Die Ressourcenuniversität. Seit 1765.

BILEVEL PROGRAMMING: REFORMULATIONS, REGULARITY, AND STATIONARITY

By the Faculty of Mathematics and Computer Science
of the Technische Universität Bergakademie Freiberg

approved

Thesis

to attain the academic degree of

Doctor rerum naturalium
(Dr.rer.nat.)

submitted by **MSc. Alain B. Zemkoho**
born on the 28th of May 1978 in Batcham, Cameroon

Reviewers: Prof. Dr. Stephan Dempe, Freiberg, Germany
Prof. Dr. Boris S. Mordukhovich, Detroit, USA

Date of the award: 12th June 2012

„Gedruckt mit Unterstützung des Deutschen Akademischen Austauschdienstes”

Versicherung

Hiermit versichere ich, dass ich die vorliegende Arbeit ohne unzulässige Hilfe Dritter und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe; die aus fremden Quellen direkt oder indirekt übernommenen Gedanken sind als solche kenntlich gemacht.

Die Hilfe eines Promotionsberaters habe ich nicht in Anspruch genommen. Weitere Personen haben von mir keine geldwerten Leistungen für Arbeiten erhalten, die nicht als solche kenntlich gemacht worden sind. Die Arbeit wurde bisher weder im Inland noch im Ausland in gleicher oder ähnlicher Form einer anderen Prüfungsbehörde vorgelegt.

15. Juni 2012 MSc. Alain B. Zemkoho

Declaration

I hereby declare that I completed this work without any improper help from a third party and without using any aids other than those cited. All ideas derived directly or indirectly from other sources are identified as such.

I did not seek the help of a professional doctorate-consultant. Only those persons identified as having done so received any financial payment from me for any work done for me. This thesis has not previously been published in the same or a similar form in Germany or abroad.

June 15, 2012 MSc. Alain B. Zemkoho

Acknowledgements

For the creation of this thesis, I would like to express all my deepest gratitude to my supervisor, Stephan Dempe, for given me at the first place, the opportunity to pursue my PhD studies here at TU Bergakademie Freiberg. Right from the beginning, he introduced me to the fascinating area of nonsmooth optimization and more generally to optimization. Before I arrived in Freiberg, I had a very narrow idea about optimization since I really never had any formal course on the topic before. Nevertheless, he was ready to work with me. Through his courses and special seminars, he taught me all the basics. He has always been there when I needed him sacrificing a lot of the little time he has to help me built a personal experience in scientific research. His insights and ideas have hugely contributed to generate the publications that provided the material for this thesis.

Through him, I also got to know a number of prominent researchers in the field of optimization and variational analysis. My special thanks go to Boris Mordukhovich, with whom we got a very intense collaboration in the last year and half, and which led to some joint works with Stephan Dempe. I am especially grateful to him for accepting to host me during part of the research in this thesis. During this time he introduced me a special class of necessary optimality conditions I did not know before, that is, the *upper subdifferential optimality conditions*, which have had a tremendous impact on this thesis. My discussions with Boris Mordukhovich from this period on have been of a great help to shape my knowledge in variational analysis. I am also very grateful to his then PhD students Hung Phan and Nghia Tran for all the logistic support they provided me with during my stay in Detroit. Robert Bruner, the associate chair of the Department of Mathematics at Wayne State University also deserves my deepest gratitude for attending to all my office needs during my stay. I would also like to say thank you to Boris Mordukhovich for agreeing to be co-referee of this thesis.

Furthermore, I am indebted to Jiří Outrata for a useful discussion about the *calmness condition*, especially for bringing the paper in reference [65] to my attention. All the colleagues and staff members of the Department of Mathematics and Computer Science of TU Bergakademie Freiberg deserve my special thanks for their multiple support and for providing me with a pleasant working environment. Among them I will mention Wolfgang Mönch for his constant support during my scholarship procedures, and for accepting to be a member of my Doctoral board.

The work in this thesis would not have been possible without the financial support of the Deutscher Akademischer Austausch Dienst (DAAD) via its prestigious "Research Grants for Doctoral Candidates and Young Academics and Scientists". Stephan Dempe also deserves a special mention for multiple financial supports, in order to enable me attend various scientific conferences, and most importantly for the travel grant that contributed to make my trip to Detroit possible.

Last but not least, I would like to sincerely thank my wife Christelle and my son Phanuel for all their emotional support and patience during these years where most of my time has been devoted to the research contained in this thesis.

Contents

Abbreviations

CQ	constraint qualification
CRCQ	constant rank constraint qualification
KKT	Karush-Kuhn-Tucker
LICQ	linear independence constraint qualification
LLVF	lower-level value function
MFCQ	Mangasarian-Fromowitz constraint qualification
MPCC	mathematical program with complementarity constraints
MPEC	mathematical program with equilibrium constraints
OPCC	optimization problem with complementarity constraints
OPEC	optimization problem with generalized equation constraint
SSOC	strong sufficient condition of second order

Notation

Spaces and orthants

$\mathbb{R}$	the real numbers
$\mathbb{R}_-$	the left-half-line
$\mathbb{R}_+$	the right-half-line
$\mathbb{R}^n$	the $n-$dimensional real vector space
$\mathbb{R}^n_-$	the nonpositive orthant in $\mathbb{R}^n$
$\mathbb{R}^n_+$	the nonnegative orthant in $\mathbb{R}^n$
$\overline{\mathbb{R}}$	$(-\infty,\infty]$, extended real line

Sets

$\mathbb{N}$	the set of natural numbers
$\mathbb{B}$	unit ball in $\mathbb{R}^n$
$\{x\}$	the set consisting of the vector x
$\{0_n\}$	origin of the space $\mathbb{R}^n$
$\operatorname{co}\Omega$	convex hull of the set Ω
$\operatorname{cl}\Omega$	topological closure of the set Ω
$\operatorname{bd}\Omega$	topological boundary of the set Ω
$\operatorname{int}\Omega$	topological interior of Ω
$\Omega_1 \subseteq \Omega_2$	Ω_1 is a subset of Ω_2
η	$\{i \mid \bar{u}_i = 0,\, g_i(\bar{x},\bar{y}) < 0\}$
θ	$\{i \mid \bar{u}_i = 0,\, g_i(\bar{x},\bar{y}) = 0\}$, biactive or degenerate index set
ν	$\{i \mid \bar{u}_i > 0,\, g_i(\bar{x},\bar{y}) = 0\}$
$\lvert\delta\rvert$	cardinality of the set δ

Vectors

$x \in \mathbb{R}^n$	column vector in $\mathbb{R}^n$
$x^\top$	the transpose of vector x
$x^\top y$	standard inner product of x and y with $x,y \in \mathbb{R}^n$
(x,y)	column vector $(x^\top,y^\top)^\top$
x_i	$i-$th component of x
x_δ	vector in $\mathbb{R}^{\lvert\delta\rvert}$ consisting of components $x_i,\ i \in \delta$
$x \leq y$	componentwise comparison $x_i \leq y_i,\ i = 1,\ldots,n$
$x < y$	strict componentwise comparison $x_i < y_i,\ i = 1,\ldots,n$
$\lVert x\rVert$	Euclidean norm of x

Functions

$\operatorname{epi}\psi$	epigraph of the function $\psi : \mathbb{R}^n \longrightarrow \overline{\mathbb{R}}$
$\widehat{\partial}\psi(\bar{x})$	Fréchet subdifferential of the function $\psi : \mathbb{R}^n \longrightarrow \overline{\mathbb{R}}$
$\partial\psi(\bar{x})$	basic subdifferential of the function $\psi : \mathbb{R}^n \longrightarrow \overline{\mathbb{R}}$
$\bar{\partial}\psi(\bar{x})$	convexified subdifferential of the function $\psi : \mathbb{R}^n \longrightarrow \overline{\mathbb{R}}$
$\nabla\psi(\bar{x})$	gradient of the function $\psi : \mathbb{R}^n \longrightarrow \mathbb{R}$ at $\bar{x}$, column vector
$\nabla\psi(\bar{x})$	the $m \times n$ Jacobian of the vector-valued function $\psi : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ (with $m \geq 2$)
$\nabla_x\psi(\bar{x},\bar{y})$	gradient of the function $\psi : \mathbb{R}^n \longrightarrow \mathbb{R}$ with respect to x
$d(x,\Omega)$	Euclidean distance between x and the set Ω
$\varphi(x)$	$\min_y \{f(x,y) \mid y \in K(x)\}$, lower-level value function
$\mathscr{L}(x,y,u)$	$\nabla_y f(x,y) + \sum_{i=1}^p u_i \nabla_y g_i(x,y)$, gradient of the lower-level Lagrangian

Set-valued mappings

$\operatorname{gph}\Psi$	graph of the set-valued mapping $\Psi : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$
$\widehat{D}^*\Psi(\bar{x},\bar{y})$	Fréchet coderivative of the set-valued mapping $\Psi : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ at $(\bar{x},\bar{y})$
$D^*\Psi(\bar{x},\bar{y})$	Mordukhovich coderivative of the set-valued mapping $\Psi : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ at $(\bar{x},\bar{y})$
$\overline{D}^*\Psi(\bar{x},\bar{y})$	convexified coderivative of the set-valued mapping $\Psi : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ at $(\bar{x},\bar{y})$
$K(x)$	$\{y \mid g(x,y) \leq 0\}$, lower-level feasible set-valued mapping
$S(x)$	$\arg\min_y \{f(x,y) \mid y \in K(x)\}$, lower-level solution set-valued mapping
$\Lambda(x,y)$	$\{u \mid u \geq 0,\ u^\top g(x,y) = 0,\ \mathscr{L}(x,y,u) = 0\}$, lower-level multipliers set-valued mapping

Cones

$\widehat{N}_\Omega(\bar{x})$	Fréchet normal cone to the set $\Omega \subseteq \mathbb{R}^n$ at the point $\bar{x}$
$N_\Omega(\bar{x})$	basic normal cone to the set $\Omega \subseteq \mathbb{R}^n$ at the point $\bar{x}$
$\overline{N}_\Omega(\bar{x})$	convexified normal cone to the set $\Omega \subseteq \mathbb{R}^n$ at the point $\bar{x}$
$T_\Omega(\bar{x})$	Boulingand tangent cone to the set $\Omega \subseteq \mathbb{R}^n$ at the point $\bar{x}$
$\overline{T}_\Omega(\bar{x})$	Clarke tangent cone to the set $\Omega \subseteq \mathbb{R}^n$ at the point $\bar{x}$

Sequences

x^k	sequence in $\mathbb{R}^n$
$x^k \to x$	the sequence x^k converges to x
$x^k \downarrow x$	sequence in $\mathbb{R}$ converging to x and $x^k > x$ for all $k \in \mathbb{N}$

Problems

(P)	classical optimistic bilevel programming problem
(P_o)	original optimistic bilevel programming problem
(P_p)	pessimistic bilevel programming problem

1 Introduction

The focus of this thesis is on the hierarchical optimization problem

$$\text{“}\min_x\text{”}\{F(x,y)|\, x\in X,\; y\in S(x)\} \tag{1.1}$$

commonly known as *bilevel programming problem*, where the upper-level player (leader) intends to minimize his/her cost function F with respect to the variable x while taking into account the reaction y of the lower-level player (follower). Here $S : X \subset \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ is a set-valued mapping defined by

$$S(x) := \arg\min_y \{f(x,y)|\, y\in K(x)\}, \tag{1.2}$$

which describes sets of optimal solutions of the lower-level parametric optimization problem

$$\min_y \{f(x,y)|\, y\in K(x)\} \tag{1.3}$$

for any given choice $x \in X$ of the leader. The sets X and $K(x)$ are usually called upper-level and lower-level feasible sets, respectively. For simplicity in expressing the main ideas, we confine ourselves to the case where the upper and lower-level constraint sets are given explicitly as

$$X := \{x\in\mathbb{R}^n|\, G(x)\le 0\} \;\text{ and }\; K(x) := \{y\in\mathbb{R}^m|\, g(x,y)\le 0\}, \tag{1.4}$$

respectively, with $G : \mathbb{R}^n \to \mathbb{R}^k$ and $g : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^p$. Furthermore, all the functions involved are assumed to be continuously differentiable. Nonetheless, discussions are provided on how to extend most of the results to the case of equality and other types of constraints as well as to the case of nonsmooth functions via appropriate subdifferential constructions. Also recall that $F : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ and $f : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ are (single-valued) upper-level and lower-level objective/cost functions, respectively.

One way to understand the general bilevel program (1.1) is to treat it as the following set-valued optimization problem:

$$\min_{x\in X} F(x,S(x)) := \bigcup_{y\in S(x)} \{F(x,y)\}, \tag{1.5}$$

where the minimization is considered with respect to some ordering cone. Such an approach is discussed in [29] while related developments are given in [4, 7, 121] to derive necessary optimality conditions for multiobjective bilevel programs. In order to investigate problem (1.1) from the viewpoint of scalar-objective optimization, observe that (1.5) becomes a usual optimization problem

$$\min \{F(x,S(x))|\; x\in X\} \tag{1.6}$$

provided that $S(x)$ is single-valued for all $x \in X$; see [21] and the references therein for details on algorithms and necessary optimality conditions via this approach. It is worth mentioning that the main difficulty in both approaches (1.5) and (1.6) is the implicit nature of the objective.

If we can not ensure the uniqueness of optimal solutions for the lower-level problem (1.3), it has been noted earlier (see, e.g., [21]) that the formulation of problem (1.1) is ambiguous from the view point of scalar-objective optimization. This is the reason for the quotation marks in (1.1). To overcome this, two main approaches have been suggested in the literature. On one hand, we have the *optimistic formulation*

$$(\mathrm{P}_o) \qquad \min_{x\in X} \varphi_o(x) := \min_y \{F(x,y)|\, y\in S(x)\}. \tag{1.7}$$

From the economics viewpoint this corresponds to a situation where the follower participates in the profit of the leader, i.e., some cooperation is possible between both players on the upper and lower levels. However, it would not always be possible for the leader to convince the follower to make choices that are favorable for him or her. Hence it is necessary for the upper-level player to bound damages resulting from undesirable selections on the lower-level. This gives the *pessimistic formulation* of the bilevel optimization problem as follows:

$$(\mathrm{P}_p) \qquad \min_{x\in X} \varphi_p(x) := \max_{y} \{F(x,y)|\, y\in S(x)\}. \tag{1.8}$$

The latter problem is a special class of minimax problems. Static minimax problems, corresponding in our case to a situation where the feasible set of the inner problem $S(x)$ is independent of x, have been intensively investigated in the literature; see, e.g., [39, 112, 113]. At the same time, it has been well recognized that when $S(x)$ stands for varying sets of solutions to another optimization problem, the pessimistic formulation above faces many challenges. Some of them are highlighted in [21] and the references therein.

The optimistic formulation (P_o) is much simpler to handle and has therefore been the most investigated, however in the simplified version

$$(\mathrm{P}) \qquad \min_{x,y} \{F(x,y)|\, x\in X,\, y\in S(x)\}, \tag{1.9}$$

where the difficulty in (P_o) is shifted from the objective to the constraints. Problem (P) that we label here as *classical/auxiliary bilevel program* has concentrated the most attention in the literature where very often it is simply called "bilevel optimization problem" or "optimistic bilevel optimization problem". Three monographs have been devoted to it by Shimizu et al. [113], Bard [5] and Dempe [21]. About 50 pages of the Encyclopedia [52, Pages 135-185] provide a thorough overview on the theory and applications of problem (P). A more recent overview on the problem is provided by Colson et al. [18]. An annotated bibliography on bilevel optimization, mostly devoted to (P) and the related mathematical programs with equilibrium constraints (MPECs) is given by Dempe [22]. An earlier bibliography review was given by Vicente and Calamai [115] which includes discussions on the more general multilevel programming problem. The books by Luo et al. [77] and Outrata et al. [101] on MEPCs are also important sources of information as far as problem (P) is concerned.

Obviously, problem (P) and the original optimistic bilevel program (P_o) are equivalent when global solutions are in question. We show in Chapter 5 (see Section 5.3) that a local optimal solution of problem (P_o) corresponds to a local optimal solution of (P) without any assumption. However, it is very easy to find examples where a local optimal solution of (P) does not generate a local optimal solution of (P_o), cf. Section 5.3. Thus we also show in the latter section that a local optimal solution of (P) corresponds to that of (P_o) under a rather strong condition, namely, the inner semicontinuity of the solution map related to the value function in (1.7). This clearly means that there is a big gap between the original optimistic bilevel optimization problem (P_o) and the auxiliary model (P), most importantly because they are both nonconvex optimization problems, thus finding local optimal solutions is the most likely achievable goal in practical solution methods.

As far as (P_o) and (P_p) are concerned, it is also easy to construct examples of bilevel programs where (P_o) has an optimal solution while (P_p) does not have any. Even when both problems have optimal solutions, they may differ from each other; see, e.g., [21, 73] for more discussions. In contrary to (P), problems (P_o) and (P_p) have been given very little attention in the literature. For some algorithmic issues in particular classes of problem (P_p) we refer the reader to [10, 73, 74] and the bibliographies therein. For the first time, a detailed investigation on necessary optimality conditions for the original optimistic formulation (P_o) is conducted in [31] and in [30] for the pessimistic problem (P_p). Before, we were not familiar with any work on optimality conditions, apart from [19], where several reformulations of (P_p) were suggested and necessary conditions are derived in terms of the coderivative of certain set-valued mappings in some special cases. In [21, Chapter 3] some optimality conditions are suggested for the linear pessimistic bilevel programming problem while assuming among other things that the functions φ_o and φ_p coincide at the reference point. It has been well recognized that the optimistic and pessimistic reformulations of the bilevel program (1.1) are optimization problems with objectives of the marginal/value function type. In fact, in both cases (P_o) and (P_p) the objectives are described by *two-level value function*, i.e., as the optimal value function of a parametric optimization problem partly constrained by another one. Most of the material in Chapters 4–6 of this thesis is drawn from the articles [30, 31].

Clearly, one of the main concerns of the thesis is to further the understanding of the bilevel programming problem (1.1) while providing a unified framework for the three major reformulations (P), (P_o) and (P_p) spread throughout the literature. Most attention here is given to necessary optimality conditions. Let us recall that the first attempt to derive optimality conditions in bilevel programming (precisely for (P)) was made in 1984 by Bard [5]. Later, part of his results were shown to contain an error [15]. Considering the case where the lower-level problem is strongly stable, first results were obtained by Dempe [20] (in a primal form) and by Outrata [99] (in a dual form) via the implicit function approach (1.6). Results on optimality conditions accommodating nonunique lower-level solution mappings can be dated back to the papers by Ye and Zhu [128] and Chen and Florian [13]. In both papers, the so-called *lower-level value function* (LLVF) reformulation, introduced by Outrata [98] for the purpose of a numerical method, is considered. In the paper [13] the authors later switch to the classical KKT reformulation. Hence, the first result of KKT-type necessary optimality conditions for (P) via the LLVF reformulation are due to Ye and Zhu [128]. Since then this reformulation has attracted a lot of interest, see e.g. [2, 26, 27, 91, 129, 130] and their references for a number of important contributions. Chen and Florian [13] are also credited to be among the first to have pointed out the failure of some well-known CQs for a class of problem (P).

The next chapter presents basic notions and results of variational analysis and generalized differentiation widely used in the subsequent chapters. A general background on optimization problems with geometric and operator constraints is also provided in this chapter. In Chapter 3, the theory on necessary optimality conditions for (P) is revisited strictly from the view point of nonsmooth and variational analysis, especially via the generalized differentiation calculus rules by Boris Mordukhovich. We consider the three major one-level representations of the problem, i.e. the LLVF, KKT and OPEC (optimization problem with generalized equation constraints) reformulations. The material in this chapter, mostly taken from the articles [35, 33, 34], is at the heart of the developments in the thesis, in the sense that the reformulations and stationarity concepts discussed here pave the way for a full understanding of Chapters 5 and 6 devoted to the original optimistic and pessimistic problems, respectively.

Chapter 4 is mainly concerned with sensitivity analysis of OPCC (optimization problems with complementarity constraints) and OPEC value functions. Here we derive upper estimates of the limiting subdifferential for such functions from various perspectives, depending on the type of optimality/stationary conditions of interest for the the original bilevel model (P_o). It should be mentioned that the results in Section 4.1 can stand on their own. Indeed, they also provide efficient rules to obtain estimates of the coderivative and the fulfillment of the Lipschitz-like property for mappings of special structures (inequality and equality systems with complementarity constraints) important for other classes of optimization-related problems, not just for bilevel programming. Sensitivity analysis of OPEC value functions, which is of its own interest as well, is conducted in Section 4.2. This chapter is a real intersection between Chapter 3 and 5 as it provides background for a second approach to generate the necessary optimality conditions obtained in the Chapter 3 via the coderivative estimates of the lower-level solution set-valued mapping. On the other hand it provides sensitivity results for value functions readily applicable to stability analysis of the two-level value function, which then lead to the corresponding stationarity conditions for (P_o) in Chapter 5.

Part of Chapter 5 (see Subsection 5.1.1) mainly deals with applications of the results from the previous chapter to sensitivity analysis of the two-level value function φ_o (1.7) via the OPCC and OPEC approaches. Secondly, in Subsection 5.1.2, we develop lower-level value function approach to analyze φ_o. Here a detailed discussion is given on rules to derive subdifferential estimates and establish the local

Lipschitz continuity of φ_o from a perspective completely different from the previous ones. In the concluding section of this chapter we employ the results obtained above to deriving necessary optimality conditions for the original optimistic formulation (P_o) in the various forms of stationarity concepts.

In Chapter 6, we consider the pessimistic bilevel programming problem (P_p). One major difficulty to handle this problem relates to the fact that the objective function φ_p in (1.8) is usually only upper semicontinuous, which makes it hard to detect optimal solutions by using the conventional "lower subdifferential" objects that work well for lower semicontinuous functions. Secondly, observe that (P_p) is a special minimax problem. Although many publications have been devoted to optimality conditions for the latter class of problem, most of them can not be applied to our pessimistic program because the corresponding inner problem $\max_y \{F(x,y) | y \in S(x)\}$ violates the imposed constraint qualifications (CQs). In this chapter we develop two approaches to establish necessary optimality conditions for pessimistic bilevel programs. Our first approach to derive *lower* (i.e., conventional) subdifferential optimality conditions for (P_p) is to get the lower semicontinuity of the two-level value function φ_p by constructing frameworks where it is actually locally Lipschitz continuous. This would be possible by applying the results obtained in Chapter 5 for the two-level value function approach of the optimistic problem (P_o). Thus we obtain various types of first-order necessary optimality conditions via the LLVF and KKT reformulations of φ_p. The second approach considered in Section 6.3 seems to be more suitable for particular structures of the pessimistic program. Namely, we derive the so-called *upper* subdifferential optimality conditions for (P_p) in the sense of [84], which do not just require the function φ_p to be lower semicontinuous, but eventually leads to much stronger necessary optimality conditions in certain important settings.

Chapter 7 deals with applications of bilevel programming in transportation. After providing a general introduction to the field, we focus on deriving stationarity conditions for two representative problems, i.e. the bilevel road pricing problem and the Origin-Destination (O-D) matrix adjustment problem. Using material from Chapter 3, we develop implementable necessary optimality conditions for the road pricing problem containing all the necessary information. The issue of estimating the (fixed) demand required for the road pricing problem is a quite difficult problem which has been also addressed in recent years using bilevel programming. Thus we show how the ideas used in the previous case can be applied to obtain optimality conditions for the O-D matrix estimation problem. The material here is drawn from the articles [35, 36].

2 Preliminaries

2.1 Background material

In this section we introduce the basic tools from variational analysis that will be used throughout the thesis. The generalized differentiation objects that will be mostly utilized are the subdifferentials, normal cones and coderivatives by Mordukhovich. However, their interrelations with the corresponding Fréchet and Clarke nonsmooth tools will often be needed. More details on the material briefly discussed here can be found in the books [16, 89, 109, 111] and the references therein. Let us start by recalling that the notion of *Painlevé-Kuratowski outer/upper limit* plays a central role in variational analysis when evaluating nearby values of a given set-valued mapping around a reference point. It is defined for a set-valued mapping $\Psi : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ as $x \to \bar{x}$ by

$$\operatorname*{Limsup}_{x\to\bar{x}} \Psi(x) := \{v \in \mathbb{R}^m | \exists x^k \to \bar{x},\, v^k \to v \text{ with } v^k \in \Psi(x^k) \text{ as } k \to \infty\}. \tag{2.1}$$

2.1.1 Subdifferentials

Given an extended real-valued function $\psi : \mathbb{R}^n \to \overline{\mathbb{R}} := (-\infty, \infty]$, the *Fréchet/regular* (lower) subdifferential of ψ at $\bar{x} \in \operatorname{dom}\psi := \{x \in \mathbb{R}^n |\, \psi(x) < \infty\}$ is given by

$$\widehat{\partial}\psi(\bar{x}) := \left\{ v \in \mathbb{R}^n | \liminf_{x\to\bar{x}} \frac{\psi(x) - \psi(\bar{x}) - \langle v, x - \bar{x}\rangle}{\|x - \bar{x}\|} \geq 0 \right\}$$

whereas the the *Fréchet upper* subdifferential of ψ at $\bar{x} \in \operatorname{dom}\psi$ is defined by

$$\widehat{\partial}^+\psi(\bar{x}) := -\widehat{\partial}(-\psi)(\bar{x}). \tag{2.2}$$

The *basic/Mordukhovich* (lower) subdifferential of ψ at $\bar{x} \in \operatorname{dom}\psi$ is defined via the Kuratowski-Painlevée upper limit (2.1) by

$$\partial\psi(\bar{x}) := \operatorname*{Limsup}_{x\to\bar{x}} \widehat{\partial}\psi(x), \tag{2.3}$$

while its *upper* counterpart $\partial^+\psi$ can be defined analogously to (2.2) via the above lower basic subdifferential (2.3)

$$\partial^+\psi(\bar{x}) := -\partial(-\psi)(\bar{x}). \tag{2.4}$$

Provided the function ψ is Lipschitz continuous around $\bar{x} \in \operatorname{dom}\psi$ the *convexified/Clarke* subdifferential can be defined by

$$\bar{\partial}\psi(\bar{x}) := \operatorname{co}\partial\psi(\bar{x}), \tag{2.5}$$

where “co" stands for the convex hull of the set in question. Thanks to this link between the basic and convexified subdifferentials, we have the following *convex hull property* which plays a determinant role in this thesis:

$$\operatorname{co}\partial(-\psi)(\bar{x}) = -\operatorname{co}\partial\psi(\bar{x}). \tag{2.6}$$

For this equality to hold, ψ should be Lipschitz continuous near $\bar{x}$.

If the function ψ is strictly differentiable at $\bar{x}$, that is

$$\lim_{v\to\bar{x},\, x\to\bar{x}} \frac{\psi(v) - \psi(x) - \langle\nabla\psi(\bar{x}), v - x\rangle}{\|v - x\|} = 0 \tag{2.7}$$

(with $\nabla\psi(\bar{x})$ denoting the classical gradient of ψ at $\bar{x}$), then we have

$$\bar{\partial}\psi(\bar{x}) = \widehat{\partial}\psi(\bar{x}) = \partial\psi(\bar{x}) = \{\nabla\psi(\bar{x})\}. \tag{2.8}$$

It is worth mentioning that these three subdifferentials also coincide with the the subdifferential in the sense of convex analysis provided ψ is convex, precisely, if the latter holds, then we have

$$\bar{\partial}\psi(\bar{x}) = \widehat{\partial}\psi(\bar{x}) = \partial\psi(\bar{x}) := \{v \in \mathbb{R}^n |\ \psi(x) - \psi(\bar{x}) \geq \langle v, x - \bar{x}\rangle, \forall x \in \mathbb{R}^n\}. \tag{2.9}$$

Also of importance, note that every function which is locally Lipschitzian around $\bar{x}$ is strictly differentiable at $\bar{x}$ provided that its basic subdifferential (2.3) is a singleton.

The function ψ is said to be lower (resp. upper) regular at $\bar{x}$ if one has

$$\widehat{\partial}\psi(\bar{x}) = \partial\psi(\bar{x}) \quad (\text{resp. } \widehat{\partial}^+\psi(\bar{x}) = \partial^+\psi(\bar{x})). \tag{2.10}$$

Obviously, ψ is upper regular if and only if $-\psi$ is lower regular at the point in question. Also note that if ψ is strictly differentiable at $\bar{x}$ then it is upper and lower level regular, cf. (2.8). On the other hand the function ψ is lower (resp. upper) level regular provided it is convex (resp. concave), cf. (2.9). For the remainder of this thesis, recall that by convention, if $\bar{x} \notin \operatorname{dom}\psi$, then we have

$$\bar{\partial}\psi(\bar{x}) = \widehat{\partial}\psi(\bar{x}) = \partial\psi(\bar{x}) = \emptyset.$$

However, in case ψ is Lipschitz continuous around $\bar{x}$, then $\partial\psi(\bar{x})$ is nonempty and compact. In general though for a given function $\psi : \mathbb{R}^n \to \overline{\mathbb{R}} := (-\infty, \infty]$, we always have

$$\widehat{\partial}\psi(\bar{x}) \subseteq \partial\psi(\bar{x}) \subseteq \bar{\partial}\psi(\bar{x}). \tag{2.11}$$

The most important subdifferential calculus rules, that will also be useful in this thesis, are the sum and chain rules. Consider two functions ψ_1 and ψ_2 which are locally Lipschitz continuous around $\bar{x}$, and let λ_1 and λ_2 be two nonnegative real numbers. Then we have the sum rule

$$\partial(\lambda_1\psi_1 + \lambda_2\psi_2)(\bar{x}) \subseteq \lambda_1\partial\psi_1(\bar{x}) + \lambda_2\partial\psi_2(\bar{x}), \tag{2.12}$$

where equality holds if ψ_1 or ψ_2 is continuously differentiable at $\bar{x}$ [87, Corollary 4.6]. As for the chain rule, let $\psi_1 : \mathbb{R}^n \to \mathbb{R}^m$ be Lipschitz continuous around $\bar{x}$, and $\psi_2 : \mathbb{R}^m \to \bar{\mathbb{R}}$ Lipschitz continuous around $\bar{y} = \psi_1(\bar{x}) \in \operatorname{dom}\psi_2$. Then we have

$$\partial(\psi_2 \circ \psi_1)(\bar{x}) \subseteq \bigcup\big\{\partial\langle u, \psi_1\rangle(\bar{x}) \,\big|\, u \in \partial\psi_2(\bar{y})\big\}, \tag{2.13}$$

cf. [83, Proposition 2.10]. More details on these results and related properties can be found in the papers [83, 87] and also the books [89, 109].

2.1.2 Normal cones

Given a nonempty set $\Omega \subseteq \mathbb{R}^n$, the *basic/Mordukhovich normal cone* to it at $\bar{x} \in \Omega$ corresponding to the subdifferential construction (2.3) is defined by

$$N_\Omega(\bar{x}) := \operatorname*{Limsup}_{x \to \bar{x}\,(x \in \Omega)} \widehat{N}_\Omega(x) \tag{2.14}$$

via the outer limit (2.1) of the regular counterpart

$$\widehat{N}_\Omega(x) := \left\{ v \in \mathbb{R}^n |\ \limsup_{u \to x (u \in \Omega)} \frac{\langle v, u - x\rangle}{\|u - x\|} \leq 0 \right\} \tag{2.15}$$

at points $x \in \Omega$ near $\bar{x}$. Note that for sets $\Omega \subseteq \mathbb{R}^n$ locally closed around $\bar{x}$ the given definition (2.14) reduces to the original one

$$N_\Omega(\bar{x}) = \operatorname*{Limsup}_{x \to \bar{x}} [\text{cone}(x - \Pi_\Omega(x))].$$

introduced in [85], where the symbol "cone" stands for the conic hull of the corresponding set, and where Π denotes the Euclidean projection on the set in question. The *convexified/Clarke normal cone* is defined via the convex closure of the basic normal cone by

$$\overline{N}_\Omega(\bar{x}) := \text{clco} N_\Omega(\bar{x}). \tag{2.16}$$

Consider the Boulingand/contingent and Clarke's tangent cone, respectively, to Ω at some point $\bar{x} \in \Omega$ defined by

$$\begin{aligned} T_\Omega(\bar{x}) &:= \{u \in \mathbb{R}^n |\ \exists t^k \downarrow 0,\ u^k \to u:\ \bar{x} + t^k u^k \in \Omega\}, \\ \overline{T}_\Omega(\bar{x}) &:= \{u \in \mathbb{R}^n |\ \forall t^k \downarrow 0,\ x^k \to \bar{x}\ (x^k \in \Omega);\ \exists u^k \to u:\ x^k + t^k u^k \in \Omega\}. \end{aligned} \tag{2.17}$$

Then according to [109, Theorem 6.28] and [16], respectively, the regular and convexified normal cones can also be obtained as the polar of the Boulingand and the Clarke tangent cones, respectively:

$$\begin{aligned} \widehat{N}_\Omega(\bar{x}) &= \{v \in \mathbb{R}^n | \langle v, u \rangle \leq 0,\ \forall u \in T_\Omega(\bar{x})\}, \\ \overline{N}_\Omega(\bar{x}) &= \{v \in \mathbb{R}^n | \langle v, u \rangle \leq 0,\ \forall u \in \overline{T}_\Omega(\bar{x})\}. \end{aligned} \tag{2.18}$$

Furthermore, the regular and the convexified normal cones are always convex sets, while the basic normal cone is in general nonconvex. For this reason the latter cone cannot be polar to any tangential approximation of Ω [89]. Note however that the basic normal cone possesses full calculus (cf. e.g. [89, 109]), which is not the case for the other cones. Moreover we always have the following chain of inclusion at any point $\bar{x} \in \Omega$

$$\widehat{N}_\Omega(\bar{x}) \subseteq N_\Omega(\bar{x}) \subseteq \overline{N}_\Omega(\bar{x}). \tag{2.19}$$

Hence, proceeding with the basic normal cone one obtains sharper optimality conditions as with the convexified one. In case the set Ω is convex, all the above normal cones coincide with the normal cone in the sense of convex analysis; this corresponds to the normal cone counterpart of (2.9):

$$\widehat{N}_\Omega(\bar{x}) = N_\Omega(\bar{x}) = \overline{N}_\Omega(\bar{x}) = \{v \in \mathbb{R}^n |\ \langle v, u - \bar{x} \rangle \leq 0, \forall u \in \Omega\}. \tag{2.20}$$

Using the normal cone (2.14), we can equivalently describe the basic subdifferential (2.3) by

$$\partial \psi(\bar{x}) = \{v \in \mathbb{R}^n |\ (v, -1) \in N_{\text{epi}\psi}(\bar{x}, \psi(\bar{x}))\} \tag{2.21}$$

for lower semicontinuous (l.s.c.) functions via the normal cone to the epigraph of ψ (denoted by epiψ). Similarly, the *singular subdifferential* of ψ at $\bar{x} \in \text{dom}\,\psi$ can be defined

$$\partial^\infty \psi(\bar{x}) := \{v \in \mathbb{R}^n |\ (v, 0) \in N_{\text{epi}\psi}(\bar{x}, \psi(\bar{x}))\}. \tag{2.22}$$

It is worth mentioning that for a function ψ l.s.c. around $\bar{x}$, we have $\partial^\infty \psi(\bar{x}) = \{0\}$ if and only if ψ is locally Lipschitzian around this point. Analogously to (2.21) the regular and convexified normal cones can equivalently be defined via the regular (2.15) and convexified (2.16) normal cones, respectively.

We now conclude this subsection by recalling some useful properties related to normal cones. A set Ω will be said to be regular at a point $\bar{x} \in \Omega$ if we have

$$N_\Omega(\bar{x}) = \widehat{N}_\Omega(\bar{x}), \tag{2.23}$$

which is obviously the case if Ω is a convex set (2.20). For completeness note that analogously to (2.11), it is a convention that all the above normal cones be empty provided $\bar{x} \notin \Omega$, while for $\bar{x} \in \text{int}\,\Omega$, they all coincide with the origin of the space $\{0\}$.

2.1.3 Coderivatives

In this subsection we consider a set-valued mapping $\Psi : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ and its graph

$$\text{gph}\,\Psi := \{(x,y) \in \mathbb{R}^n \times \mathbb{R}^m | y \in \Psi(x)\}.$$

The Mordukhovich coderivative of Ψ at $(\bar{x},\bar{y}) \in \text{gph}\,\Psi$ introduced in [88] is a positively homogeneous mapping $D^*\Psi(\bar{x},\bar{y}) : \mathbb{R}^m \rightrightarrows \mathbb{R}^n$ defined by

$$D^*\Psi(\bar{x},\bar{y})(v) := \{u \in \mathbb{R}^n | (u,-v) \in N_{\text{gph}\,\Psi}(\bar{x},\bar{y})\} \text{ for } v \in \mathbb{R}^m, \tag{2.24}$$

via the normal cone (2.14) to the graph of Ψ. If Ψ is single-valued and locally Lipschitzian around $\bar{x}$, its coderivative can be represented analytically as

$$D^*\Psi(\bar{x})(v) = \partial\langle v, \Psi\rangle(\bar{x}) \text{ for } v \in \mathbb{R}^m,$$

via the basic subdifferential (2.3) of the Lagrange scalarization $\langle v, \Psi\rangle(x) := \langle v, \Psi(x)\rangle$, where the component $\bar{y}$ $(= \Psi(\bar{x}))$ is omitted in the coderivative notation for single-valued mappings. This implies the coderivative representation

$$D^*\Psi(\bar{x})(v) = \{\nabla\Psi(\bar{x})^\top v\} \text{ for } v \in \mathbb{R}^m,$$

when Ψ is strictly differentiable at $\bar{x}$ as in (2.8) with $\nabla\Psi(\bar{x})$ standing for its Jacobian matrix at $\bar{x}$ and with the symbol "$^\top$" standing for transposition. Similarly, the *Fréchet/regular* and *Clarke/convexified* coderivatives can respectively be defined as

$$\begin{aligned} \widehat{D}^*\Psi(\bar{x},\bar{y})(v) &:= \{u \in \mathbb{R}^n | (u,-v) \in \widehat{N}_{\text{gph}\,\Psi}(\bar{x},\bar{y})\} \text{ for } v \in \mathbb{R}^m, \\ \overline{D}^*\Psi(\bar{x},\bar{y})(v) &:= \{u \in \mathbb{R}^n | (u,-v) \in \overline{N}_{\text{gph}\,\Psi}(\bar{x},\bar{y})\} \text{ for } v \in \mathbb{R}^m, \end{aligned} \tag{2.25}$$

via the regular (2.15) and convexified (2.16) normal cones, respectively. Details on these three coderivative concepts and relationships between them can be found in [87, 89].

2.1.4 Continuity concepts of set-valued mappings

Throughout this subsection we consider a set-valued mapping $\Psi : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$. If Ψ is a positively homogeneous mapping, its outer norm (resp. inner norm) is defined by

$$\|\Psi\|^+ := \sup_{x\in\mathbb{B}} \sup_{u\in\Psi(x)} \|u\| \quad \big(\text{resp. } \|\Psi\|^- := \sup_{x\in\mathbb{B}} \inf_{u\in\Psi(x)} \|u\|\big). \tag{2.26}$$

Moreover, considering a cone C, the restricted inner norm of Ψ on C defined by

$$\|\Psi|_C\|^- := \sup_{x\in\mathbb{B}\cap C} \inf_{u\in\Psi(x)} \|u\| \tag{2.27}$$

will also play an important role in what follows. In (2.26) and (2.27), $\mathbb{B}$ stands for the unit ball in $\mathbb{R}^n$.

Inner semicompactness. Ψ is inner semicompact at $\bar{x}$ with $\Psi(\bar{x}) \neq \emptyset$ if for every sequence $x_k \to \bar{x}$ with $\Psi(x_k) \neq \emptyset$ there is a sequence of $y_k \in \Psi(x_k)$ that contains a convergent subsequence as $k \to \infty$. It follows that the inner semicompactness holds in finite dimensions whenever Ψ is uniformly bounded around $\bar{x}$, i.e., there exists a neighborhood U of $\bar{x}$ and a bounded set $\Omega \subseteq \mathbb{R}^m$ such that

$$\Psi(x) \subseteq \Omega \;\text{ for all }\; x \in U. \tag{2.28}$$

The inner semicompactness of Ψ at $\bar{x}$ is also automatical, provided the following inner semicontinuity is satisfied at $(\bar{x},\bar{y}) \in \text{gph}\,\Psi$.

Inner semicontinuity. The mapping Ψ is inner semicontinuous at $(\bar{x},\bar{y}) \in \text{gph}\,\Psi$ if for every sequence

$x_k \to \bar{x}$ there is a sequence of $y_k \in \Psi(x_k)$ that converges to $\bar{y}$ as $k \to \infty$. If Ψ is inner semicompact at $\bar{x}$ and $\Psi(\bar{x}) = \{\bar{y}\}$, then Ψ is inner semicontinous at $(\bar{x}, \bar{y})$. The latter property reduces to the usual continuity for single-valued mappings while in the general set-valued case it is implied by the Lipschitz-like property of Ψ around $(\bar{x}, \bar{y}) \in \text{gph}\,\Psi$.

Lipschitz-like/Aubin property. Ψ is Lipschitz-like around $(\bar{x}, \bar{y}) \in \text{gph}\,\Psi$ if there are neighborhoods U of $\bar{x}$, V of $\bar{y}$, and a constant $\ell > 0$ such that

$$\Psi(x) \cap V \subseteq \Psi(u) + \ell \|x - u\| \mathbb{B} \;\text{ for all } x, u \in U, \tag{2.29}$$

with $\mathbb{B}$ denoting the unit ball in $\mathbb{R}^m$. When $V = \mathbb{R}^m$ in (2.29), this property reads as to the classical local Lipschitz continuity of Ψ around $\bar{x}$. A complete characterization of the Lipschitz-like property (2.29), and hence a sufficient condition for the inner semicontinuity of Ψ at $(\bar{x}, \bar{y})$, is given for closed-graph mappings by the following *coderivative/Mordukhovich criterion* (see [89, Theorem 4.10] and [109, Theorem 9.40]):

$$D^*\Psi(\bar{x}, \bar{y})(0) = \{0\}. \tag{2.30}$$

Furthermore, the infimum of all $\ell > 0$ for which (2.29) holds, also known as the Lipschitz modulus of Ψ at $(\bar{x}, \bar{y})$, is given via the outer norm of the coderivative of Ψ (2.26):

$$\text{lip}\,\Psi(\bar{x}, \bar{y}) := \inf\{\ell \in (0, \infty) |\; (2.29) \;\text{ holds for some }\; U \;\text{ and }\; V\} = \|D^*\Psi(\bar{x}, \bar{y})\|^+. \tag{2.31}$$

Calmness property. Ψ will be said to be calm at some point $(\bar{x}, \bar{y}) \in \text{gph}\,\Psi$, if there exist neighborhoods U of $\bar{x}$, V of $\bar{y}$, and a constant $\ell > 0$ such that

$$\Psi(x) \cap V \subseteq \Psi(\bar{x}) + \ell \|x - \bar{x}\| \mathbb{B} \;\text{ for all } x \in U. \tag{2.32}$$

Clearly, if we fix $x = \bar{x}$ in (2.29), we get (2.32). Thus Ψ is automatically calm at $(\bar{x}, \bar{y})$, if it Lipschitz-like around the same point. Moreover if $V = \mathbb{R}^m$ in (2.32), the resulting property corresponds to the upper Lipschitz property of Robinson [106]. An equivalent formulation of relationship (2.32) that will often be used in this thesis is:

$$d(y, \Psi(\bar{x})) \leq \ell \|x - \bar{x}\| \;\text{ for all }\; x \in U \;\text{ and }\; y \in \Psi(x) \cap V, \tag{2.33}$$

cf. e.g. [61]. A similar reformulation corresponding to the Lipschitz-like property (2.29) is self-evident. Similarly to (2.31), the calmness modulus of Ψ at the point $(\bar{x}, \bar{y})$ can be expressed in terms of the restricted inner norm of the coderivative of Ψ (2.27):

$$\begin{aligned} \text{cal}\,\Psi(\bar{x}, \bar{y}) &:= \inf\{\ell \in (0, \infty) |\; (2.32) \;\text{ holds for some }\; U \;\text{ and }\; V\} \\ &= \limsup_{u \in \Psi(\bar{x}),\, u \to \bar{y}} \|D^*\Psi(\bar{x}, u)|_{-\widehat{N}_{\Psi(\bar{x})}(u)}\|^-, \end{aligned} \tag{2.34}$$

provided Ψ is weakly L-subsmooth, which holds in particular if the graph of Ψ is convex, cf. [134, Corollary 4.10]. For more details on the latter estimate of the calmness modulus and the definition of L-subsmooth and weakly L-subsmooth mappings, see the paper [134].

Local upper Lipschitzian selection. First recall that a function $f : D \subseteq \mathbb{R}^n \to \mathbb{R}^m$ is local upper Lipschitzian at $\bar{x}$ (in the sense of Robinson [107]) if there is a neighborhood U of $\bar{x}$ and a constant $\ell \geq 0$ such that

$$\|f(x) - f(\bar{x})\| \leq \ell \|x - \bar{x}\| \text{ for all } x \in U \cap D. \tag{2.35}$$

A set-valued mapping $\Psi : D \subseteq \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ is said to admit a *local upper Lipschitzian selection* at $(\bar{x}, \bar{y})$ if there is a single-valued mapping $f : D \subseteq \mathbb{R}^n \to \mathbb{R}^m$, which is local upper Lipschitzian at $\bar{x}$ satisfying $f(\bar{x}) = \bar{y}$ and $f(x) \in \Psi(x)$ for all $x \in D$ in a neighborhood of $\bar{x}$.

2.2 A general optimization problem

We first consider the optimization problem with the very general geometric constraint

$$\min \{f(x) |\, x \in \Omega\}, \tag{2.36}$$

where the function $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ and the set $\Omega \subseteq \mathbb{R}^n$. Next we provide the upper subdifferential necessary optimality condition (2.37) and the lower subdifferential necessary optimality condition (2.39), in the sense of Mordukhovich, see [84, Theorem 3.1] or [89, Chapter 5].

Theorem 2.2.1 (upper and lower subdifferential necessary optimality conditions under a geometric constraint)**.** *Let $\bar{x}$ be a local optimal solution to problem* (2.36) *with $|f(\bar{x})| < \infty$. Then the following assertions are satisfied:*

(i) *The following* upper subdifferential necessary optimality condition *is automatical:*

$$-\widehat{\partial}^+ f(\bar{x}) \subseteq \widehat{N}_\Omega(\bar{x}). \tag{2.37}$$

(ii) *If in addition, f is lower semicontinuous, Ω is closed around $\bar{x}$ and the qualification condition*

$$\partial^\infty f(\bar{x}) \cap (-N_\Omega(\bar{x})) = \{0\} \tag{2.38}$$

holds at $\bar{x}$, which is the case, in particular, when f is Lipschitz continuous around $\bar{x}$. Then, we have the following lower subdifferential necessary optimality condition:

$$0 \in \partial f(\bar{x}) + N_\Omega(\bar{x}). \tag{2.39}$$

Clearly, the lower subdifferential necessary optimality condition (2.39) is the necessary optimality condition of problem (2.36) in the usual sense. Hence, from time to time, the prefix "lower subdifferential" will be omitted when such conditions are investigated. However, the upper subdifferential necessary optimality condition (2.37) appears to be usually stronger, as it will be clear in Chapter 6, in the context of the pessimistic bilevel optimization problem. More details on these optimality conditions and related commentaries are provided in [84] (also see [89, Chapter 5]). In the the next subsection, we focus our attention on lower subdifferential necessary optimality conditions of a problem with *operator constraint* in the sense of Mordukhovich [84]. Results on upper subdifferential necessary optimality conditions of such a problem can also be found in [84, 89].

2.2.1 Problem with operator constraint

In this subsection, we consider the *optimization problem with operator constraint*

$$\min \{f(x) |\, x \in \Omega \cap \psi^{-1}(\Lambda)\}, \tag{2.40}$$

where $f : \mathbb{R}^n \to \mathbb{R}$ and $\psi : \mathbb{R}^n \to \mathbb{R}^m$ are locally Lipschitz continuous functions, and the sets $\Omega \subseteq \mathbb{R}^n$, $\Lambda \subseteq \mathbb{R}^m$ are closed. Obviously, since f is locally Lipschitz continuous, condition (2.38) holds at any reference point. Nonetheless, with the structure of the feasible set in (2.40), a constraint qualification (CQ) is needed in order to derive detailed KKT type dual optimality conditions in terms of problem data. The *basic CQ* defined at a feasible point $\bar{x}$ of problem (2.40) by

$$\left.\begin{array}{r} 0 \in \partial\langle u, \psi\rangle(\bar{x}) + N_\Omega(\bar{x}) \\ u \in N_\Lambda(\psi(\bar{x})) \end{array}\right\} \Longrightarrow u = 0 \tag{2.41}$$

is one of the most powerful CQs in optimization. It is in fact generic and stable in the sense that if it holds at a point $\bar{x}$, it also holds at all the points of some neighborhood of the same point. Moreover the satisfaction of CQ (2.41) implies that the corresponding set of Lagrange multipliers is nonempty and bounded. The basic CQ was introduced in [86] and studied for example in [82, 109, 108]. Interesting results on this CQ and related extensions can also be found in the papers [58, 70].

Remark 2.2.2 (basic CQ and MFCQ)**.** *If we set* $\Omega := \mathbb{R}^n$, $\Lambda := \mathbb{R}^l_- \times \{0_{m-l}\}$ *while the function* ψ *is a continuously differentiable function at* $\bar{x}$*, then the basic CQ coincides with the dual form of the well-known Mangasarian-Fromovitz constraint qualification (MFCQ) [56]. Hence, the basic CQ is a generalization of the MFCQ to the optimization problem with operator constraint.*

The following perturbation map $\Psi : \mathbb{R}^m \rightrightarrows \mathbb{R}^n$ of the operator constraint will play a central role in generating the necessary optimality conditions for the bilevel optimization problem in this thesis:

$$\Psi(u) := \{x \in \Omega |\ \psi(x) + u \in \Lambda\}. \tag{2.42}$$

The next lemma, which is a consequence of [87, Theorem 6.10] and [82, Corollary 4.2], shows that this set-valued mapping is Lipschitz-like at $(0,\bar{x}) \in \text{gph}\,\Psi$ provided the basic CQ is satisfied at $\bar{x}$.

Lemma 2.2.3 (variational properties of Ψ (2.42))**.** *Consider* $\bar{x} \in \Psi(0)$*, then we have*

$$D^*\Psi(0,\bar{x})(v) \subseteq \{u \in N_\Lambda(\psi(\bar{x}))| - v \in \partial\langle u, \psi\rangle(\bar{x}) + N_\Omega(\bar{x})\} \tag{2.43}$$

and the following chain of implications is also satisfied:

$$\text{basic CQ at } \bar{x} \Longrightarrow \Psi \text{ is Lipschitz-like at } (0,\bar{x}) \Longrightarrow \Psi \text{ is calm at } (0,\bar{x}). \tag{2.44}$$

Since we are only interested in deriving necessary optimality conditions for local optimal solutions, it is important to show that the Lipschitz-like and calmness properties defined in the previous section are locally preserved for the set-valued mapping Ψ (2.42).

Lemma 2.2.4 (local preservation of Lipschitz and calmness properties)**.** *Let* $\bar{x} \in \Psi(\bar{u})$ *and let* V *be a neighborhood of* $\bar{x}$*. If* Ψ (2.42) *is calm (resp. Lipschitz-like) at* $(\bar{u},\bar{x})$*, then the set-valued mapping*

$$\Psi_V(u) := \{x \in \Omega \cap V |\ \psi(x) + u \in \Lambda\} \tag{2.45}$$

is also calm (resp. Lipschitz-like) at $(\bar{u},\bar{x})$*.*

Proof. Let us set $\psi_u(x) := \psi(x) + u$. Then we have $\Psi(u) = \Omega \cap \psi_u^{-1}(\Lambda)$ and $\Psi_V(u) = V \cap \Omega \cap \psi_u^{-1}(\Lambda)$. Hence the result follows from the definition of calmness (2.32) (resp. Lipschitz-like property (2.29)), while noting that for A, B and $C \subseteq \mathbb{R}^n$, $A \subseteq B \Longrightarrow A \cap C \subseteq B \cap C$. □

We are now ready to state a KKT type optimality condition for problem (2.40) under the basic CQ. The technique utilized in the proof is inspired by [127, Lemma 3.1], in the framework of an optimization problem with a generalized equation constraint (OPEC). For the latter class of optimization problems, this approach has also been used in [89, 133]. The statement of this result and many proofs exist in the literature, cf. [43, 59, 89, 109, 108]; but we were unable to find any reference where the multiplier vector u is bounded, except as already mentioned, for an OPEC. Hence, the reason why we include the proof here.

Proposition 2.2.5 (necessary optimality conditions in an optimization problem with operator constraint)**.** *Let* $\bar{x}$ *be a local optimal solution of problem* (2.40) *and assume that* EITHER *the basic CQ* (2.41) *holds at* $\bar{x}$ OR *the set-valued mapping* Ψ (2.42) *is calm at* $(0,\bar{x})$*. Then, there exists* $\mu > 0$ *such that for any* $r \geq \mu$*, one can find* $u \in r\mathbb{B} \cap N_\Lambda(\psi(\bar{x}))$ *such that we have*

$$0 \in \partial f(\bar{x}) + \partial\langle u, \psi\rangle(\bar{x}) + N_\Omega(\bar{x}). \tag{2.46}$$

Proof. Since the basic CQ implies the calmness property (2.44), it suffices to provide the proof under the latter condition. Without loss of generality, let $\bar{x}$ be an optimal solution of problem (2.40) in a closed neighborhood V of $\bar{x}$. By Lemma 2.2.4, the mapping Ψ_V (2.45) is calm at $(0,\bar{x})$. Denote by

$\ell_V := \mathrm{cal}\,\Psi_V(0,\bar{x})$ its calmness modulus (2.34) and let ℓ_f be the Lipschitz modulus of the objective function f. We claim that for any $r \geq \mu := \ell_V \ell_f$, the point $(0,\bar{x})$ is a local optimal solution of the problem

$$\min_{u,x} \left\{ f_r(u,x) := f(x) + r\|u\| \,\middle|\, (u,x) \in \mathrm{gph}\,\Psi_V \right\}. \tag{2.47}$$

In fact, note that by the definition of the calmness of Ψ_V (2.45), there exist neighborhoods U_1 of 0 and V_1 of $\bar{x}$ such that $\forall u \in U_1$, $\forall x \in \Psi(u) \cap V_1$, we can find $v \in \Psi_V(0)$ with

$$\|x - v\| \leq \ell_V \|u\|. \tag{2.48}$$

In addition we have $f(\bar{x}) \leq f(v)$, since $v \in \Psi_V(0) = \Omega \cap V \cap \psi^{-1}(\Lambda)$. Now, let $r \geq \mu$ and consider a point $(u,x) \in \mathrm{gph}\,\Psi_V \cap (U_1 \times V_1)$, then it follows that

$$\begin{aligned} f_r(0,\bar{x}) = f(\bar{x}) &\leq [f(v) - f(x)] + f(x) \\ &\leq \ell_f \|x - v\| + f(x) \;\; \text{(cf. Lipschitz continuity of } f) \\ &\leq \ell_V \ell_f \|u\| + f(x) \leq f_r(u,x) \;\; \text{(cf. inequality (2.48))}. \end{aligned}$$

Applying Theorem 2.2.1 (ii) to problem (2.47), we have the following condition while taking into account the Lipschitz continuity of f:

$$(0,0) \in r\mathbb{B} \times \partial f(\bar{x}) + N_{\mathrm{gph}\,\Psi_V}(0,\bar{x}).$$

Hence, there exist $u \in r\mathbb{B}$ and $x \in \partial f(\bar{x})$ such that we have $(-u,-x) \in N_{\mathrm{gph}\,\Psi_V}(0,\bar{x})$. It follows from the definition of the coderivative (2.24) and the inclusion (2.43), that we have the inclusions $-u \in N_\Lambda(\psi(\bar{x}))$ and $-x \in \partial\langle -u, \psi\rangle(\bar{x}) + N_\Omega(\bar{x})$, which concludes the proof. □

Remark 2.2.6 (link with the approach by Ye and Ye [127]). *The optimality condition* (2.46) *also follows from [127, Theorem 3.1], where one has to set* $\Phi(x) := -\psi(x) + \Lambda$. *However, the approach in [127] does not allow us to detect the fact that* u *also belongs to* $N_\Lambda(\psi(\bar{x}))$, *which is an important component of Proposition 2.2.5, regarding the structure of problem* (2.40). *Furthermore, as it will be clear in the subsequent parts of the thesis, the inclusion* $u \in N_\Lambda(\psi(\bar{x}))$ *plays an important role in the applications.*

A major consequence of Proposition 2.2.5 on a standard optimization problem with inequality and equality constraints is as follows: Let the point $\bar{x}$ be a local optimal solution of the problem

$$\min \{ f(x) |\, g(x) \leq 0,\; h(x) = 0 \},$$

with $f : \mathbb{R}^n \to \mathbb{R}$, $g : \mathbb{R}^n \to \mathbb{R}^p$ and $h : \mathbb{R}^n \to \mathbb{R}^q$ locally Lipschitz continuous functions around $\bar{x}$, and assume that the set-valued mapping $\Psi(u,v) := \{x |\, g(x) \leq u,\; h(x) = v\}$ is calm at $(0,0,\bar{x})$ (resp. there is no nonzero vector (u,v) such that $0 \in \sum_{i=1}^p u_i \partial g_i(\bar{x}) + \partial\langle v,h\rangle(\bar{x})$ and $u_i \geq 0$, $u_i g_i(\bar{x}) = 0$, $i = 1,\ldots p$). Then, there exists (u,v) such that we have:

$$\begin{aligned} 0 \in \partial f(\bar{x}) + \sum_{i=1}^p u_i \partial g_i(\bar{x}) + \partial\langle v,h\rangle(\bar{x}), \\ \forall i = 1,\ldots,p,\; u_i \geq 0,\; u_i g_i(\bar{x}) = 0, \\ \text{where } \|(u,v)\| \leq \|\ell_f\| \mathrm{cal}\,\Psi(0,\bar{x}) \text{ (resp. } \|(u,v)\| \leq \|\ell_f\| \mathrm{lip}\,\Psi(0,\bar{x})). \end{aligned}$$

Here ℓ_f denotes the Lipschitz modulus of the objective function f, while $\mathrm{cal}\,\Psi(0,\bar{x})$ and $\mathrm{lip}\,\Psi(0,\bar{x})$ are given by (2.31) and (2.34), respectively. For the expression of $\mathrm{cal}\,\Psi(0,\bar{x})$ in (2.34), we need an additional assumption, say the convexity of g and the linearity of h. Formulas for ℓ_f can be found in [109, Chapter 9]. Various results have been suggested in the literature on the nonvacuity and boundedness of the set of Lagrange multipliers, see e.g. [42, 54, 70, 95, 104]. However many of them do not provide an explicit bound for these multipliers.

Now let us recall that a second approach to obtain the result in Proposition 2.2.5 is to directly apply Theorem 2.2.1 (ii), which imply that for a local optimal solution $\bar{x}$ of problem (2.40), we have

$$0 \in \partial f(\bar{x}) + N_{\Omega \cap \psi^{-1}(\Lambda)}(\bar{x}). \tag{2.49}$$

However the CQs in Proposition 2.2.5 are also needed to compute the normal cone in (2.49) as to be sated in the following well-known result, which can be found e.g. in [89, 109] under the basic CQ (2.41), while under the calmness it follows from [59, Theorem 4.1].

Theorem 2.2.7 (normal cone to an operator constraint feasible set)**.** *Let $\bar{x} \in \Omega \cap \psi^{-1}(\Lambda)$ and assume that* EITHER *the basic CQ* (2.41) *holds at $\bar{x}$* OR *the set-valued mapping Ψ* (2.42) *is calm at $(0, \bar{x})$. Then the following inclusion is satisfied:*

$$N_{\Omega \cap \psi^{-1}(\Lambda)}(\bar{x}) \subseteq \bigcup \big\{ \partial \langle u, \psi \rangle + N_{\Omega}(\bar{x}) \big|\, u \in N_{\Lambda}(\psi(\bar{x})) \big\}. \tag{2.50}$$

Equality holds in (2.50)*, provided the set Λ is regular at $\psi(\bar{x})$, in the sense of* (2.23).

It is not clear to the author how to directly detect the bound on the multipliers derived via this approach. Nevertheless, it still appears that these multipliers will be bounded under the basic CQ, see e.g. [54, 70, 95, 104].

3 Classical optimistic bilevel programming problem

3.1 Reformulations and stationarity concepts

3.1.1 One-level reformulations

Consider problem (P) (see (1.9)), which has been mostly investigated under the name of "bilevel optimization/programming problem" and that we call here *classical* bilevel optimization/programming problem or simply *classical bilevel program*. If the optimal solutions of the lower-level problem (1.3) are not uniquely determined for some parameters $x \in X$, then many reformulations of (P) into a one-level (i.e. a standard) optimization problem have been considered in the literature, which are essentially based on the possible representations of the set of optimal solutions of problem (1.3). Let us start with the lower-level value function (LLVF) representation of S (1.2)

$$S(x) = \{y \in K(x) | \, f(x,y) \leq \varphi(x)\}, \tag{3.1}$$

where φ denotes the optimal value function of the lower-level problem (1.3)

$$\varphi(x) := \min_{y} \, \{f(x,y) | \, y \in K(x)\}, \tag{3.2}$$

then (P) can take the following so-called (lower-level) optimal value function reformulation, that we denote as *LLVF reformulation*, for short:

$$\min_{x,y} \, \{F(x,y) | \, x \in X, \, y \in K(x), \, f(x,y) \leq \varphi(x)\}. \tag{3.3}$$

In order for the value function φ (3.2) to be well-defined, it is assumed from here on that $S(x) \neq \emptyset$ for all $x \in X$. Problem (3.3) and (P) are completely equivalent as shown in the following trivial result:

Theorem 3.1.1 (link between (P) and its LLVF reformulation (3.3))**.** *The point $(\bar{x}, \bar{y})$ is a local (resp. global) optimal solution of (P) if and only if $(\bar{x}, \bar{y})$ is a local (resp. global) optimal solution of (3.3).*

The LLVF reformulation of (P) was introduced by Outrata [98] while constructing a numerical approach to solve the problem. Work on necessary optimality conditions for (P) via (3.3) was pioneered by Ye and Zhu [128] by means of the now well-known concept of partial calmness that will be discussed in the next section.

Next we assume the parametric problem (1.3) to be convex in the sense that the functions $f(x,.)$ and $g(x,.)$ are convex for all $x \in X$, then S (1.2) takes the *generalized equation* form

$$S(x) = \{y \in \mathbb{R}^m | \, 0 \in \nabla_y f(x,y) + N_{K(x)}(y)\}, \tag{3.4}$$

where $N_{K(x)}(y)$ denotes the normal cone in the sense of convex analysis (2.20), to $K(x)$ at y, provided $y \in K(x)$, whereas $N_{K(x)}(y) := \emptyset$, if $y \notin K(x)$. Hence, (P) can be interpreted as an optimization problem with generalized equation constraint (OPEC):

$$\min_{x,y} \, \{F(x,y) | \, x \in X, \;\, 0 \in \nabla_y f(x,y) + N_{K(x)}(y)\}. \tag{3.5}$$

The link between (P) and its OPEC/primal KKT reformulation (3.5) is also obvious:

Theorem 3.1.2 (link between (P) and its OPEC reformulation (3.5))**.** *The lower-level problem* (1.3) *is assumed to be convex. Then a point* $(\bar{x},\bar{y})$ *is a local (resp. global) optimal solution of (P) if and only if* $(\bar{x},\bar{y})$ *is a local (resp. global) optimal solution of* (3.5).

Problem (3.5) has been studied for example in [89, 93, 127], usually under the name of "optimization problem with variational inequality constraint (OPVIC)". The reason for the latter appellation is that under the convexity of $K(x)$ for $x \in X$, one has the following well-known expression for the normal cone in the sense of convex analysis (see (2.20))

$$N_{K(x)}(y) := \{v \in \mathbb{R}^m | \langle v, u-y\rangle \leq 0, \ \forall u \in K(x)\}.$$

If we insert this formula in (3.5), the aforementioned vocabulary is clearly justified.

The last reformulation of (P) that will be considered here is the so-called *KKT reformulation*. In addition to the convexity assumption, we need a CQ to proceed. The Slater CQ will be said to hold at a parameter $\bar{x} \in X$ if we have

$$K^{<}(\bar{x}) := \{y | \ g_i(\bar{x},y) < 0, \ i = 1,\dots,p\} \neq \emptyset. \tag{3.6}$$

Under this CQ, the normal cone $N_{K(x)}(y)$ can be represented by

$$N_{K(x)}(y) = \Big\{ \sum_{i=1}^{p} u_i \nabla_y g_i(x,y) \Big| \ u \geq 0, u^\top g(x,y) = 0 \Big\}. \tag{3.7}$$

Inserting the latter formula in the OPEC reformulation (3.5), we get the following problem

$$\min_{x,y,u} \{F(x,y) | x \in X, \mathscr{L}(x,y,u) = 0, \ g(x,y) \leq 0, \ u \geq 0, \ u^\top g(x,y) = 0\}, \tag{3.8}$$

often labeled as Karush-Kuhn-Tucker (KKT) reformulation, given that the above formula for the normal cone (3.7) induces the Karush-Kuhn-Tucker type necessary optimality conditions for the lower-level optimization problem (1.3). In (3.8), $\mathscr{L}$ stands for the gradient of the Lagrange function of the lower-level problem:

$$\mathscr{L}(x,y,u) := \nabla_y f(x,y) + \sum_{i=1}^{p} u_i \nabla_y g_i(x,y). \tag{3.9}$$

Clearly, problem (3.8) is also a mathematical programming problem with complementarity constraints (MPCC), considering the embedded complementarity problem:

$$g(x,y) \leq 0, \ u \geq 0, \ u^\top g(x,y) = 0. \tag{3.10}$$

We have the following theorem, resulting from a recent study by Dempe and Dutta [24], on the relationship between (P) and its KKT reformulation. To simplify the presentation of the result, denote the set of Lagrange multipliers of the lower-level problem by

$$\Lambda(x,y) := \{u | \ u \geq 0, \ u^\top g(x,y) = 0, \ \mathscr{L}(x,y,u) = 0\}. \tag{3.11}$$

Theorem 3.1.3 (link between (P) and its KKT reformulation (3.8))**.** *Let the lower-level problem* (1.3) *be convex. Then the following assertions hold:*

(i) *Let* $(\bar{x},\bar{y})$ *be a global* (*resp. local*) *optimal solution of problem (P) and assume that the Slater CQ* (3.6) *is satisfied at* $\bar{x}$. *Then, for each* $\bar{u} \in \Lambda(\bar{x},\bar{y})$, *the point* $(\bar{x},\bar{y},\bar{u})$ *is a global* (*resp. local*) *optimal solution of problem* (3.8).

(ii) *Let the Slater CQ be satisfied at all* $x \in X$ (*resp. at* $\bar{x}$). *Assume that* $(\bar{x},\bar{y},\bar{u})$ *is a global optimal solution* $\big($*resp. local optimal solution for all* $\bar{u} \in \Lambda(\bar{x},\bar{y})\big)$ *of problem* (3.8). *Then,* $(\bar{x},\bar{y})$ *is a global (resp. local) optimal solution of problem (P).*

It is clear from this result that for global solutions, problem (3.8) is equivalent to the initial problem (P) in the sense that, $(\bar{x},\bar{y})$ is a global optimal solution of (P) if and only if there exists a vector $\bar{u} \in \Lambda(\bar{x},\bar{y})$ such that $(\bar{x},\bar{y},\bar{u})$ is a global optimal solution of problem (3.8). However, for local solutions, the requirement that $(\bar{x},\bar{y},\bar{u})$ be a local optimal solution of (3.8) for all $\bar{u} \in \Lambda(\bar{x},\bar{y})$, before one can be sure that $(\bar{x},\bar{y})$ is a local optimal solution for (P), is too strong. Not only it is not realistic at a computational level, but one can easily construct examples of problems where a local optimal solution of problem (3.8) is not a local optimal solution of (P), cf. [24]. Hence, it may be fair to say that computing the normal cone in (3.1.2) via (3.7) destroys the nice link, stated in Theorem 3.1.2, between the bilevel program (P) and its primal KKT reformulation (3.5).

It is important to mention that the CQ (Slater) can not be dropped in any segment of the above theorem [24]. This is the case for the convexity of the lower-level problem as well; otherwise a solution of the bilevel program need not even be a stationary point of problem (3.8), see Mirrlees [81]. To circumvent this convexity assumption and some possibly unwanted behavior that may occur using the optimal value reformulation, Ye and Zhu [129] recently suggested a combination of the KKT and the optimal value reformulations in order to obtain optimality conditions for the bilevel programming problem. Concretely, it is assumed in [129], that the KKT conditions of the lower level problem be satisfied without necessarily requiring the convexity of the lower level problem and hence the following one level optimization problem is considered:

$$\begin{aligned}\min_{x,y,u} \{F(x,y)| \quad & x \in X,\ f(x,y) \le \varphi(x),\\ & \mathscr{L}(x,y,u) = 0,\ g(x,y) \le 0,\ u \ge 0,\ u^\top g(x,y) = 0\}.\end{aligned}$$

An equivalence between this problem and the initial one (P) was established in [129] around a predefined neighborhood of the considered point. For optimality conditions, techniques known for MPCCs were applied under calmness and partial calmness concepts tailored for the problem. We will not insist on this approach, since we are interested but in considering the LLVF, OPEC and KKT reformulations separately, and looking at possible links between them. The interested reader is referred to the aforementioned paper for further details. Also see [103] for other approaches to derive optimality conditions for the problem.

3.1.2 Stationarity concepts

Taking into account the reformulations of (P) in the previous subsection, we now introduce the main stationarity concepts that will be derived in the next section. The concepts that we introduce here will also provide a general pattern of the stationarity conditions that will be discussed in Chapter 5 for problems (P_o) and (P_p). Basically, for the LLVF reformulation (3.3), we will have the “KM” and “KN” type stationarity conditions, reflecting the difference between the **K**KT-type optimality conditions obtained via the inner semico**M**pactness and inner semico**N**tinuity, respectively, of the optimal solution map S (1.2) for the lower-level problem. Moreover, for more precision, the prefix "P" will be added to differentiate between the stationarity concepts of problem (P) and those of the other problems, namely (P_o) and (P_p).

Definition 3.1.4 (KM-stationarity for (P))**.** $(\bar{x},\bar{y})$ *is* P-KM-STATIONARY *if there exist* $(\alpha,\beta) \in \mathbb{R}^{k+p}$,

$\lambda \in \mathbb{R}_+$, $(\mu_s, \upsilon_s) \in \mathbb{R}^{k+1}$ and $y_s \in S(\bar{x})$ with $s = 1,\dots,n+1$ such that the following conditions hold:

$$\nabla_x F(\bar{x},\bar{y}) + \sum_{j=1}^{k} \alpha_j \nabla G_j(\bar{x}) + \lambda\Big(\nabla_x f(\bar{x},\bar{y}) - \sum_{s=1}^{n+1} \upsilon_s \nabla_x f(\bar{x},y_s)\Big) + \sum_{i=1}^{p} \beta_i \nabla_x g_i(\bar{x},\bar{y})$$
$$-\lambda \sum_{s=1}^{n+1} \upsilon_s \sum_{i=1}^{p} \mu_{is} \nabla_x g(\bar{x},y_s) = 0, \tag{3.12}$$
$$\nabla_y F(\bar{x},\bar{y}) + \lambda \nabla_y f(\bar{x},\bar{y}) + \sum_{i=1}^{p} \beta_i \nabla_y g_i(\bar{x},\bar{y}) = 0, \tag{3.13}$$
$$\forall s = 1,\dots,n+1,\ \nabla_y f(\bar{x},y_s) + \sum_{i=1}^{p} \mu_{is} \nabla_y g(\bar{x},y_s) = 0, \tag{3.14}$$
$$\forall s = 1,\dots,n+1,\ i = 1,\dots,p,\ \mu_{is} \geq 0,\ \mu_{is} g_i(\bar{x},y_s) = 0, \tag{3.15}$$
$$\forall j = 1,\dots,k,\ \alpha_j \geq 0,\ \alpha_j G_j(\bar{x}) = 0, \tag{3.16}$$
$$\forall i = 1,\dots,p,\ \beta_i \geq 0,\ \beta_i g_i(\bar{x},\bar{y}) = 0, \tag{3.17}$$
$$\forall s = 1,\dots,n+1,\ \upsilon_s \geq 0,\ \sum_{s=1}^{n+1} \upsilon_s = 1. \tag{3.18}$$

Relationships (3.12)–(3.18) *are called the* KM-STATIONARITY CONDITIONS.

Definition 3.1.5 (KN-stationarity for (P))**.** *$(\bar{x},\bar{y})$ is* P-KN-STATIONARY *if there exist $(\alpha,\beta,\mu) \in \mathbb{R}^{k+2p}$ and $\lambda \in \mathbb{R}_+$ such that relationships* (3.13) *and* (3.16)–(3.17) *together with the following ones hold:*

$$\nabla_x F(\bar{x},\bar{y}) + \sum_{j=1}^{k} \alpha_j \nabla G_j(\bar{x}) + \sum_{i=1}^{p} (\beta_i - \lambda \mu_i) \nabla_x g_i(\bar{x},\bar{y}) = 0, \tag{3.19}$$
$$\nabla_y f(\bar{x},\bar{y}) + \sum_{i=1}^{p} \mu_i \nabla_y g_i(\bar{x},\bar{y}) = 0, \tag{3.20}$$
$$\forall i = 1,\dots,p,\ \mu_i \geq 0,\ \mu_i g_i(\bar{x},\bar{y}) = 0. \tag{3.21}$$

All the relationships (3.13)*,* (3.16)–(3.17)*, and* (3.19)–(3.21) *considered together are called the* KN-STATIONARITY CONDITIONS.

In order to introduce the stationarity concepts related to the OPEC and KKT reformulations, we first consider the following well-known partition of the indices of the functions involved in the complementarity system (3.10)

$$\begin{array}{lllll} \eta & := & \eta(\bar{x},\bar{y},\bar{u}) & := & \{i|\ \ \bar{u}_i = 0,\ g_i(\bar{x},\bar{y}) < 0\}, \\ \theta & := & \theta(\bar{x},\bar{y},\bar{u}) & := & \{i|\ \ \bar{u}_i = 0,\ g_i(\bar{x},\bar{y}) = 0\}, \\ \nu & := & \nu(\bar{x},\bar{y},\bar{u}) & := & \{i|\ \ \bar{u}_i > 0,\ g_i(\bar{x},\bar{y}) = 0\}, \end{array} \tag{3.22}$$

where the middle set θ in (3.22) is known as the *biactive* or *degenerate index set*. The difference between the following stationarity concepts depends on the structure of the components of the multipliers corresponding to the biactive set θ. Namely, we will consider *M*(ordukhovich), *C*(larke) and *S*(trong) stationarity conditions, tailored to problem (P). As far as the OPEC and KKT reformulations are concerned, many other surrogates of well-known types could also be defined for (P). In this thesis we will focus our attention only on these three (M, C and S) since they are the most important ones.

Definition 3.1.6 (M-stationarity concepts for (P))**.** *$(\bar{x},\bar{y})$ is* SP-M-STATIONARY (*resp.* P-M-STATIONARY) *if for every $\bar{u} \in \Lambda(\bar{x},\bar{y})$* (*resp. for some $\bar{u} \in \Lambda(\bar{x},\bar{y})$*) *we can find a triple $(\alpha,\beta,\gamma) \in \mathbb{R}^{k+p+m}$ such that re-*

lationship (3.16) *together with the following conditions are satisfied:*

$$\nabla_x F(\bar{x},\bar{y}) + \sum_{j=1}^{k} \alpha_j \nabla G_j(\bar{x}) + \sum_{i=1}^{p} \beta_i \nabla_x g_i(\bar{x},\bar{y}) + \sum_{l=1}^{m} \gamma_l \nabla_x \mathscr{L}_l(\bar{x},\bar{y},\bar{u}) = 0, \tag{3.23}$$

$$\nabla_y F(\bar{x},\bar{y}) + \sum_{i=1}^{p} \beta_i \nabla_y g_i(\bar{x},\bar{y}) + \sum_{l=1}^{m} \gamma_l \nabla_y \mathscr{L}_l(\bar{x},\bar{y},\bar{u}) = 0, \tag{3.24}$$

$$\nabla_y g_\nu(\bar{x},\bar{y})\gamma = 0,\ \beta_\eta = 0, \tag{3.25}$$

$$\forall i \in \theta,\ \big(\beta_i > 0 \wedge \sum_{l=1}^{m} \gamma_l \nabla_{y_l} g_i(\bar{x},\bar{y}) > 0\big) \vee \beta_i \sum_{l=1}^{m} \gamma_l \nabla_{y_l} g_i(\bar{x},\bar{y}) = 0. \tag{3.26}$$

All the relationships (3.16) *and* (3.23)–(3.26) *considered together are called the* M-STATIONARITY CONDITIONS.

Here the term "SP-M-STATIONARY" stands for STRONG P-M-STATIONARY. In what follows, "SP-C-STATIONARY" and "SP-S-STATIONARY" will define similar concepts. Also recall that since

$$\nabla_y g(\bar{x},\bar{y})\gamma = \Big[\sum_{l=1}^{m} \gamma_l \nabla_{y_l} g_1(\bar{x},\bar{y}), \ldots, \sum_{l=1}^{m} \gamma_l \nabla_{y_l} g_p(\bar{x},\bar{y})\Big]^\top, \tag{3.27}$$

the vector $\nabla_y g_\nu(\bar{x},\bar{y})\gamma$ denotes the components of the vector in the right-hand-side for which $i \in \nu$. Furthermore, note that with the expression of the lower-level Lagrangian in (3.9), we have

$$\mathscr{L}_l(x,y,u) := \nabla_{y_l} f(x,y) + \sum_{i=1}^{p} u_i \nabla_{y_l} g_i(x,y) \ \text{ for } l = 1,\ldots,m. \tag{3.28}$$

Definition 3.1.7 (C-stationarity concepts for (P))**.** $(\bar{x},\bar{y})$ *is* SP-C-STATIONARY (*resp.* P-C-STATIONARY) *if for every* $\bar{u} \in \Lambda(\bar{x},\bar{y})$ (*resp. for some* $\bar{u} \in \Lambda(\bar{x},\bar{y})$) *we can find a triple* $(\alpha,\beta,\gamma) \in \mathbb{R}^{k+p+m}$ *such that we have relationships* (3.16) *and* (3.23)–(3.25) *to be satisfied together with the following condition:*

$$\forall i \in \theta,\ \beta_i \sum_{l=1}^{m} \gamma_l \nabla_{y_l} g_i(\bar{x},\bar{y}) \geq 0. \tag{3.29}$$

All the relationships (3.16), (3.23)–(3.25) *and* (3.29) *considered together are called the* C-STATIONARITY CONDITIONS.

Definition 3.1.8 (S-stationarity concepts for (P))**.** $(\bar{x},\bar{y})$ *is* SP-S-STATIONARY (*resp.* P-S-STATIONARY) *if for every* $\bar{u} \in \Lambda(\bar{x},\bar{y})$ (*resp. for some* $\bar{u} \in \Lambda(\bar{x},\bar{y})$) *we can find a triple* $(\alpha,\beta,\gamma) \in \mathbb{R}^{k+p+m}$ *such that we have relationships* (3.16) *and* (3.23)–(3.25) *to be satisfied together with the following condition:*

$$\forall i \in \theta,\ \beta_i \geq 0 \wedge \sum_{l=1}^{m} \gamma_l \nabla_{y_l} g_i(\bar{x},\bar{y}) \geq 0. \tag{3.30}$$

All the relationships (3.16), (3.23)–(3.25) *and* (3.30) *considered together are called the* S-STATIONARITY CONDITIONS.

The above M, C and S-stationarity concepts are specifically tailored to (P). Usually, in the MPCC theory, the term $\nabla_y g_i(\bar{x},\bar{y})\gamma$ in (3.26), (3.29) and (3.30), is in fact a multiplier. These representations of the necessary optimality conditions of (P) obtained via the OPEC and KKT reformulations was introduced in [34]. Additionally, the quantifier "for all" attached to $\bar{u}$ is related to the nature of the link between (P) and its KKT reformulation (3.8) as stated in Theorem 3.1.3. Note the adjustment made here on the position of the quantifier "for all" as compare to the initial Definitions 1.3 and 1.4 in [34].

Theorem 3.1.9 (relationships between the stationarity concepts)**.** *For a given feasible point* $(\bar{x},\bar{y})$ *of problem (P), we have the following chain of implications:*

$$\begin{array}{ccccccc} \text{P-KM-stationary} & & \text{SP-S-stationary} & \Longrightarrow & \text{SP-M-stationary} & \Longrightarrow & \text{SP-C-stationary} \\ (1)\Updownarrow & & \Downarrow & & \Downarrow & & \Downarrow \\ \text{P-KN-stationary}\,(\lambda \textit{ free}) & \overset{(2)}{\Longleftarrow} & \text{P-S-stationary} & \Longrightarrow & \text{P-M-stationary} & \Longrightarrow & \text{P-C-stationary} \end{array} \quad (3.31)$$

where the assumptions (1) *and* (2) *are given by:*

$$\begin{cases} (1): & S(\bar{x}) = \{\bar{y}\},\, \Lambda(\bar{x},\bar{y}) = \{\mu\}, \\ (2): & \sum_{l=1}^{m} \gamma_l \nabla_{x,y}\mathscr{L}_l(\bar{x},\bar{y},\bar{u}) = 0. \end{cases}$$

Proof. That SP-S-stationary $\Longrightarrow$ P-S-stationary is trivial. It is the same for the M- and C-stationarity. The implications "P-S-stationary $\Longrightarrow$ P-M-stationary $\Longrightarrow$ P-C-stationary" are also obvious and well-known; and the stronger ones "SP-S-stationary $\Longrightarrow$ SP-M-stationary $\Longrightarrow$ SP-C-stationary" follow analogously.

For the equivalence between the KM- and KN-stationarity in (3.31), note that saying that a vector $(\bar{x}, y_s\alpha, \beta, \lambda, \mu_s, \upsilon_s)$ with $s = 1,\dots,n+1$ satisfies the KM-stationarity conditions (3.12)–(3.18) is equivalent to say that $\gamma_s \in \Lambda(\bar{x}, y_s)$, $y_s \in S(\bar{x})$ with $s = 1,\dots,n+1$ such that (3.12)–(3.13) and (3.16)–(3.18). Hence, with $S(\bar{x}) = \{\bar{y}\}$ and $\Lambda(\bar{x},\bar{y}) = \{\mu\}$, we have $s \in \{1\}$ and $\upsilon_s = 1$. Thus the resulting conditions are identical to the KN-stationarity conditions in the sense of Definition 3.1.5.

It therefore remains to show that we have

$$\text{P-KN-stationary (with } \lambda \text{ free)} \overset{(2)}{\Longleftarrow} \text{P-S-stationary.}$$

For this first note that if we insert the value of $\nabla_y f(\bar{x},\bar{y}) = -\sum_{i=1}^{p} \mu_i \nabla_y g(\bar{x},\bar{y})$ from equation (3.20) in (3.13), we get the relationship

$$\nabla_y F(\bar{x},\bar{y}) + \sum_{i=1}^{p} (\beta_i - \lambda\mu_i)\nabla_y g(\bar{x},\bar{y}) = 0. \quad (3.32)$$

Hence, it now suffices to show that $(\bar{x},\bar{y})$ is P-S-stationary if and only if there exists $(\bar{u},\alpha,\beta,\gamma,\lambda)$ such that the vector $(\bar{x},\bar{y},\bar{u},\alpha,\beta,\gamma,\lambda)$ satisfies the conditions (3.16)–(3.17) together with the following ones:

$$\nabla_x F(\bar{x},\bar{y}) + \sum_{j=1}^{k} \alpha_j \nabla G_j(\bar{x}) + \sum_{i=1}^{p} (\beta_i - \lambda\bar{u}_i)\nabla_x g_i(\bar{x},\bar{y}) + \sum_{l=1}^{m} \gamma_l \nabla_x \mathscr{L}_l(\bar{x},\bar{y},\bar{u}) = 0, \quad (3.33)$$

$$\nabla_y F(\bar{x},\bar{y}) + \sum_{i=1}^{p} (\beta_i - \lambda\bar{u}_i)\nabla_y g_i(\bar{x},\bar{y}) + \sum_{l=1}^{m} \gamma_l \nabla_y \mathscr{L}_l(\bar{x},\bar{y},\bar{u}) = 0, \quad (3.34)$$

$$\nabla_y f(\bar{x},\bar{y}) + \sum_{i}^{p} \bar{u}_i \nabla_y g_i(\bar{x},\bar{y}) = 0, \quad (3.35)$$

$$\forall i = 1,\dots,p,\ \bar{u}_i \geq 0,\ \bar{u}_i g_i(\bar{x},\bar{y}) = 0, \quad (3.36)$$

$$\forall i = 1,\dots,p,\ \sum_{l=1}^{m} \gamma_l \nabla_{y_l} g_i(\bar{x},\bar{y}) - \lambda g_i(\bar{x},\bar{y}) \geq 0,\ \ \bar{u}_i \sum_{l=1}^{m} \gamma_l \nabla_{y_l} g_i(\bar{x},\bar{y}) = 0. \quad (3.37)$$

Let $(\bar{u},\alpha,\beta,\gamma,\lambda)$ be such that $(\bar{x},\bar{y},\bar{u},\alpha,\beta,\gamma,\lambda)$ satisfies (3.16)–(3.17) and (3.33)–(3.37). We then set $\alpha^* := \alpha$, $\beta^* := \beta - \lambda\bar{u}$ and $\gamma^* := \gamma$. By exploiting the definition of η (3.22) and the complementarity system (3.17), we have $\beta^*_\eta = 0$. Analogously, we have $\nabla_y g_\nu(\bar{x},\bar{y})\gamma = 0$ using the definition of ν (3.22) and the equality $\bar{u}_i \sum_{l=1}^{m} \gamma_l \nabla_{y_l} g_i(\bar{x},\bar{y}) = 0$ from (3.37). Finally, by combining the complementarity system (3.17), the definition of θ (3.22), and inequality $\nabla_y g(\bar{x},\bar{y})\gamma - \lambda g(\bar{x},\bar{y}) \geq 0$ from (3.37), it follows that $\beta^*_i \geq 0$ and $\nabla_y g_i(\bar{x},\bar{y})\gamma \geq 0$ for $i \in \theta$. Hence, the point $(\bar{x},\bar{y},\bar{u},\alpha^*,\beta^*,\gamma^*)$ satisfies the S-stationarity conditions (3.16), (3.23)–(3.25) and (3.30). Thus $(\bar{x},\bar{y})$ is P-S-stationary while taking into account that the conditions (3.35)–(3.36) represent inclusion $\bar{u} \in \Lambda(\bar{x},\bar{y})$.

Conversely, let $(\bar{x},\bar{y})$ be P-S-stationary. Then there exists $(\bar{u},\alpha,\beta,\gamma)$ such that conditions (3.16), (3.23)–(3.25) together with (3.30) and (3.35)–(3.36) are satisfied. Now observe that the set of conditions (3.23)–(3.25) and (3.30) can equivalently be represented by the the following system where the new multiplier vector $\zeta \in \mathbb{R}^p$:

$$\begin{bmatrix} \nabla_x F(\bar{x},\bar{y}) \\ \nabla_y F(\bar{x},\bar{y}) \\ 0 \end{bmatrix} + \sum_{j=1}^{k} \alpha_j \begin{bmatrix} \nabla G_j(\bar{x}) \\ 0 \\ 0 \end{bmatrix} + \sum_{l=1}^{m} \gamma_l \begin{bmatrix} \nabla_x \mathscr{L}_l(\bar{x},\bar{y},\bar{u}) \\ \nabla_y \mathscr{L}_l(\bar{x},\bar{y},\bar{u}) \\ \nabla_{y_l} g(\bar{x},\bar{y}) \end{bmatrix} - \sum_{i=1}^{p} \zeta_i \begin{bmatrix} 0 \\ 0 \\ e^i \end{bmatrix}$$
$$+ \sum_{i=1}^{p} \beta_i \begin{bmatrix} \nabla_x g_i(\bar{x},\bar{y}) \\ \nabla_y g_i(\bar{x},\bar{y}) \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix},$$
$$\zeta_\nu = 0,\ \beta_\eta = 0,$$
$$\forall i \in \theta,\ \beta_i \geq 0 \wedge \zeta_i \geq 0.$$

Here, the vector e^i denotes the i-th vector of the canonical basis of $\mathbb{R}^p$. It follows from [49, Proof of Proposition 4.2] that by setting $\alpha^* := \alpha$, $\gamma^* := \gamma$, $\zeta^*_{\theta\cup\nu} := \zeta_{\theta\cup\nu}$, $\beta^*_{\zeta\cup\theta} := \beta_{\zeta\cup\theta}$, $\zeta^*_\eta := \zeta_\eta - \lambda^* g_\eta(\bar{x},\bar{y})$ and $\beta^*_\nu := \beta_\nu + \lambda^* \bar{u}_\nu$, with $\lambda^* := \max\{\max_{i\in\eta}\{\frac{\zeta_i}{g_i(\bar{x},\bar{y})}\}, \max_{i\in\nu}\{\frac{-\beta_i}{\bar{u}_i}\}\}$, then $(\bar{x},\bar{y},\bar{u},\alpha^*,\beta^*,\gamma^*,\zeta^*,\lambda^*)$ satisfies (3.16)–(3.17), (3.35)–(3.36) together with the following conditions:

$$\begin{bmatrix} \nabla_x F(\bar{x},\bar{y}) \\ \nabla_y F(\bar{x},\bar{y}) \\ 0 \end{bmatrix} + \sum_{j=1}^{k} \alpha^*_j \begin{bmatrix} \nabla G_j(\bar{x}) \\ 0 \\ 0 \end{bmatrix} + \sum_{l=1}^{m} \gamma^*_l \begin{bmatrix} \nabla_x \mathscr{L}_l(\bar{x},\bar{y},\bar{u}) \\ \nabla_y \mathscr{L}_l(\bar{x},\bar{y},\bar{u}) \\ \nabla_{y_l} g(\bar{x},\bar{y}) \end{bmatrix} - \sum_{i=1}^{p} \zeta^*_i \begin{bmatrix} 0 \\ 0 \\ e^i \end{bmatrix}$$
$$+ \sum_{i=1}^{p} \beta^*_i \begin{bmatrix} \nabla_x g_i(\bar{x},\bar{y}) \\ \nabla_y g_i(\bar{x},\bar{y}) \\ 0 \end{bmatrix} - \lambda^* \begin{bmatrix} \sum_{i=1}^{p} \bar{u}_i \nabla_x g_i(\bar{x},\bar{y}) \\ \sum_{i=1}^{p} \bar{u}_i \nabla_y g_i(\bar{x},\bar{y}) \\ g(\bar{x},\bar{y}) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix},$$
$$\forall i = 1,\dots,p,\ \zeta^*_i \geq 0,\ \zeta^*_i \bar{u}_i = 0.$$

By observing from the latter system that we have $\zeta^* = \nabla_y g(\bar{x},\bar{y})\gamma^* - \lambda^* g(\bar{x},\bar{y})$, it clearly follows that the vector $(\bar{x},\bar{y},\bar{u},\alpha^*,\beta^*,\gamma^*,\lambda^*)$ satisfies the conditions (3.33)–(3.37). Now set $\sum_{l=1}^{m} \gamma_l \nabla_x \mathscr{L}_l(\bar{x},\bar{y},\bar{u}) = 0$ in (3.33) and $\sum_{l=1}^{m} \gamma_l \nabla_y \mathscr{L}_l(\bar{x},\bar{y},\bar{u}) = 0$ in (3.34), then the resulting conditions together with (3.35)–(3.36) and (3.16)–(3.17) reduce to the KN-stationarity conditions (with $\mu := \bar{u}$) while replacing equation (3.13) by the equivalent condition (3.32). This completes the proof of the theorem. □

Note that the term "λ free" in (3.31) means that in the framework of Theorem 3.1.9 the multiplier λ must not be nonnegative as it is the case in Definition 3.1.5. Assumption (1) in Theorem 3.1.9 is made of two components: $S(\bar{x}) = \{\bar{y}\}$ is satisfied if a certain second order sufficient condition is satisfied for the lower-level problem (1.3), cf. Remark 3.2.15 for a related discussion. $\Lambda(\bar{x},\bar{y}) = \{\mu\}$ is obtained, provided the linear independence constraint qualification (LICQ) holds for the same problem (1.3). As for assumption (2), it is automatical if the functions f and g defining the lower-level problem (1.3) take the form $f(x,y) := a(x) + b^\top y$ and $g(x,y) := C(x) + D^\top y$, respectively. Here $a : \mathbb{R}^n \to \mathbb{R}$ and $C : \mathbb{R}^n \to \mathbb{R}^p$, while $b \in \mathbb{R}^m$ and $D \in \mathbb{R}^{p\times m}$. Note that in general the stationarity conditions for problem (P) obtained via the LLVF reformulation differ significantly from those derived via the OPEC and KKT reformulations, especially due to the second order term appearing in the latter case. Implication "P-KN-stationary (with λ free) $\overset{(2)}{\Longleftarrow}$ P-S-stationary" in the above theorem establishes a clear link between both classes of conditions provided the aforementioned second order term is the zero vector.

3.2 Optimal value reformulation

3.2.1 Failure of the basic CQ and a possible adjustment

Considering the LLVF reformulation (3.3) of (P), we respectively define Ω and ψ by

$$\Omega := \{(x,y)|x \in X, y \in K(x)\} \text{ and } \psi(x,y) := f(x,y) - \varphi(x). \tag{3.38}$$

Hence, problem (3.3) takes the following operator constraint representation

$$\min_{x,y} \{F(x,y)|\,(x,y) \in \Omega \cap \psi^{-1}(\mathbb{R}_-)\} \tag{3.39}$$

and we can now define the corresponding basic CQ, which is shown, in the next theorem, to fail at any feasible point, under a mild assumption

$$\partial\psi(\bar{x},\bar{y}) \cap (-N_\Omega(\bar{x},\bar{y})) = \emptyset. \tag{3.40}$$

Theorem 3.2.1 (failure of the basic CQ for the LLVL reformulation of (P))**.** *Let* $(\bar{x},\bar{y})$ *be an arbitrary feasible point of problem* (3.3)*, where the sum rule*

$$\partial(\psi + \delta_\Omega)(\bar{x},\bar{y}) \subseteq \partial\psi(\bar{x},\bar{y}) + \partial\delta_\Omega(\bar{x},\bar{y}) \tag{3.41}$$

holds, which is automatical if the lower-level value function φ (3.2) *is Lipschitz continuous around* $\bar{x}$*. Then the basic CQ* (3.40) *fails at* $(\bar{x},\bar{y})$*.*

Proof. $(\bar{x},\bar{y})$ being a feasible point of (3.3), we have $(\bar{x},\bar{y}) \in \text{gph}\,S$, and it follows that $\psi(\bar{x},\bar{y}) = 0$. On the other hand, we have for all $x \in X$ that $f(x,y) \geq \varphi(x), \forall y \in K(x)$. Hence, $\psi(x,y) \geq 0, \forall (x,y) \in \Omega$. That is $\psi(\bar{x},\bar{y}) = 0 \leq \psi(x,y), \forall (x,y) \in \Omega$. Hence from (3.41) it follows that $0 \in \partial\psi(\bar{x},\bar{y}) + N_\Omega(\bar{x},\bar{y})$ or equivalently $\partial\psi(\bar{x},\bar{y}) \cap (-N_\Omega(\bar{x},\bar{y})) \neq \emptyset$. □

A similar result was already shown by Ye and Zhu [128] where basically, they chose $\Omega := \mathbb{R}^n \times \mathbb{R}^m$ and $\psi(x,y) := [G(x), g(x,y), f(x,y) - \varphi(x)]$ and used the Clarke subdifferential (2.5) to define their basic CQ, which leads to a stronger CQ. Furthermore, not only this choice of Ω and ψ has as consequence that they needed stronger assumptions to obtained the failure of the basic CQ, but it does not allow for the improvements of the latter CQ that we discuss below. In fact, following the work by Henrion and Outrata [61], we introduce the following weak form of CQ (3.40) by passing to the boundary of the normal cone:

$$\partial\psi(\bar{x},\bar{y}) \cap (-\text{bd}N_\Omega(\bar{x},\bar{y})) = \emptyset. \tag{3.42}$$

As it will be clear in the next section, condition (3.42) is a CQ for LLVF reformulation (3.2), provided Ω is *semismooth* at the reference point $(\bar{x},\bar{y}) \in \Omega$, i.e. for any sequence $x_k \to \bar{x}$ with $x_k \in \Omega$ and $(x_k - \bar{x})\|x_k - \bar{x}\|^{-1} \to d$, it holds that $\langle \bar{x}_k^*, d\rangle \to 0$ for all selections $\bar{x}_k^* \in \bar{\partial} d_\Omega(x_k)$. Note that by choosing Ω and ψ as in [128], the resulting counterpart of CQ (3.42) will coincide with its counterpart of (3.40); and hence still fail.

To illustrate the validity of the weak basic CQ (3.42) for problem (3.3), we now introduce the class of *simple convex bilevel programming problems* studied for example in [23]:

$$\min \{F(x)|\, x \in S := \text{argmin}\{f(x)|x \in \Omega\}\}, \tag{3.43}$$

where Ω is a closed and convex set, and the upper and lower level objective functions F and f are all convex real-valued function. Denoting by $\pi := \min\{f(x)|x \in \Omega\}$, problem (3.43) can be reformulated as

$$\min \{F(x)|\, x \in \Omega,\, f(x) \leq \pi\} \tag{3.44}$$

via the LLVF reformulation (3.3). This is a convex optimization problem, but the MFCQ also fails at any feasible point [23]. In this case, the weak basic CQ (3.42) reduces to:

$$\partial f(\bar{x}) \cap (-\mathrm{bd} N_\Omega(\bar{x})) = \emptyset. \tag{3.45}$$

The following example from the class of simple convex bilevel programming problems shows that the weak basic CQ (3.42) could be a quite useful tool to derive KKT-type necessary optimality conditions via the LLVF reformulation.

Example 1 (validity of the weak basic CQ)**.** *We consider the simple convex bilevel program:*

$$\min \{x^2 + y^2 |\ (x,y) \in S := \mathrm{argmin}\{x + y |\ x, y \geq 0\}\}.$$

We have $f(x,y) := x + y$ *and* $\Omega := \mathbb{R}^2_+$. *The point* $(0,0)$ *is the unique optimal solution of the problem.* Ω *is semismooth as a convex set, and the normal cone to it is* $N_\Omega(0,0) = \mathbb{R}^2_-$. *Hence, we have*

$$\partial f(0,0) = (1,1) \notin \{(x,0)|\ x \geq 0\} \cup \{(0,y)|\ y \geq 0\} = -\mathrm{bd} N_\Omega(0,0).$$

Clearly, condition (3.45) is fulfilled at $(0,0)$ *while the basic CQ does not hold at the same point.*

In the next section, it will be shown that the weak basic CQ (3.42) strictly implies the *partial calmness*, which we introduce next.

3.2.2 The concept of partial calmness

Let $(\bar{x}, \bar{y})$ be a feasible point of the LLVF reformulation (3.3). This problem is said to be *partially calm* at $(\bar{x}, \bar{y})$ if there is a number $\lambda > 0$ and a neighborhood U of $(\bar{x}, \bar{y}, 0)$ such that we have:

$$\begin{aligned} &F(x,y) - F(\bar{x},\bar{y}) + \lambda |u| \geq 0, \\ &\forall (x,y,u) \in U: \ x \in X, y \in K(x),\ f(x,y) - \varphi(x) + u = 0. \end{aligned} \tag{3.46}$$

The following result from [128, Proposition 3.3] emphasizes the importance of the partial calmness concept in the process of deriving necessary optimality for (P) via the LLVF reformulation.

Theorem 3.2.2 (partial exact penalization via partial calmness)**.** *Let* $(\bar{x}, \bar{y})$ *be a local optimal solution of problem* (3.3)*. This problem is partially calm at* $(\bar{x}, \bar{y})$ *if and only if there exists a number* $\lambda > 0$ *such that* $(\bar{x}, \bar{y})$ *is a local optimal solution of the partially penalized problem*

$$\min_{x,y} \{F(x,y) + \lambda (f(x,y) - \varphi(x))|\ x \in X, y \in K(x)\}. \tag{3.47}$$

Clearly, the partial calmness has drawn a lot of attention (see e.g. [26, 36, 35, 65, 91, 122, 128, 130]) because of its capacity to move the optimal value constraint function $f - \varphi$ (responsible for the failure of the basic CQ, cf. Theorem 3.2.1) from the feasible set to the upper-level objective function (3.47), thus paving the way to more tractable constraints in the perspective of KKT-type optimality conditions.

In their seminal paper [128] where Ye and Zhu introduced the partial calmness, it was proven that a bilevel programming problem with a lower-level problem linear in (x,y), is partially calm. We show in the next theorem that this proof can be adapted to the case where the follower's problem is linear only in the lower-level variable y. Precisely we consider the classical optimistic bilevel programming problem

$$\min_{x,y} \left\{F(x,y)|\ x \in X,\ y \in S(x) := \arg\min_y \{a(x)^\top y + b(x)|\ C(x)y \leq d(x)\}\right\} \tag{3.48}$$

with $a : \mathbb{R}^n \to \mathbb{R}^m$, $b : \mathbb{R}^n \to \mathbb{R}$, $C : \mathbb{R}^n \to \mathbb{R}^{p \times m}$ and $d : \mathbb{R}^n \to \mathbb{R}^p$. Its LLVF reformulation is obtained as

$$\min \{F(x,y)|x \in X,\ \ C(x)y \leq d(x),\ \ a(x)^\top y + b(x) \leq \varphi(x)\}, \tag{3.49}$$

where the lower-level value function φ is given by $\varphi(x) := \min_y \{a(x)^\top y + b(x)|C(x)y \leq d(x)\}$.

Theorem 3.2.3 (partial calmness in presence of linearity in the follower's problem with respect to the lower-level variable)**.** *Problem* (3.49) *is partially calm at any of its optimal solutions.*

Proof. Let $(\bar{x},\bar{y})$ be an optimal solution of problem (3.49) and consider a neighborhood U of $(\bar{x},\bar{y},0)$. Furthermore let $(x,y,u) \in U$ such that $x \in X$ and

$$C(x)y \leq d(x),\ a(x)^\top y + b(x) - \varphi(x) + u = 0. \tag{3.50}$$

Since $S(x) \neq \emptyset$ (by the general assumption made in Subsection 3.1.1), let $y(x) \in S(x)$ be a projection of y on $S(x)$. Denoting by $e = (1,\dots,1)^\top$ a m-dimensional vector, we have

$$\begin{aligned}\|y - y(x)\| &= \min_z \{\|y - z\| : z \in S(x)\} \\ &= \min_{\varepsilon,z} \{\varepsilon :\ \|y - z\| \leq \varepsilon,\ z \in S(x)\} \\ &= \min_{\varepsilon,z} \{\varepsilon :\ -\varepsilon e \leq y - z \leq \varepsilon e,\ z \in S(x)\}.\end{aligned}$$

The last equality describes the linear program

$$\min_{\varepsilon,z} \{\varepsilon| \quad -\varepsilon e + z \leq y,\ -\varepsilon e - z \leq -y,\ C(x)z \leq d(x),\ a(x)^\top y \leq \varphi(x) - b(x)\},$$

having as dual the problem

$$\begin{aligned}\max_{\xi_1,\xi_2,\xi_3,\xi_4} \ &\{y^\top \xi_1 - y^\top \xi_2 + (\varphi(x) - b(x))\xi_3 + d(x)^\top \xi_4 |\ -e^\top \xi_1 - e^\top \xi_2 = 1, \\ &\xi_1 - \xi_2 + \xi_3 a(x) + C(x)^\top \xi_4 = 0,\ (\xi_1,\xi_2,\xi_3,\xi_4) \in \mathbb{R}^m_- \times \mathbb{R}^m_- \times \mathbb{R}_- \times \mathbb{R}^p_-\}.\end{aligned}$$

By inserting the constraint $\xi_1 - \xi_2 + \xi_3 a(x) + C(x)^\top \xi_4 = 0$ in the objective function of the latter problem, we have the following equivalent problem

$$\begin{aligned}\max_{\xi_1,\xi_2,\xi_3,\xi_4} \ &\Big\{\xi_3\big(\varphi(x) - a(x)^\top y - b(x)\big) + \big(d(x) - C(x)y\big)^\top \xi_4 \big| \\ &-e^\top \xi_1 - e^\top \xi_2 = 1,\ \xi_1,\xi_2,\xi_3,\xi_4 \leq 0\Big\}.\end{aligned}$$

Thus there is at least one vertex $(\xi_1^o,\xi_2^o,\xi_3^o,\xi_4^o)$ of the system

$$-e^\top \xi_1 - e^\top \xi_2 = 1,\ \xi_1,\xi_2,\xi_3,\xi_4 \leq 0 \tag{3.51}$$

such that

$$\|y - y(x)\| = \xi_3^o\big(\varphi(x) - a(x)^\top y - b(x)\big) + \big(d(x) - C(x)y\big)^\top \xi_4^o,$$

which implies

$$\|y - y(x)\| \leq \xi_3^o u$$

given that (x,y) satisfies (3.50). Also notice that $u \leq 0$ considering the definition of φ. Since the number of vertices satisfying (3.51) is finite, let $\xi_3^B \in \mathbb{R}$ be the smallest $(2m+1)^{\text{th}}$ component of such vertices, then

$$\|y - y(x)\| \leq |\xi_3^B||u|. \tag{3.52}$$

On the other hand we recall that F is Lipschitz continuous. Denote by ℓ_F its Lipschitz modulus, then given that $(\bar{x},\bar{y})$ is an optimal solution to problem (3.49) and $(x,y(x))$ being a feasible point, we have

$$\begin{aligned}F(x,y) - F(\bar{x},\bar{y}) &\geq F(x,y) - F(x,y(x)) \\ &\geq -\ell_F\|y - y(x)\| \\ &\geq -\lambda|u|,\end{aligned}$$

with $\lambda := \ell_F|\xi_3^B|$ while taking into account inequality (3.52). □

It was later shown in [122] that problem (3.49) has a *uniform weak sharp minimum* (the definition is given below) provided there exists $\lambda > 0$ such that

$$c(x) := \sup_{w,y,I} \left\{ w_{p+1} \;\middle|\; \begin{array}{l} \|\sum_{i\in I} w_i \nabla_y \bar{g}_i(x,y)\|_1 = 1,\ y \in S(x),\ w_i > 0,\ \bar{g}_i(x,y) = 0,\ \forall i \in I, \\ \text{vectors } \{\nabla_y \bar{g}_i(x,y) |\, i \in I\} \text{ are linearly independent,} \\ \{p+1\} \subseteq I \subseteq \{1,\dots,p+1\} \end{array} \right\} \leq \lambda, \quad (3.53)$$

for all $x \in X$ such that there exists I as in the previous line. Here, $\bar{g}(x,y) := (g(x,y),\ f(x,y) - \varphi(x))$ with $g(x,y) := C(x)y - d(x)$ and $f(x,y) := a(x)^\top y + b(x)$. Not only this assumption is difficult to check (see [78]), but it is clear from Theorem 3.2.3 that in the perspective of deriving necessary optimality conditions, it is in fact not necessary, cf. Corollary 3.2.13.

In the general case, to obtain sufficient conditions for partial calmness, the concept of weak sharp minimum by Ferris [45] (also see e.g. [9]) was used in [128]. We now introduce a notion of weak sharp minimum taken from Henrion et al. [59], which generalizes the one used by Ye and Zhu [128]. For this purpose, we consider the general optimization problem (2.36), where Ω is closed and f is continuous.

Definition 3.2.4 (weak sharp minima)**.** *In* (2.36)*, the function f is said to have S as set of weak sharp minima with respect to $\Omega \cap \mathscr{N}$, if there exists $\lambda > 0$ such that we have*

$$f(x) - f_* \geq \lambda\, d(x,S) \;\; \textit{for all} \;\; x \in \Omega \cap \mathscr{N},$$

where $f_* := \inf\{f(x) |\, x \in \Omega\}$ and $\mathscr{N}$ is a neighborhood of S.

By replacing the set $\mathscr{N}$ by the whole space $\mathbb{R}^n$, we obtain the definition used by Ye and Zhu [128]. As we will see in what follows, Definition 3.2.4 can lead to a new sufficient condition for partial calmness. Before heading to that, it seems appropriate to first recall the link between the partial calmness and the notion of weak sharp minimum. For this, we bring the previous definition to the context of the parametric optimization problem (1.3). The family of parametric problems $\{(1.3) |\, x \in X\}$ will be said to have a *uniformly weak sharp minimum*, if there exist $\lambda > 0$ and a neighborhood $\mathscr{N}(x)$ of $S(x)$ for $x \in X$ such that we have

$$f(x,y) - \varphi(x) \geq \lambda\ d(y, S(x)),\ \forall y \in K(x) \cap \mathscr{N}(x),\ \forall x \in X. \quad (3.54)$$

The term *uniformly weak sharp minimum* was first used by Ye and Zhu [128]. We now present the following result, without the proof, since it can easily follow as in [128, Proposition 5.1].

Theorem 3.2.5 (uniform weak sharp minimum implies partial calmness)**.** *Let $(\bar{x},\bar{y})$ be an optimal solution of problem* (3.3)*. Assume that F is Lipschitz continuous in y uniformly in $x \in X$, and the family $\{(1.3) |\, x \in X\}$ has a uniformly weak sharp minimum* (3.54)*. Then problem* (3.3) *is partially calm at $(\bar{x},\bar{y})$.*

Now, we define the family of functions $\psi_x(y) := f(x,y) - \varphi(x)$ and set-valued mappings

$$\Psi_x(z) := \{y \in K(x) |\, \psi_x(y) + z \leq 0\} \quad (3.55)$$

for all $x \in X$. The family of mappings $\{\Psi_x | x \in X\}$ will be said to be *uniformly calm* if for each $x \in X$, Ψ_x is calm on $\{0\} \times S(x)$ and there is a family $\{\ell_x(y) |\, y \in S(x), x \in X\}$ of calmness constants satisfying

$$\ell_x(y) \leq \lambda,\ \forall y \in S(x),\ x \in X; \quad (3.56)$$

where λ is a positive number. It should be clear in this definition that when we fix $x \in X$, and a vector $y \in S(x)$, we consider only a certain calmness constant $\ell_x(y)$ satisfying the corresponding calmness condition (2.33) and such that $\ell_x(y) \leq \lambda$. It is also clear from (3.55) that $\Psi_x(0) = S(x)$. Next we give a new sufficient condition for problem (3.3) to be partially calm. The proof is inspired from [59].

Theorem 3.2.6 (uniform calmness implies uniform weak sharp minimum)**.** *Let the family* $\{\Psi_x|\ x\in X\}$ *be uniformly calm and the set-valued mapping* S (1.2) *be compact-valued. Furthermore, assume that the follower's cost function* f *is continuous in* y*. Then the family* $\{(1.3)|\ x\in X\}$ *has a uniformly weak sharp minimum* (3.54).

Proof. Fix $x_o\in X$, since $S(x_o)\neq\emptyset$, let $y\in S(x_o)$. Ψ_{x_o} (3.55) is calm at $(0,y)$ and it follows from the definition of calmness (2.33), that there exist $\varepsilon_{x_o},\delta_{x_o}>0$ such that

$$d(u,\Psi_{x_o}(0))\leq \ell_{x_o}(y)|z|, \forall z:|z|<\delta_{x_o}, \forall u\in\Psi_{x_o}(z)\cap B(y,\varepsilon_{x_o}). \tag{3.57}$$

Since f is continuous in u, then we can choose ε_{x_o} small enough such that

$$|f(x_o,u)-f(x_o,y)|<\delta_{x_o}, \forall u\in B(y,\varepsilon_{x_o}),$$

with $f(x_o,y)=\varphi(x_o)$. Thus we have $|\psi_{x_o}(u)|<\delta_{x_o}, \forall u\in B(y,\varepsilon_{x_o})$. Hence, by taking $z=\psi_{x_o}(u)$ in (3.57) and observing that $\Psi_{x_o}(0)=S(x_o)$, we have

$$d(u,S(x_0))\leq \ell_{x_o}(y)|\psi_{x_o}(u)|,\ \forall u\in K(x_o)\cap B(y,\varepsilon_{x_o}).$$

Since the solution set-valued mapping S is compact-valued; then, there is a finite number of vectors $y^i\in S(x_o)$ and real numbers $\delta^i_{x_o}>0$ and $\ell_{x_o}(y^i)>0$ such that $S(x_o)\subseteq\underset{i}{\cup}B(y^i,\delta^i_{x_o})$, and we have

$$d(u,S(x_0))\leq \ell_{x_o}(y^i)(f(x_o,u)-\varphi(x_o)),\ \forall u\in K(x_o)\cap B(y^i,\delta^i_{x_o}),\ \forall i.$$

By taking $\mathscr{N}(x_o)=\underset{i}{\cup}B(y^i,\delta^i_{x_o})$, it follows that

$$d(u,S(x_0))\leq c(x_o)(f(x_o,u)-\varphi(x_o)),\ \forall u\in K(x_o)\cap\mathscr{N}(x_o),$$

where $c(x_o)=\max\limits_i \ell_{x_o}(y^i)$. Since the family of multifunctions $\{\Psi_x|x\in X\}$ is uniformly calm, we assume without lost of generality that the family of calmness constants $\{\ell_x(y)|\ y\in S(x),\ x\in X\}$ is chosen in such a way that inequality (3.56) is satisfied. Hence, there exists $\lambda>0$ with $c(x)<\lambda,\forall x\in X$ such that

$$f(x,y)-\varphi(x)\geq\lambda^{-1}d(y,S(x)),\ \forall y\in K(x)\cap\mathscr{N}(x),\ \forall x\in X.$$

This completes the proof of the result. □

A second approach to obtain a result similar to Theorem 3.2.6 for the family $\{(1.3)|x\in X\}$ to have a uniformly weak sharp minimum is to consider but the family of multifunctions

$$\Psi_x(z):=\{y\in K(x)|\,f(x,y)\leq z\},\ \text{ for }\ x\in X$$

instead of that in (3.55). Hence, in the definition of the uniform calmness of the family of multifunctions $\{\Psi_x|x\in X\}$, consider but the calmness of each Ψ_x on $\{\varphi(x)\}\times S(x)$ instead of $\{0\}\times\Psi(x)$. A result similar to Theorem 3.2.6 can then be directly obtained from [59, Lemma 4.7].

In the next result we provide a sufficient condition for the uniform calmness while providing a formula for the *calmness modulus of the set-valued-mapping* Ψ_x (3.55), which can equivalently be defined at a point $(0,y)\in\operatorname{gph}\Psi_x$ by

$$\operatorname{cal}\Psi_x(0,y):=\inf\{\ell_x\in(0,\infty)|\ d(u,\ \Psi_x(0))\leq\ell_x\|v\|,\ \forall v\in\mathbb{B}(0,\delta_x),\ \forall u\in\Psi_x(0)\cap\mathbb{B}(y,\delta_x)\},$$

where $\mathbb{B}(a,\delta_x)$ denotes the ball of center a and radius δ_x.

Proposition 3.2.7 (sufficient condition for uniform calmness)**.** *For all $x \in X$, assume that the functions $f(x,.)$ and $g(x,.)$ are convex while the set-valued mapping Ψ_x (3.55) is calm on $\{0\} \times S(x)$. Then the family $\{\Psi_x |\ x \in X\}$ is uniformly calm provided there exists a number $\lambda > 0$ such that we have*

$$\begin{aligned} \mathrm{cal}\,\Psi_x(0,y) &= \limsup_{v \in S(x),\, v \to y} \|D^*\Psi_x(0,v)|_{-N_{S(x)}(v)}\|^- \\ &= \min\{\ell_x \in (0,\infty) |\ d(u,\, \Psi_x(0)) \le \ell_x \|v\|,\ \forall v \in \mathbb{B}(0,\delta_x),\ \forall u \in \Psi_x(0) \cap \mathbb{B}(y,\delta_x)\} \\ &\le \lambda,\ \forall y \in S(x),\ x \in X. \end{aligned} \tag{3.58}$$

Proof. Fix $x \in X$, if $f(x,.)$ and $g(x,.)$ are convex, then $\mathrm{gph}\,\Psi_x$ is a convex set. Hence, the first equality in (3.58) follows from (2.34), given that $\Psi_x(0) = S(x)$ and $S(x)$ is convex. Finally, if $\mathrm{cal}\,\Psi_x(0,y)$ is achieved as a minimum for all $y \in S(x)$, $x \in X$ (cf. second equality in (3.58)), then we choose $\ell_x(y) = \mathrm{cal}\,\Psi_x(0,y)$, $y \in \Psi(x)$, $x \in X$ and it follows from the inequality in (3.58) that the family $\{\Psi_x |\ x \in X\}$ is uniformly calm. □

Theorem 3.2.8 (weak basic CQ implies partial calmness)**.** *Let $(\bar{x},\bar{y})$ be a local optimal solution of problem* (3.3)*, where Ω* (3.38) *is semismooth while φ is Lipschitz continuous around $\bar{x}$. Then, the following implication holds true at $(\bar{x},\bar{y})$*

$$\textit{weak basic CQ } (3.42) \Longrightarrow \textit{partial calmness}. \tag{3.59}$$

Proof. Recall that under the convexity of Ω and the Lipschitz continuity of φ around $\bar{x}$, condition (3.42) is well-defined to be a CQ at a given point $(\bar{x},\bar{y})$. Moreover, it follows from [61] that the set-valued mapping

$$\Psi(u) = \{(x,y) \in \Omega |\ f(x,y) - \varphi(x) \le u\} \tag{3.60}$$

is calm at $(0,\bar{x},\bar{y})$. Also observe that $\Psi(0)$ coincides with the feasible set of problem (3.3). Hence, it follows from [16] that, if $(\bar{x},\bar{y})$ is a local optimal solution of (3.3), then there exist $\lambda > 0$ and a neighborhood W of $(\bar{x},\bar{y})$, such that

$$F(\bar{x},\bar{y}) \le F(x,y) + \lambda\, d((x,y),\Psi(0)),\ \forall (x,y) \in W. \tag{3.61}$$

On the other hand, the calmness of Ψ at $(0,\bar{x},\bar{y})$ implies that there exist neighborhoods U of 0, V of $(\bar{x},\bar{y})$ and a constant $\ell > 0$ such that

$$d((x,y),\Psi(0)) \le \ell |u|,\ \forall u \in U,\ \forall (x,y) \in \Psi(u) \cap V. \tag{3.62}$$

By combining (3.61) and (3.62), it follows that

$$F(x,y) - F(\bar{x},\bar{y}) + \lambda \ell |u| \ge 0, \forall u \in U, \forall (x,y) \in W \cap V \cap \Psi(u),$$

which coincides with the definition of partial calmness given in (3.46). □

In the next example, we show that the converse of implication (3.59) is not always possible.

Example 2 (implication (3.59) is strict)**.** *Consider the bilevel program*

$$\min\ \left\{x^2 + y^2 |\ x \in \mathbb{R},\ y \in \operatorname{argmin}\{x^2 y + y |\ y \ge 0\}\right\}. \tag{3.63}$$

Set $f(x,y) := x^2 y + y$ and $\Omega := \mathbb{R} \times \mathbb{R}_+$. We have $\varphi(x) = 0, \forall x \in X := \mathbb{R}$, and $\psi(x,y) = x^2 y + y$. We can easily check that $(0,0)$ is an optimal solution of problem (3.63)*. Hence $\partial\psi(0,0) = (0,1)$ and $N_\Omega(0,0) = \{0\} \times \mathbb{R}_-$. It follows that $(0,1) \in \{0\} \times \mathbb{R}_+ = -\mathrm{bd} N_\Omega(0,0)$. This means that CQ* (3.42) *fails at $(0,0)$. On the other hand, it follows from Theorem 3.2.3 that problem (3.63) is partially calm at $(0,0)$.*

To conclude this subsection, we summarize the relationships between the the constraint qualifications discussed above in the following diagram, where "u.w." is an abbreviation of "uniform weak":

$$\begin{array}{ccccc} \text{uniform calmness} & & \text{weak basic CQ (3.42)} & & \\ \Downarrow (1) & & \Downarrow & & \\ \text{u.w. sharp minimum} & \Longrightarrow & \text{calmness of } \Psi \text{ (3.60)} & \Longrightarrow & \text{partial calmness} \end{array} \tag{3.64}$$

with (1) standing for all the assumptions in Theorem 3.2.6.

3.2.3 Necessary optimality conditions

In this section we derive the KM- and KN-stationarity conditions introduced in Subsection 3.1.2 while paying attention to the special case where the lower-level problem is linear in the follower's variable or strongly stable in the sense of Kojima [71]. Moreover, necessary optimality conditions for the simple convex bilevel programming problem will also be discussed. Recall that the particularity of the LLVF reformulation of (P) relies on the value function (3.2), which is a typical nonsmooth function. Hence, to proceed we first provide the sensitivity analysis of $-\varphi$ that will be useful here. For this purpose, let us introduce the *lower-level regularity* that will also be needed in the other parts of the thesis:

$$\Big[\sum_{i=1}^{p}\beta_i\nabla_y g_i(\bar{x},\bar{y})=0,\ \beta_i\geq 0,\ \beta_i g_i(\bar{x},\bar{y})=0,\ i=1,\ldots,p\Big]\Longrightarrow \beta_i=0,\ i=1,\ldots,p. \tag{3.65}$$

Theorem 3.2.9 (sensitivity analysis of the negative value function in the lower-level problem)**.** *The following assertions hold for the negation of the value function φ in* (3.2)*:*

(i) *If the solution map S* (1.2) *is inner semicompact at $\bar{x}$, and $(\bar{x},y)$ is lower-level regular* (3.65) *for all $y\in S(\bar{x})$, then φ is Lipschitz continuous around $\bar{x}$ and we have the inclusion*

$$\partial(-\varphi)(\bar{x})\subseteq\ \Big\{-\textstyle\sum_{s=1}^{n+1}\upsilon_s\Big(\nabla_x f(\bar{x},y_s)+\sum_{i=1}^{p}\mu_{is}\nabla_x g_i(\bar{x},y_s)\Big)\Big|\ \sum_{s=1}^{n+1}\upsilon_s=1,\ \forall s=1,\ldots,n+1,\ \upsilon_s\geq 0,\ y_s\in S(\bar{x}),\ \mu_s\in\Lambda(\bar{x},y_s)\Big\}. \tag{3.66}$$

(ii) *Assume that $(\bar{x},\bar{y})$ is lower-level regular* (3.65)*, and that* EITHER *S* (1.2) *is inner semicontinuous at this point* OR *f and g_i, $i=1,\ldots,p$ are fully convex. Then φ is Lipschitz continuous around $\bar{x}$ and the basic subdifferential of $-\varphi$ is estimated as:*

$$\partial(-\varphi)(\bar{x})\subseteq\bigcup_{\mu\in\Lambda(\bar{x},\bar{y})}\Big\{-\nabla_x f(\bar{x},\bar{y})-\sum_{i=1}^{p}\mu_i\nabla_x g_i(\bar{x},\bar{y})\Big\}. \tag{3.67}$$

Proof. The local Lipschitz continuity of φ is justified in [90] under the fulfillment of the lower-level regularity (3.65) in both inner semicontinuous and inner semicompactness cases. If the functions f and g_i, $i=1,\ldots,p$ are fully convex, then the value function φ is convex as well; in this case the Lipschitz continuity follows from [16]. To prove the subdifferential inclusion in (i), recall that

$$\partial\varphi(\bar{x})\subseteq\bigcup_{y\in S(\bar{x})}\ \bigcup_{\mu\in\Lambda(\bar{x},y)}\Big\{\nabla_x f(\bar{x},y)+\sum_{i=1}^{p}\mu_i\nabla_x g_i(\bar{x},\bar{y})\Big\},$$

by [92] under the assumptions made in (i). The claimed estimate of $\partial(-\varphi)$ follows from here by the classical Carathéodory's theorem while noting that we have

$$\partial(-\varphi)(\bar{x})\subseteq \mathrm{co}\,\partial(-\varphi)(\bar{x})=-\mathrm{co}\,\partial\varphi(\bar{x}), \tag{3.68}$$

where the second equality follows from the convex hull property (2.6).

When S is inner semicontinuous at $(\bar{x},\bar{y})$, we have by [91] that

$$\bar{\partial}\varphi(\bar{x})\subseteq\bigcup_{\mu\in\Lambda(\bar{x},\bar{y})}\Big\{\nabla_x f(\bar{x},\bar{y})+\sum_{i=1}^{p}\mu_i\nabla_x g_i(\bar{x},\bar{y})\Big\}, \tag{3.69}$$

which implies the subdifferential inclusion in (ii) by the combination of (2.5) and (3.68). For the estimate of $\partial(-\varphi)(\bar{x})$ in the convex case, note that with expression of the subdifferential in the sense of convex analysis (2.9) and the fact that $(\bar{x},\bar{y})\in\mathrm{gph}\,S$, a vector $u\in\partial\varphi(\bar{x})$ implies

$$f(x,y)-\langle u,x\rangle\geq f(\bar{x},\bar{y})-\langle u,\bar{x}\rangle,\ \forall(x,y):\ g(x,y)\leq 0,$$

which means that $(\bar{x},\bar{y})$ is an optimal solution of the problem

$$\min_{x,y}\{f(x,y)-\langle u,x\rangle\,|\,g(x,y)\le 0\}.$$

Hence, applying the Lagrange multiplier rule to the latter problem, it follows, under the lower-level regularity (3.65), that there exists $\mu\in\Lambda(\bar{x},\bar{y})$ such that we have $u=\nabla_x f(\bar{x},\bar{y})+\sum_{i=1}^{p}\mu_i\nabla_x g_i(\bar{x},\bar{y})$. Thus an upper bound of the basic subdifferential of φ can be obtained as

$$\partial\varphi(\bar{x})\subseteq\bigcup_{\mu\in\Lambda(\bar{x},\bar{y})}\Big\{\nabla_x f(\bar{x},\bar{y})+\sum_{i=1}^{p}\mu_i\nabla_x g_i(\bar{x},\bar{y})\Big\} \tag{3.70}$$

and inclusion (3.67) follows from the combination of (3.70) and (3.68), while noting that we have $\bar{\partial}\varphi(\bar{x}):=\mathrm{co}\,\partial\varphi(\bar{x})=\partial\varphi(\bar{x})$ in this case. □

Note that in the fully convex (even nonsmooth) case, the lower-level regularity assumption in Theorem 3.2.9 can be replaced by a much weaker qualification condition [40] requiring that the set

$$\mathrm{epi}\,f^*+\mathrm{cone}\,\Big(\bigcup_{i=1}^{p}\mathrm{epi}\,g_i^*\Big)\text{ is closed on }\mathbb{R}^n\times\mathbb{R}^m\times\mathbb{R}, \tag{3.71}$$

where $\mathrm{epi}\,\vartheta^*$ denotes the conjugate function for an extended-real-valued convex function ϑ. It is also important to mention that inclusion (3.66) can similarly be deduced from the well-known (see e.g. [46, 55]) upper estimate of the Clarke subdifferential of the value function φ.

To state the first result on the stationarity conditions for (P) via the LLVF reformulation (3.3), we also need the following *upper-level regularity* at $\bar{x}$:

$$\Big[\sum_{j=1}^{k}\alpha_j\nabla G_j(\bar{x})=0,\ \ \alpha_j\ge 0,\ \alpha_j G_j(\bar{x})=0,\ \ j=1,\dots,k\Big]\Longrightarrow\alpha_j=0,\ j=1,\dots,k. \tag{3.72}$$

Theorem 3.2.10 (KM- and KN-stationarity conditions for (P) via the LLVF reformulation and under the weak basic CQ)**.** *Let* $(\bar{x},\bar{y})$ *be a local optimal solution of problem* (3.3)*, and assume that* Ω (3.38) *is semismooth at the point* $(\bar{x},\bar{y})$*, where CQ* (3.42) *also holds. Furthermore, let* $\bar{x}$ *be upper-level regular* (3.72)*. Then the following assertions are satisfied:*

(i) *If the solution map S in* (1.2) *is inner semicompact at* $\bar{x}$*, while the lower-level regularity* (3.65) *holds at* $(\bar{x},y)$*,* $y\in S(\bar{x})$*. Then* $(\bar{x},\bar{y})$ *is KM-stationary with* $\lambda\in[0,r]$*, for some* $r>0$*.*

(ii) *Assume that* $(\bar{x},\bar{y})$ *is lower-level regular* (3.65)*, and that* EITHER *S is inner semicontinuous at this point* OR *f and g are fully convex. Then* $(\bar{x},\bar{y})$ *is KN-stationary with* $\lambda\in[0,r]$*, for some* $r>0$*.*

Proof. We provide here only the proof for (i), since the (ii) case follows similarly. Consider the operator constraint representation (3.39) of problem (3.3). We have from Theorem 3.2.9 (i) that the value function (3.2) is Lipschitz continuous around $\bar{x}$. In addition to the continuity of G and g (1.4), $\Omega\cap\psi^{-1}(\mathbb{R}_-)$ is closed around $(\bar{x},\bar{y})$. Furthermore, recall that the set-valued mapping Ψ (3.60) is calm at $(\bar{x},\bar{y})$, given that Ω (3.38) is semismooth at $(\bar{x},\bar{y})$, where CQ (3.42) also holds, cf. [61, Theorem 3.1]. Thus we have from Proposition 2.2.5 that there exists $\lambda\in[0,r]$ (for some $r>0$) such that

$$0\in\nabla F(\bar{x},\bar{y})+\lambda\partial\psi(\bar{x},\bar{y})+N_\Omega(\bar{x},\bar{y}) \tag{3.73}$$

with Ω and ψ given in (3.38). The basic subdifferential of ψ and the normal cone to Ω are respectively given by:

$$\partial\psi(\bar{x},\bar{y})=\nabla f(\bar{x},\bar{y})+\partial(-\varphi)(\bar{x})\times\{0\}, \tag{3.74}$$

$$N_\Omega(\bar{x},\bar{y})=\left\{\left[\begin{array}{r}\sum_{j=1}^{k}\alpha_i\nabla G_i(\bar{x})+\sum_{i=1}^{p}\beta_i\nabla_x g_i(\bar{x},\bar{y})\\ \sum_{i=1}^{p}\beta_i\nabla_y g_i(\bar{x},\bar{y})\end{array}\right]\middle|\ \begin{array}{l}\alpha_j\ge 0,\ \alpha_j G_j(\bar{x})=0,\ j=1,\dots,k\\ \beta_i\ge 0,\ \beta_i g_i(\bar{x},\bar{y})=0,\ i=1,\dots,p\end{array}\right\}, \tag{3.75}$$

where equality (3.74) follows from the sum rule (2.12). As for equality (3.75), it is obtained from Theorem 2.2.7 provided the following basic-type CQ (2.41) holds at $(\bar{x},\bar{y})$:

$$\left.\begin{array}{r}\sum_{j=1}^{k}\alpha_i\nabla G_i(\bar{x})+\sum_{i=1}^{p}\beta_i\nabla_x g_i(\bar{x},\bar{y})=0\\ \sum_{i=1}^{p}\beta_i\nabla_y g_i(\bar{x},\bar{y})=0\\ \alpha_j\geq 0,\ \alpha_j G_j(\bar{x})=0,\ j=1,\dots,k\\ \beta_i\geq 0,\ \beta_i g_i(\bar{x},\bar{y})=0,\ i=1,\dots,p\end{array}\right\}\Longrightarrow\left\{\begin{array}{l}\alpha_j=0,\ j=1,\dots,k,\\ \beta_i=0,\ i=1,\dots,p.\end{array}\right. \tag{3.76}$$

Now, note that under the the lower-level regularity (3.65) at $(\bar{x},\bar{y})$, the next implication is also satisfied:

$$\left.\begin{array}{r}\sum_{i=1}^{p}\beta_i\nabla_y g_i(\bar{x},\bar{y})=0\\ \beta_i\geq 0,\ \beta_i g_i(\bar{x},\bar{y})=0,\ i=1,\dots,p\end{array}\right\}\Longrightarrow\sum_{i=1}^{p}\beta_i\nabla_x g_i(\bar{x},\bar{y})=0. \tag{3.77}$$

Thus the fulfillment of the lower- and upper-level regularity conditions (3.65) and (3.72), simultaneously at $\bar{x}$ and $(\bar{x},\bar{y})$, imply the satisfaction of condition (3.76). The optimality conditions in (i) are derived from (3.66) and (3.73)–(3.75). □

In the vein of this theorem, we provide necessary optimality conditions for the simple convex bilevel programming problem (3.43).

Theorem 3.2.11 (stationarity conditions for the simple convex bilevel program (3.43))**.** *Let $\bar{x}$ be a local optimal solution of problem* (3.43)*, where $\Omega:=\{x|\,g_i(x)\leq 0,\ i=1,\dots,p\}$ with g_i, $i=1,\dots,p$ convex. Assume that CQ* (3.45) *holds at $\bar{x}$, while there exists x_o such that $g_i(x_o)<0$, $i=1,\dots,p$. Then there exist $\lambda\in[0,r]$ (for some $r>0$) and $\beta\in\mathbb{R}^p$ such that we have:*

$$\nabla F(\bar{x})+\lambda\nabla f(\bar{x})+\sum_{i=1}^{p}\beta_i\nabla g_i(\bar{x})=0,$$
$$\forall i=1,\dots,p,\ \beta_i\geq 0,\ \beta_i g_i(\bar{x})=0.$$

Proof. Problem (3.43) can be reformulated as

$$\min_x\,\{F(x)|\,x\in\Omega\cap\psi^{-1}(\mathbb{R}_-)\}\ \text{ with }\ \psi(x):=f(x)-\pi.$$

Since Ω is semismooth thanks to the convexity of Ω, proceeding as in the proof of Theorem 3.2.10, we have the result under CQ (3.45). □

Optimality conditions for the simple convex bilevel program were derived in [23] under a CQ in the stream of (3.71). Coming back to the general problem (P), we now derive the KM- and KN-stationarity conditions under the partial calmness condition (3.46).

Theorem 3.2.12 (KM- and KN-stationarity conditions for (P) under the partial calmness condition)**.** *Let $(\bar{x},\bar{y})$ be a local optimal solution of problem* (3.3)*, which is is partially calm at the same point. Furthermore, assume that $\bar{x}$ is upper-level regular* (3.72)*. Then, the following assertions hold:*

(i) *Let the set-valued mapping S* (1.2) *be inner semicompact at $\bar{x}$ while the point $(\bar{x},y)$ is lower-level regular* (3.65) *for all $y\in S(\bar{x})$. Then $(\bar{x},\bar{y})$ is KM-stationary with $\lambda>0$ and $\|(\alpha,\beta)\|\leq r$, for some $r>0$.*

(ii) *Assume that $(\bar{x},\bar{y})$ is lower-level regular* (3.65)*, and that* EITHER *S is inner semicontinuous at this point* OR *f and g_i, $i=1,\dots,p$ are fully convex. Then $(\bar{x},\bar{y})$ is KN-stationary with $\lambda>0$ and $\|(\alpha,\beta)\|\leq r$, for some $r>0$.*

Proof. Under the partial calmness condition, it follows from Theorem 3.2.2 that $(\bar{x},\bar{y})$ is a local optimal solution of the partially penalized problem

$$\min_{x,y}\,\{F(x,y)+\lambda(f(x,y)-\varphi(x))|\ \psi(x,y)\in\Lambda\}, \tag{3.78}$$

where $\psi(x,y) := [G(x),\ g(x,y)]$ and $\Lambda := \mathbb{R}^k_- \times \mathbb{R}^p_-$. Since φ is Lipschitz continuous around $\bar{x}$ (cf. Theorem 3.2.9), applying Proposition 2.2.5 to problem (3.78), there exists $(\alpha,\beta) \in N_\Lambda(\psi(\bar{x},\bar{y}))$ with $\|(\alpha,\beta)\| \le r$ (for some $r > 0$) such that we have

$$0 \in \nabla F(\bar{x},\bar{y}) + \lambda \nabla f(\bar{x},\bar{y}) + \lambda \partial(-\varphi)(\bar{x}) \times \{0\} + \partial\langle(\alpha,\beta),\psi\rangle(\bar{x},\bar{y}), \tag{3.79}$$

provided condition (3.76) holds at $(\bar{x},\bar{y})$. Recall that the latter holds given that the lower- and upper-level regularity conditions are satisfied. Further observe that we respectively have

$$N_\Lambda(\psi(\bar{x},\bar{y})) = \left\{(\alpha,\beta) \;\middle|\; \begin{array}{l} \alpha_j \ge 0,\ \alpha_j G_j(\bar{x}) = 0,\ j = 1,\dots,k \\ \beta_i \ge 0,\ \beta_i g_i(\bar{x},\bar{y}) = 0,\ i = 1,\dots,p \end{array}\right\}, \tag{3.80}$$

$$\partial\langle(\alpha,\beta),\psi\rangle(\bar{x},\bar{y}) = \left[\begin{array}{r} \sum_{j=1}^k \alpha_i \nabla G_i(\bar{x}) + \sum_{i=1}^p \beta_i \nabla_x g_i(\bar{x},\bar{y}) \\ \sum_{i=1}^p \beta_i \nabla_y g_i(\bar{x},\bar{y}) \end{array}\right]. \tag{3.81}$$

Hence, the results in Theorem 3.2.12 (i) and (ii) follow from the combination of (3.79)–(3.81) and Theorem 3.2.9 (i) and (ii), respectively. □

As a first application of this result, we consider the bilevel programming problem (3.48), where the lower-level problem is linear in the follower's variable y. To proceed, we specify the components of the matrix $C(x)$ in (3.48) by $C(x) := (c_{il}(x))_{1\le i\le p,\,1\le l\le m}$, and the expression of the lower-level solution set-valued mapping by

$$S(x) := \arg\min_y \{a(x)^\top y + b(x) \,|\, C(x)y \le d(x)\}. \tag{3.82}$$

Corollary 3.2.13 (stationarity conditions for a bilevel program where the lower-level is linear in the follower's variable)**.** *Let $(\bar{x},\bar{y})$ be an optimal solution of problem* (3.48) *and assume that $\bar{x}$ is upper-level regular, while there exists $y^{\bar{x}}$ such that $\sum_{l=1}^m y_l^{\bar{x}} c_{il}(\bar{x}) < d_i(\bar{x})$, $i = 1,\dots,p$. Then, the following assertions are satisfied:*

(i) *Let the solution set-valued mapping S* (3.82) *be inner semicompact at $\bar{x}$. Then, there exist $(\alpha,\beta) \in \mathbb{R}^{k+p}$ with $\|(\alpha,\beta)\| < r$ (for some $r > 0$), $\lambda \in \mathbb{R}^*_+$, $(\mu_s,\upsilon_s) \in \mathbb{R}^{k+1}$ and $y_s \in S(\bar{x})$ with $s = 1,\dots,n+1$ such that relationships* (3.16) *and* (3.18) *together with the following conditions are satisfied:*

$$\begin{aligned}\nabla_x F(\bar{x},\bar{y}) + \sum_{j=1}^k \alpha_j \nabla G_j(\bar{x}) + \lambda\Big(\sum_{l=1}^m \bar{y}_l \nabla a_l(\bar{x}) - \sum_{s=1}^{n+1} \upsilon_s \sum_{l=1}^m y_{ls} \nabla a_l(\bar{x})\Big) \\ + \sum_{i=1}^p \beta_i \Big(\sum_{l=1}^m \bar{y}_l \nabla c_{il}(\bar{x}) - \nabla d_i(\bar{x})\Big) - \lambda \sum_{s=1}^{n+1} \upsilon_s \sum_{i=1}^p \mu_{is} \Big(\sum_{l=1}^m \bar{y}_{ls} \nabla c_{il}(\bar{x}) - \nabla d_i(\bar{x})\Big) = 0,\end{aligned} \tag{3.83}$$

$$\nabla_y F(\bar{x},\bar{y}) + \lambda a(\bar{x}) + \sum_{i=1}^p \beta_i [c_{i1}(\bar{x}),\dots,c_{im}(\bar{x}))]^\top = 0, \tag{3.84}$$

$$\forall s = 1,\dots,n+1,\ \nabla a(\bar{x}) + \sum_{i=1}^p \mu_{is} [c_{i1}(\bar{x}),\dots,c_{im}(\bar{x}))]^\top = 0, \tag{3.85}$$

$$\forall s = 1,\dots,n+1,\ i = 1,\dots,p,\ \mu_{is} \ge 0,\ \mu_{is}\Big(\sum_{l=1}^m y_{ls} c_{il}(\bar{x}) - d_i(\bar{x})\Big) = 0, \tag{3.86}$$

$$\forall i = 1,\dots,p,\ \beta_i \ge 0,\ \beta_i \Big(\sum_{l=1}^m y_l c_{il}(\bar{x}) - d_i(\bar{x})\Big) = 0. \tag{3.87}$$

(ii) *Let the solution set-valued mapping S* (3.82) *be inner semicontinuous at $(\bar{x},\bar{y})$, which is the case, in particular, if $a(x) := c$, $b(x) := b$, $C(x) := C$ and $d(x) := x$, for all $x \in X$. Then there exists $(\alpha,\beta,\mu) \in \mathbb{R}^{k+2p}$ with $\|(\alpha,\beta)\| < r$ (for some $r > 0$), and $\lambda \in \mathbb{R}^*_+$ such that relationships* (3.16), (3.83) *and* (3.87) *together with the following ones hold:*

$$\nabla_x F(\bar{x},\bar{y}) + \sum_{j=1}^{k} \alpha_j \nabla G_j(\bar{x}) + \sum_{i=1}^{p} (\beta_i - \lambda \mu_i)\Big(\sum_{l=1}^{m} \bar{y}_l \nabla c_{il}(\bar{x}) - \nabla d_i(\bar{x})\Big) = 0,$$

$$\nabla a(\bar{x}) + \sum_{i=1}^{p} \mu_i [c_{i1}(\bar{x}),\dots,c_{im}(\bar{x}))]^\top = 0,$$

$$\forall i = 1,\dots,p,\ \mu_i \geq 0,\ \mu_i\Big(\sum_{l=1}^{m} y_l c_{il}(\bar{x}) - d_i(\bar{x})\Big) = 0.$$

Proof. We have from Theorem 3.2.3 that problem (3.48) is partially calm. Hence, by applying Theorem 3.2.12, we have the result. □

As second application of Theorem 3.2.12, we consider the situation where the lower-level problem is strongly stable in the sense of Kojima. For a fixed parameter $x := \bar{x}$ the vector $\bar{y} \in S(\bar{x})$ is a strongly stable local optimal solution of the parametric optimization problem (1.3), in the sense of Kojima [71], if there exist open neighborhoods U of $\bar{x}$, and V of $\bar{y}$, and a uniquely determined continuous vector-valued function $y(.)\ : U \to V$ such that $y(x)$ is the unique local optimal solution of problem (1.3) in V for all $x \in U$. This is automatical if the point $(\bar{x},\bar{y})$ is lower-level regular (3.65) and the following *strong sufficient condition of second order* (SSOC) is satisfied at the same point:

$$d^\top \nabla_y \mathscr{L}(\bar{x},\bar{y},u)d > 0,\ \ \forall u \in \Lambda(\bar{x},\bar{y}),\ \ \forall d \neq 0:\ \ \nabla_y g_\nu(\bar{x},\bar{y})d = 0. \tag{3.88}$$

In order to ensure the Lipschitz continuity of the function $y(.)$ we additionally need the *constant rank constraint qualification* (CRCQ) (cf. [105]), which holds at a point $(\bar{x},\bar{y})$ if there exists a neighborhood W of $(\bar{x},\bar{y})$ such that:

$$\nabla_y g_I(x,y)\ \text{ has the same rank }\ \forall I \subseteq \theta \cup \nu,\ \forall (x,y) \in W. \tag{3.89}$$

Recall that the expressions of function $\mathscr{L}$ and the set $\Lambda(\bar{x},\bar{y})$ used in (3.88) and (3.89) are given in (3.9) and (3.11), respectively. As for the index sets ν and θ, they are defined in (3.22).

Corollary 3.2.14 (stationarity conditions for (P) under the strong stability of the follower's problem)**.** *Let* $(\bar{x},\bar{y})$ *be a local optimal solution of problem* (3.3)*. Assume that the SSOC, CRCQ, and the lower-level regularity* (3.65) *are satisfied at* $(\bar{x},\bar{y})$*, while the upper-level regularity* (3.72) *holds at* $\bar{x}$*. Furthermore, assume that there exists* $\lambda > 0$ *such that we have*

$$f(x,y) - f(x,y(x)) \geq \lambda\ \|y - y(x)\|,\ \forall y \in K(x) \cap V(x),\ \forall x \in X, \tag{3.90}$$

where $V(x)$ *is a neighborhood of* $y(x)$ *with* $x \in X$*. Then, there exist* $\lambda > 0$ *and* $(\alpha,\beta) \in \mathbb{R}^{k+p}$ *with* $\|(\alpha,\beta)\| \leq r$ *(for some* $r > 0$*), such that relationships* (3.13) *and* (3.16)–(3.17) *together with the following condition are satisfied:*

$$0 \in \nabla_x F(\bar{x},\bar{y}) + \sum_{j=1}^{k} \alpha_j \nabla G_j(\bar{x}) + \sum_{i=1}^{p} \beta_i \nabla_x g_i(\bar{x},\bar{y}) - \bar{\partial} y(\bar{x})^\top \nabla_y f(\bar{x},\bar{y}),$$

where $\bar{\partial} y(\bar{x})$ *denotes the Clarke generalized Jacobian of the function* $y(.)$ *at the point* $\bar{x}$*.*

Proof. Under the lower-level regularity and the SSOSC, we have the LLVF reformulation (3.3) with the lower-level value function (3.2) takes the form

$$\varphi(x) := f(x,y(x)) \tag{3.91}$$

where $y(.)$ is a continuous function near $\bar{x}$. Replacing $S(x)$ in (3.54) by $y(x)$, this condition reduces to (3.90), which ensures that the corresponding reformulation of problem (3.3) is partially calm at $(\bar{x},\bar{y})$. Thus we have (3.79) with φ in (3.91). It then remains to compute the basic subdifferential of this function. Recall that in addition to the lower-level regularity and the SSOSC, the CRCQ implies the Lipschitz continuity of $y(.)$ around $\bar{x}$, cf. [105, Theorem 4.10]. Hence, we have from the chain rule in [89, Theorem 1.110] that

$$\partial\varphi(\bar{x}) = \nabla_x f(\bar{x},\bar{y}) + D^* y(\bar{x})(\nabla_y f(\bar{x},\bar{y})). \tag{3.92}$$

Applying the convex hull operator "co" to this equality, the convexified subdifferential of φ (3.91) is obtained as

$$\bar{\partial}\varphi(\bar{x}) = \nabla_x f(\bar{x},\bar{y}) + \operatorname{co} D^* y(\bar{x})(\nabla_y f(\bar{x},\bar{y})). \tag{3.93}$$

We have from [87] that the convex hull part of this formula can be represented as

$$\operatorname{co} D^* y(\bar{x})(\nabla_y f(\bar{x},\bar{y})) = \bar{\partial} y(\bar{x})^\top \nabla_y f(\bar{x},\bar{y}). \tag{3.94}$$

Combing (3.79)–(3.81) and (3.93)–(3.94), while taking into account that $\partial\varphi(\bar{x}) \subseteq \bar{\partial}\varphi(\bar{x})$, we have the result. □

For the computation of the generalized Jacobian of the function y(.), the interested reader is referred to [32]. Optimality conditions for problem (P), under strong stability, can also be derived by using the implicit function approach in (1.6). This approach is considered in the papers [20, 99]. However the optimality conditions there are completely different from those in Corollary 3.2.14.

Remark 3.2.15 (inner semicompactness, inner semicontinuity and convexity)**.** *As noted in the above results, the inner semicompactness and inner semicontinuity of the lower-level solution set-valued mapping S* (1.2) *have played a major role. For the inner semicompactness, as mentioned in Subsection 2.1.4, it is automatically satisfied if S is uniformly bounded* (2.28)*, which is a weak requirement. A practical framework where the latter is satisfied is the bilevel road pricing problem considered in Chapter 7. As for the inner semicontinuity, it is obtained if S is Lipschitz-like around the point in question. Conditions ensuring that the solution set-valued mapping of an optimization problem is Lipschitz-like are developed in Chapter 4 (also see Chapter 5) under various settings. This condition also automatically holds at $(\bar{x},\bar{y})$ provided $S(\bar{x}) = \{\bar{y}\}$ or if S is the solution set-valued mapping of a parametric linear program with additive right-hand-side perturbations. For the single-valuedness, it is satisfied if the SSOC* (3.88) *is modified to take the form: $d^\top \nabla_y \mathscr{L}(\bar{x},\bar{y},u)d > 0,\ \forall u \in \Lambda(\bar{x},\bar{y}),\ \forall d \neq 0:\ \nabla_y g_i(\bar{x},\bar{y})d \leq 0\ (g_i(\bar{x},\bar{y}) = 0)$, $\nabla_y g_i(\bar{x},\bar{y})d = 0\ (u_i > 0)$. Now observe that the estimate of the basic subdifferential of the lower-level value function φ* (3.2) *coincide when either S is inner semicontinuous or φ is convex, cf. Theorem 3.2.9 (ii). It should however be clear that both conditions are in general not related, cf. [26, Remark 3.2]. Further discussions and references on the inner semicontinuity of solution set-valued mappings are given in the latter reference.*

3.3 KKT reformulation

Consider the KKT reformulation (3.8) as a usual nonlinear optimization problem with equality and inequality constraints; in the context of the operator constraint formulation (2.40), it consists to set $\Omega := \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p$, $\psi(x,y,u) := [G(x),\ \mathscr{L}(x,y,u),\ g(x,y),\ u,\ u^\top g(x,y)]$ and $\Lambda := \mathbb{R}^k_- \times \{0_m\} \times \mathbb{R}^p_- \times \mathbb{R}^p_+ \times \{0\}$, then one can show, cf. [13, 48, 110, 130], that the basic CQ/MFCQ fails at any feasible point. In order to avoid this, one needs to consider a new constructive representation of the feasible set of problem (3.8). To motivate our discussion, we suggest an example of problem where the validity of the basic CQ is reinstated by reformulating the feasible set into an operator constraint form (2.40) with $\psi(x,y,u) := [G(x), \mathscr{L}(x,y,u)]$, $\Lambda := \mathbb{R}^k_- \times \{0_m\}$ and $\Omega := \{(x,y,u) |\, u \geq 0,\ g(x,y) \leq 0,\ u^\top g(x,y) = 0\}$.

Example 3 (validity of the basic CQ via a reformulation of the feasible set)**.** *We consider the bilevel programming problem*

$$\min_{x,y} \left\{x^2+y^2 |\, x \geq 0, y \in S(x) := \arg\min\{xy+y|\, y \geq 0\}\right\}.$$

One can easily check that $(0,0)$ *is the optimal solution of the above problem. The classical KKT reformulation of this problem is:*

$$\min_{x,y,u} \{x^2+y^2 |\, x \geq 0, x-u+1=0,\ u \geq 0, y \geq 0, uy=0\}$$

It is obvious that the lower level multiplier corresponding to the optimal solution is $\bar{u}=1$*; and hence that the MFCQ fails to hold at* $(0,0,1)$*. We are now going to show that the basic CQ is satisfied if we set* $\psi(x,y,u)=(-x,x-u+1)$*,* $\Lambda=\mathbb{R}_- \times \{0\}$ *and* $\Omega=\{(x,y,u)\in\mathbb{R}^3 | y\geq 0,\ u\geq 0,\ yu=0\}$*. For some point* $(\alpha,\beta)\in N_\Lambda(\psi(0,0,1))$*, i.e.* $(\alpha,\beta)\in\mathbb{R}_+\times\mathbb{R}$*,* $(0,0,0)\in\langle\nabla\psi(0,0,1),(\alpha,\beta)\rangle+N_\Omega(0,0,1)$ *if and only if* $\alpha-\beta=0$ *and* $(0,-\beta)\in N_\Pi(0,1)$ *(with* $\Pi:=\{(y,u)\in\mathbb{R}^2 | y\geq 0,\ u\geq 0,\ yu=0\}$*). It follows from Lemma 3.3.1 below that* $\beta=0$ *and hence that* $\alpha=0$*. This shows that the basic CQ holds at* $(0,0,1)$*.*

Proceeding as in this example will lead us to M-type necessary optimality conditions for problem (3.8), that we discuss in the next subsection. By a different representation of the feasible set in terms of the operator constraint (2.40), we will derive C-type conditions in Subsection 3.3.2. The S-type conditions will be discussed in the last subsection.

3.3.1 M-stationarity conditions

By setting $v := -g(x,y)$ and hence introducing a new (dummy) variable, the idea in the above example can be extended to the more general KKT reformulation (3.8). The technicality behind this is that the new constraint $g(x,y)+v=0$ is moved to the function ψ (in the operator constraint representation (2.40)) and thus allowing just the computation of the normal cone to the polyhedral set

$$\Pi := \{(u,v)\in\mathbb{R}^{2p} | u\geq 0,\ v\geq 0,\ u^\top v=0\}, \tag{3.95}$$

which is possible without any qualification condition, see e.g. [50, Proposition 2.1]:

Lemma 3.3.1 (formula of the normal cone to Π)**.** *Let* $(\bar{u},\bar{v})\in\Pi$*, then we have*

$$N_\Pi(\bar{u},\bar{v}) = \left\{ (u,v)\in\mathbb{R}^{2p}: \begin{array}{ll} u_i=0 & \forall i: \bar{u}_i>0=\bar{v}_i \\ v_i=0 & \forall i: \bar{u}_i=0<\bar{v}_i \\ (u_i<0\wedge v_i<0)\vee u_iv_i=0 & \forall i: \bar{u}_i=0=\bar{v}_i \end{array} \right\}. \tag{3.96}$$

Next we first exploit this formula to derive M-type necessary optimality conditions for the KKT reformulation (3.8), considered as an independent optimization problem. However, these conditions would latter be interpreted as P-M-stationarity conditions in an appropriate framework. As for the SP-M-stationarity conditions they will be deduced at the end of this subsection.

Theorem 3.3.2 (M-stationarity conditions for the KKT reformulation (3.8), I)**.** *Let* $(\bar{x},\bar{y},\bar{u})$ *be a local optimal solution of problem* (3.8) *and assume that the following CQ holds at* $(\bar{x},\bar{y},\bar{u})$*:*

$$\left.\begin{array}{r} \sum_{j=1}^k \alpha_j\nabla G_j(\bar{x})+\sum_{i=1}^p \beta_i\nabla_x g_i(\bar{x},\bar{y})+\sum_{l=1}^m \gamma_l\nabla_x\mathscr{L}_l(\bar{x},\bar{y},\bar{u})=0 \\ \sum_{i=1}^p \beta_i\nabla_y g_i(\bar{x},\bar{y})+\sum_{l=1}^m \gamma_l\nabla_y\mathscr{L}_l(\bar{x},\bar{y},\bar{u})=0 \\ \forall j=1,\ldots,k,\ \alpha_j\geq 0,\ \alpha_jG_j(\bar{x})=0 \\ \beta_\nu=0,\ \nabla_y g_\eta(\bar{x},\bar{y})\gamma=0 \\ \forall i\in\theta,\ \left(\beta_i>0\wedge\sum_{l=1}^m\gamma_l\nabla_{y_l}g_i(\bar{x},\bar{y})>0\right)\vee\beta_i\sum_{l=1}^m\gamma_l\nabla_{y_l}g_i(\bar{x},\bar{y})=0 \end{array}\right\} \Longrightarrow \left\{\begin{array}{l} \alpha=0, \\ \beta=0, \\ \gamma=0. \end{array}\right. \tag{3.97}$$

Then, the M-stationarity conditions (3.16) *and* (3.23)–(3.26) *hold with* $\|(\alpha,\beta,\gamma)\|\leq r$*, for some* $r>0$*.*

Proof. Let us set $\psi(x,y,u,v) := [G(x),\ g(x,y)+v,\ \mathscr{L}(x,y,u)]$, $\Lambda := \mathbb{R}^k_- \times \{0_{p+m}\}$ and $\Omega := \mathbb{R}^{n+m} \times \Pi$. One can easily verify that there is a vector $\bar{v}$ such that $(\bar{x},\bar{y},\bar{u},\bar{v})$ is a local optimal solution of the problem

$$\min_{x,y,u,v} \{F(x,y)|\ (x,y,u,v) \in \Omega \cap \psi^{-1}(\Lambda)\}. \tag{3.98}$$

The normal cone to Ω, that to Λ, and the gradient of ψ can respectively be obtained as

$$N_\Omega(\bar{x},\bar{y},\bar{u},\bar{v}) = \{0_{n+m}\} \times N_\Pi(\bar{u},\bar{v}), \tag{3.99}$$

$$N_\Lambda(\psi(\bar{x},\bar{y},\bar{u},\bar{v})) = \{(\alpha,\beta,\gamma)|\ \alpha_j \geq 0,\ \alpha_j G_j(\bar{x}) = 0,\ j = 1,\dots,k\}, \tag{3.100}$$

$$\nabla\psi(\bar{x},\bar{y},\bar{u},\bar{v})^\top(\alpha,\beta,\gamma) = \begin{bmatrix} \triangle(\alpha,\beta,\gamma) \\ \beta \end{bmatrix}, \tag{3.101}$$

with the matrix $\triangle(\alpha,\beta,\gamma)$ in (3.101) defined by

$$\triangle(\alpha,\beta,\gamma) := \begin{bmatrix} \sum_{j=1}^k \alpha_j \nabla G_j(\bar{x}) + \sum_{i=1}^p \beta_i \nabla_x g_i(\bar{x},\bar{y}) + \sum_{l=1}^m \gamma_l \nabla_x \mathscr{L}_l(\bar{x},\bar{y},\bar{u}) \\ \sum_{i=1}^p \beta_i \nabla_y g_i(\bar{x},\bar{y}) + \sum_{l=1}^m \gamma_l \nabla_y \mathscr{L}_l(\bar{x},\bar{y},\bar{u}) \\ \nabla_y g(\bar{x},\bar{y})\gamma \end{bmatrix} \tag{3.102}$$

with the expression of $\nabla_y g(\bar{x},\bar{y})\gamma$ given in (3.27). It follows from equalities (3.99)–(3.102) that the basic CQ applied to problem (3.98) at $(\bar{x},\bar{y},\bar{u},\bar{v})$ can equivalently be formulated as: there is no nonzero vector $(\alpha,\beta,\gamma) \in \mathbb{R}^{k+p+m}$ such that we have

$$\sum_{j=1}^k \alpha_j \nabla G_j(\bar{x}) + \sum_{i=1}^p \beta_i \nabla_x g_i(\bar{x},\bar{y}) + \sum_{l=1}^m \gamma_l \nabla_x \mathscr{L}_l(\bar{x},\bar{y},\bar{u}) = 0, \tag{3.103}$$

$$\sum_{i=1}^p \beta_i \nabla_y g_i(\bar{x},\bar{y}) + \sum_{l=1}^m \gamma_l \nabla_y \mathscr{L}_l(\bar{x},\bar{y},\bar{u}) = 0, \tag{3.104}$$

$$(-\nabla_y g(\bar{x},\bar{y})\gamma, -\beta) \in N_\Pi(\bar{u},\bar{v}). \tag{3.105}$$

By noting that $\bar{v}_i = -g_i(\bar{x},\bar{y})$, for $i := 1,\dots,p$, it follows from Lemma 3.3.1 that the basic CQ applied to problem (3.98) is equivalent to CQ (3.97). Hence, from Proposition 2.2.5 there exists $(\alpha,\beta,\gamma) \in \mathbb{R}^{k+p+m}$, with $\|(\alpha,\beta,\gamma)\| \leq r$ (for some $r > 0$) such that (3.16) and (3.23)–(3.24) together with (3.105) are satisfied, taking into account that the objective function of problem (3.98) is independent of the couple (u,v). The result then follows by interpreting inclusion (3.105) via equation (3.96). □

The technique used in the proof of Theorem 3.3.2, i.e. to transform the nonlinear complementarity system in (3.10) into a linear one, has been used in various occasions, for the MPCC; see e.g. [126, 50]. One can easily check that the M-stationarity conditions obtained here are identical to those in [127] or [124] under various CQs, among which CQ (b) of [127, Theorem 4.1] or (b) of [124, Theorem 5.1] coincides with the CQ in Theorem 3.3.2. But, it should be mentioned that in the latter case, this CQ is recovered from a perspective different from that of [127, 124], where an enhanced generalized equation formulation of the KKT conditions of the lower-level problem was used to design the CQ.

We now introduce a different way to choose ψ, Ω and Λ; that would lead to a new and weaker CQ allowing us to obtain the same optimality conditions as in Theorem 3.3.2. To proceed, let us recall that the complementarity system (3.10) is equivalent to

$$u_i \geq 0,\ g_i(x,y) \leq 0,\ u_i g_i(x,y) = 0,\ i = 1,\dots,p,$$

meaning that it can be converted into an inclusion of the form

$$(u_i, -g_i(x,y)) \in \Lambda_i := \{(a,b) \in \mathbb{R}^2 | a \geq 0,\ b \geq 0,\ ab = 0\},\ i = 1,\dots,p. \tag{3.106}$$

Theorem 3.3.3 (M-stationarity conditions for the KKT reformulation (3.8), II)**.** *Let* $(\bar{x},\bar{y},\bar{u})$ *be a local optimal solution of problem* (3.8) *and assume that the following assertions are satisfied:*

$$\text{The map } \mathscr{M}_1(v_1,v_2) := \{(x,y,u)|\, G(x)+v_1 \le 0, \mathscr{L}(x,y,u)+v_2 = 0\} \text{ is calm at } (0,0,\bar{x},\bar{y},\bar{u}), \tag{3.107}$$

$$\left.\begin{array}{r} \sum_{j=1}^{k} \alpha_j \nabla G_j(\bar{x}) + \sum_{i=1}^{p} \beta_i \nabla_x g_i(\bar{x},\bar{y}) + \sum_{l=1}^{m} \gamma_l \nabla_x \mathscr{L}_l(\bar{x},\bar{y},\bar{u}) = 0 \\ \sum_{i=1}^{p} \beta_i \nabla_y g_i(\bar{x},\bar{y}) + \sum_{l=1}^{m} \gamma_l \nabla_y \mathscr{L}_l(\bar{x},\bar{y},\bar{u}) = 0 \\ \forall j = 1,\dots,k,\ \alpha_j \ge 0,\ \alpha_j G_j(\bar{x}) = 0 \\ \beta_\nu = 0,\ \nabla_y g_\eta(\bar{x},\bar{y})\gamma = 0 \\ \forall i \in \theta,\ \big(\beta_i > 0 \wedge \sum_{l=1}^{m} \gamma_l \nabla_{y_l} g_i(\bar{x},\bar{y}) > 0\big) \vee \beta_i \sum_{l=1}^{m} \gamma_l \nabla_{y_l} g_i(\bar{x},\bar{y}) = 0 \end{array}\right\} \Longrightarrow \left\{ \begin{array}{l} \beta = 0, \\ \nabla_y g(\bar{x},\bar{y})\gamma = 0. \end{array}\right. \tag{3.108}$$

Then, the M-stationarity conditions (3.16) *and* (3.23)–(3.26) *hold with* $\|\beta\| \le r$*, for some* $r > 0$.

Proof. We consider $\Omega := \{(x,y,u)|\, G(x) \le 0, \mathscr{L}(x,y,u) = 0\}$ and $\psi(x,y,u) = (u_i, -g_i(x,y))_{i=1,\dots,p}$. If $(\bar{x},\bar{y},\bar{u})$ is a local optimal solution of problem (3.8), it means, in other words that $(\bar{x},\bar{y},\bar{u})$ is a local optimal solution of the problem

$$\min_{x,y,u} \{F(x,y)|\ (x,y,u) \in \Omega \cap \psi^{-1}(\Lambda)\}, \tag{3.109}$$

where $\Lambda = \Lambda_1 \times \ldots \times \Lambda_p$, with $\Lambda_i = \{(a,b) \in \mathbb{R}^2 | a \ge 0, b \ge 0, ab = 0\}$ for $i = 1,\dots,p$. Applying Proposition 2.2.5 to (3.109), there exists a vector $(\delta,\beta) \in \mathbb{R}^{2p}$ with $\|(\delta,\beta)\| \le r$ (for some $r > 0$) such that we have

$$\forall i = 1,\dots,p,\ (\delta_i,\beta_i) \in N_{\Lambda_i}(\psi_i(\bar{x},\bar{y},\bar{u})), \tag{3.110}$$

$$(0,0) \in \begin{bmatrix} \nabla F(\bar{x},\bar{y}) \\ 0 \end{bmatrix} + \begin{bmatrix} -\sum_{i=1}^{p} \beta_i \nabla g_i(\bar{x},\bar{y}) \\ \delta \end{bmatrix} + N_\Omega(\bar{x},\bar{y},\bar{u}), \tag{3.111}$$

provided there is no nonzero vector $(\delta,\beta) \in \mathbb{R}^{2p}$ satisfying condition (3.110) together with inclusion

$$(0,0) \in \begin{bmatrix} -\sum_{i=1}^{p} \beta_i \nabla g_i(\bar{x},\bar{y}) \\ \delta \end{bmatrix} + N_\Omega(\bar{x},\bar{y},\bar{u}). \tag{3.112}$$

It follows, under assumption (3.107) (see Theorem 2.2.7), that we have

$$N_\Omega(\bar{x},\bar{y},\bar{u}) = \Big\{ \triangle(\alpha,\beta,\gamma) - \big[\sum_{i=1}^{p} \beta_i \nabla g_i(\bar{x},\bar{y}),\ 0\big]^\top \big|\ \alpha_j \ge 0,\ \alpha_j G_j(\bar{x}) = 0,\ j = 1,\dots,k \Big\},$$

where $\triangle(\alpha,\beta,\gamma)$ denotes the matrix given in (3.102). Hence, either from (3.111) or from (3.112), it follows that there exists $\gamma \in \mathbb{R}^m$ such that $\delta = -\nabla_y g(\bar{x},\bar{y})\gamma$; a fortiori, (3.110) implies that there exists $\gamma \in \mathbb{R}^m$ such that $(-\nabla_y g_i(\bar{x},\bar{y})\gamma, \beta_i) \in N_{\Lambda_i}(\psi_i(\bar{x},\bar{y},\bar{u}))$, $i := 1,\dots,p$. The result then follows by taking into account the value of $N_{\Lambda_i}(\psi_i(\bar{x},\bar{y},\bar{u}))$ in Lemma 3.3.1. □

The approach in the above result is similar to the one used in [66], for a mathematical program with vanishing constraints.

Remark 3.3.4 (links between Theorem 3.3.2 and Theorem 3.3.3)**.** *The following assertions hold:*

(i) *The qualification conditions* (3.107) *and* (3.108) *in Theorem 3.3.3 are satisfied, provided CQ* (3.97) *holds at* $(\bar{x},\bar{y},\bar{u})$. *In fact, it is obvious that CQ* (3.97) *implies CQ* (3.108). *On the other hand, CQ* (3.97) *can equivalently be written as*

$$\{(0,0,0)\} = \{(\alpha,\beta,\gamma)| \begin{array}{l} \sum_{j=1}^{k} \alpha_j \nabla G_j(\bar{x}) + \sum_{i=1}^{p} \beta_i \nabla_x g_i(\bar{x},\bar{y}) + \sum_{l=1}^{m} \gamma_l \nabla_x \mathscr{L}_l(\bar{x},\bar{y},\bar{u}) = 0, \\ \sum_{i=1}^{p} \beta_i \nabla_y g_i(\bar{x},\bar{y}) + \sum_{l=1}^{m} \gamma_l \nabla_y \mathscr{L}_l(\bar{x},\bar{y},\bar{u}) = 0, \\ \forall j = 1,\dots,k,\ \alpha_j \ge 0,\ \alpha_j G_j(\bar{x}) = 0, \\ \beta_\nu = 0,\ \nabla_y g_\eta(\bar{x},\bar{y})\gamma = 0, \\ \forall i \in \theta,\ \big(\beta_i > 0 \wedge \sum_{l=1}^{m} \gamma_l \nabla_{y_l} g_i(\bar{x},\bar{y}) > 0\big) \vee \beta_i \sum_{l=1}^{m} \gamma_l \nabla_{y_l} g_i(\bar{x},\bar{y}) = 0\} \end{array}$$

Furthermore, the set in the right-hand-side of the previous equality contains the following set

$$A(\bar{x},\bar{y},\bar{u}) := \{(\alpha,0,\gamma)|\ \ \begin{array}{l}\sum_{j=1}^{k}\alpha_j\nabla G_j(\bar{x})+\sum_{l=1}^{m}\gamma_l\nabla_x\mathscr{L}_l(\bar{x},\bar{y},\bar{u})=0,\\ \sum_{l=1}^{m}\gamma_l\nabla_y\mathscr{L}_l(\bar{x},\bar{y},\bar{u})=0,\ \ \nabla_y g(\bar{x},\bar{y})\gamma=0,\\ \forall j=1,\dots,k,\ \alpha_j\geq 0,\ \alpha_j G_j(\bar{x})=0\},\end{array}$$

which means that CQ (3.97) *is also a sufficient condition for* $A(\bar{x},\bar{y},\bar{u}) = \{(0,0,0)\}$. *Following Lemma 2.2.3* (*cf.* (2.44))*, the latter equality implies the fulfilment of CQ* (3.107)*.*

(ii) *A second possibility to recover CQ* (3.97) *in the above theorem is to move the constraints defining* Ω *to the function* ψ*, i.e. to set* $\Omega := \mathbb{R}^n\times\mathbb{R}^m$ *and* $\psi(x,y,u) := [G(x),\mathscr{L}(x,y,u),(u_i,-g_i(x,y))_{i=1,\dots,p}]$*. This would also help recover the bound on all the multipliers.*

To conclude this subsection, we now deduce the M-type optimality conditions for the bilevel optimization (3.48) from the above developments on the classical KKT reformulation (3.8).

Corollary 3.3.5 (SP-M-stationarity via the KKT reformulation)**.** *Let* $(\bar{x},\bar{y})$ *be a local optimal solution of problem (P), where the lower-level problem* (1.3) *is convex. Assume that the Slater CQ* (3.6) *holds at* $\bar{x}$ *while for all* $\bar{u}\in\Lambda(\bar{x},\bar{y})$*, CQ* (3.97) (*resp. CQs* (3.107)–(3.108)) *holds at* $(\bar{x},\bar{y},\bar{u})$*. Then* $(\bar{x},\bar{y})$ *is SP-M-stationary.*

Proof. Follows from the combination of Theorem 3.1.3 (i) and Theorem 3.3.2 (resp. Theorem 3.3.3). □

Observe that if CQ (3.97) or the combination of CQ (3.107) and (3.108) is satisfied at the point $(\bar{x},\bar{y},\bar{u})$, only for a single $\bar{u}\in\Lambda(\bar{x},\bar{y})$, we obtain the P-M-stationarity for $(\bar{x},\bar{y})$. The latter will also be derived in Section 3.4 via the OPEC reformulation (3.5), with the difference that the lower-level multiplier $\bar{u}$ is not known a priori, as it is the case here via the KKT reformulation (3.8).

3.3.2 C-stationarity conditions

We proceed here as in the previous subsection, i.e. we first provide C-stationarity conditions for (3.8), and then deduce the SP-C-stationarity for problem (P).

Theorem 3.3.6 (C-stationarity conditions for the KKT reformulation (3.8))**.** *Let* $(\bar{x},\bar{y},\bar{u})$ *be a local optimal solution of problem* (3.8) *and assume that the following CQ holds at* $(\bar{x},\bar{y},\bar{u})$*:*

$$\left.\begin{array}{r}\sum_{j=1}^{k}\alpha_j\nabla G_j(\bar{x})+\sum_{i=1}^{p}\beta_i\nabla_x g_i(\bar{x},\bar{y})+\sum_{l=1}^{m}\gamma_l\nabla_x\mathscr{L}_l(\bar{x},\bar{y},\bar{u})=0\\ \sum_{i=1}^{p}\beta_i\nabla_y g_i(\bar{x},\bar{y})+\sum_{l=1}^{m}\gamma_l\nabla_y\mathscr{L}_l(\bar{x},\bar{y},\bar{u})=0\\ \forall j=1,\dots,k,\ \alpha_j\geq 0,\ \alpha_j G_j(\bar{x})=0\\ \beta_\nu=0,\ \nabla_y g_\eta(\bar{x},\bar{y})\gamma=0\\ \forall i\in\theta,\ \beta_i\sum_{l=1}^{m}\gamma_l\nabla_{y_l}g_i(\bar{x},\bar{y})\geq 0\end{array}\right\}\Longrightarrow\left\{\begin{array}{l}\alpha=0,\\ \beta=0,\\ \gamma=0.\end{array}\right. \tag{3.113}$$

Then, the C-stationarity conditions (3.16) *and* (3.23)–(3.26) *hold with* $\|(\alpha,\beta,\gamma)\|\leq r$ *for some* $r>0$*.*

Proof. Following the work by Scheel and Scholtes [110], the KKT reformulation (3.8) can take the operator constraint form (2.40) with $\Omega := \mathbb{R}^n\times\mathbb{R}^m\times\mathbb{R}^p$, $\psi(x,y,u) := [G(x),\ V(x,y,u),\ \mathscr{L}(x,y,u)]$ and $\Lambda := \mathbb{R}^k_-\times\{0_{m+p}\}$. Here,

$$V_i(x,y,u) := \min\{u_i,\ -g_i(x,y)\}\ \text{ for }\ i=1,\dots,p. \tag{3.114}$$

Applying Proposition 2.2.5 to the corresponding operator constraint reformulation of (3.8), there exists (α,β,γ) with $\|(\alpha,\beta,\gamma)\|\leq r$ for some $r>0$ such that we have:

$$(\alpha,\beta,\gamma)\in N_\Lambda(\psi(\bar{x},\bar{y},\bar{u})), \tag{3.115}$$

$$0\in\begin{bmatrix}\nabla F(\bar{x},\bar{y})\\ 0\end{bmatrix}+\partial\langle(\alpha,\beta,\gamma),\psi\rangle(\bar{x},\bar{y},\bar{u}), \tag{3.116}$$

provided the following implication is satisfied at the point $(\bar{x},\bar{y},\bar{u})$:

$$\Big[0 \in \partial\langle(\alpha,\beta,\gamma),\psi\rangle(\bar{x},\bar{y},\bar{u}),\ (\alpha,\beta,\gamma) \in N_\Lambda(\psi(\bar{x},\bar{y},\bar{u}))\Big] \Longrightarrow \alpha = 0, \beta = 0, \gamma = 0. \tag{3.117}$$

Now observe that $\partial\langle(\alpha,\beta,\gamma),\psi\rangle(\bar{x},\bar{y},\bar{u})$ can be written in terms of $\triangle(\alpha,\beta,\gamma)$ (3.102) as follows

$$\partial\langle(\alpha,\beta,\gamma),\psi\rangle(\bar{x},\bar{y},\bar{u}) = \triangle(\alpha,\beta,\gamma) - \left[\begin{array}{c} \sum_{i=1}^{p}\beta_i\nabla g_i(\bar{x},\bar{y}) \\ 0 \end{array}\right] + \partial\langle\beta,V\rangle(\bar{x},\bar{y},\bar{u}), \tag{3.118}$$

while an upper estimate of $\partial\langle\beta,V\rangle(\bar{x},\bar{y},\bar{u})$ can be obtained as

$$\begin{array}{ll} \partial\langle\beta,V\rangle(\bar{x},\bar{y},\bar{u}) & \subseteq \sum_{i=1}^{p}\beta_i\bar{\partial}V_i(\bar{x},\bar{y},\bar{u}) \subseteq \left\{ \left[\begin{array}{c} -\sum_{i=1}^{p}\xi_i\nabla g_i(\bar{x},\bar{y}) \\ \zeta \end{array}\right] \middle| \begin{array}{l} \xi_\eta = 0,\ \xi_\nu = \beta_\nu,\ \zeta_\eta = \beta_\eta,\ \zeta_\nu = 0, \\ \forall i \in \theta,\ \exists t_i \in [0,1]:\ \xi_i = \beta_i(1-t_i),\ \zeta_i = \beta_i t_i \end{array} \right\} \\ & \subseteq \left\{ \left[\begin{array}{c} -\sum_{i=1}^{p}\beta_i\nabla g_i(\bar{x},\bar{y}) \\ \zeta \end{array}\right] \middle| \begin{array}{l} \beta_\eta = 0,\ \zeta_\nu = 0, \\ \forall i \in \theta,\ \beta_i\zeta_i \geq 0 \end{array} \right\} \end{array} \tag{3.119}$$

More details on this estimation can be found in the proof of Theorem 4.1.1 in Chapter 4. Combining (3.118) and (3.119), one can easily check that CQ (3.113) implies the fulfilment of condition (3.117). The C-stationarity conditions (3.16) and (3.23)–(3.26) are obtained by successively inserting (3.118) and (3.119) in inclusion (3.116). It should be noted that the formula of $N_\Lambda(\psi(\bar{x},\bar{y},\bar{u}))$ here is identical to that in (3.100). Given that $\forall i \in \theta,\ \xi_i = \beta_i(1-t_i)$ (for some $t_i \in [0,1]$) in (3.119), we effectively maintain the same bound on the multipliers: $\|(\alpha,\beta,\gamma)\| \leq r$. □

Although the C-stationarity conditions are weaker than the M-ones, cf. diagram in (3.31), it easy to check that CQ (3.113) used to obtain the former conditions is stronger than CQ (3.97) under which the latter stationarity conditions are obtained in Theorem 3.3.2. Next we deduce the SP-C-stationarity from Theorem 3.3.6.

Corollary 3.3.7 (SP-C-stationarity via the KKT reformulation)**.** *Let* $(\bar{x},\bar{y})$ *be a local optimal solution of the bilevel program (P), where the lower-level problem* (1.3) *is convex. Assume that the Slater CQ* (3.6) *holds at* $\bar{x}$ *while for all lower-level multipliers* $\bar{u} \in \Lambda(\bar{x},\bar{y})$*, CQ* (3.113) *holds at* $(\bar{x},\bar{y},\bar{u})$*. Then* $(\bar{x},\bar{y})$ *is SP-C-stationary.*

Proof. Follows from the combination of Theorem 3.1.3 (i) and Theorem 3.3.6. □

Similarly to the previous subsection, if CQ (3.113) is satisfied at the point $(\bar{x},\bar{y},\bar{u})$, only for a single $\bar{u} \in \Lambda(\bar{x},\bar{y})$, we obtain the P-C-stationarity for $(\bar{x},\bar{y})$.

3.3.3 S-stationarity conditions

In the framework of MPCCs, the Guignard CQ has been shown to be one of the few CQs to be directly applicable to (3.8) considered as a usual nonlinear optimization problem with equality and inequality constraints, cf. [48]. In the next result, we derive the S-type stationarity conditions tailored to problem (3.8) via the Guignard CQ. For an optimization problem

$$\min\ \{f(x)|\ g(x) \leq 0, h(x) = 0\}, \tag{3.120}$$

where the functions $f : \mathbb{R}^n \to \mathbb{R}$, $g : \mathbb{R}^n \to \mathbb{R}^p$ and $h : \mathbb{R}^n \to \mathbb{R}^q$ are continuously differentiable, if we denote by C the feasible set of problem (3.120), the Guignard CQ is satisfied at $\bar{x} \in C$, provided that

$$L_C(\bar{x})^* = -\widehat{N}_C(\bar{x}). \tag{3.121}$$

The set in the left-hand-side of the equality denotes the dual cone of the linearized tangent cone to C at $\bar{x} \in C$ which is defined by

$$L_C(\bar{x}) := \{d \in \mathbb{R}^n |\, \nabla g_i(\bar{x})^\top d \leq 0 \ \forall i: \ g_i(\bar{x}) = 0, \ \nabla h_i(\bar{x})^\top d = 0 \ \forall i: \ i = 1, \ldots, q\}.$$

Theorem 3.3.8 (S-stationarity conditions for the KKT reformulation (3.8) via the Guignard CQ)**.** *Let* $(\bar{x}, \bar{y}, \bar{u})$ *be a local optimal solution of problem* (3.8) *and assume that the corresponding Guignard CQ* (3.121) *holds at* $(\bar{x}, \bar{y}, \bar{u})$*. Then the S-stationarity conditions* (3.16), (3.23)–(3.25) *and* (3.30) *hold with* $\|(\alpha, \gamma)\| \leq r$ *for some number* $r > 0$.

Proof. First start by observing that in the complementarity system (3.10) the equality $u^\top g(x,y) = 0$ is equivalent to $-u^\top g(x,y) \leq 0$, given that $u \geq 0$ and $g(x,y) \leq 0$; hence the feasible set C of problem (3.8) can take the form

$$C := \{(x,y,u) |\, G(x) \leq 0, \ g(x,y) \leq 0, \ \mathscr{L}(x,y,u) = 0, \ -u \leq 0, \ -u^\top g(x,y) \leq 0\}. \tag{3.122}$$

Furthermore, $(\bar{x}, \bar{y}, \bar{u})$ being a local optimal solution of problem (3.8), we have

$$\nabla_x F(\bar{x}, \bar{y})^\top d + \nabla_y F(\bar{x}, \bar{y})^\top v \geq 0, \ \forall (d,v,w) \in T_C(\bar{x}, \bar{y}, \bar{u}), \tag{3.123}$$

where T_C denotes the Bouligand tangent cone (2.17). Thus considering (2.18), it follows from (3.123) that we have

$$-\nabla_{x,y,u} F(\bar{x}, \bar{y}) \in (T_C(\bar{x}, \bar{y}, \bar{u}))^o = \widehat{N}_C(\bar{x}, \bar{y}, \bar{u}).$$

Under the Guignard CQ (3.121), applied to our problem (3.8) with C in (3.122), it follows that

$$\nabla_x F(\bar{x}, \bar{y})^\top d + \nabla_y F(\bar{x}, \bar{y})^\top v \geq 0, \ \forall (d,v,w) \in L_C(\bar{x}, \bar{y}, \bar{u}).$$

Clearly, this implies that the triple $(0,0,0)$ is an optimal solution of the optimization problem

$$\min_{d,v,w} \{\nabla_x F(\bar{x}, \bar{y})^\top d + \nabla_y F(\bar{x}, \bar{y})^\top v |\, \psi(d,v,w) \in \Lambda\}, \tag{3.124}$$

where the function ψ and the set Λ are respectively defined by

$$\psi(d,v,w) := \begin{bmatrix} \nabla G_j(\bar{x})^\top d, \ j \in J \\ \nabla g_i(\bar{x}, \bar{y})^\top (d,v), \ i \in \theta \cup \nu \\ \nabla \mathscr{L}_l(\bar{x}, \bar{y}, \bar{u})^\top (d,v,w), \ l = 1, \ldots, m \\ -\nabla \chi(\bar{x}, \bar{y}, \bar{u})^\top (d,v,w) \\ -w_i \leq 0, \ i \in \eta \cup \theta \end{bmatrix} \ \text{and} \ \Lambda := \mathbb{R}_-^{|J|} \times \mathbb{R}_-^{|\theta \cup \nu|} \times \{0_m\} \times \mathbb{R}_- \times \mathbb{R}_-^{|\eta \cup \theta|}$$

with $\chi(x,y,u) := \sum_{i=1}^p u_i g_i(x,y)$ and $J := \{j |\, G_j(\bar{x}) = 0\}$, while η, θ and ν are given in (3.22). Now observe that the set-valued mapping

$$\Psi(\vartheta) := \{(d,v,w) |\, \psi(d,v,w) + \vartheta \in \Lambda\}$$

is calm at $(0,0,0,0)$ given that it is polyhedral, cf. [106, Proposition 1]. Hence, applying Proposition 2.2.5 to problem (3.124), it follows that there exists $(\alpha, \beta, \gamma, \lambda, \xi) \in N_\Lambda(\psi(0,0,0))$ with $\|(\alpha, \beta, \gamma, \lambda, \xi)\| \leq r$ for some $r > 0$ such that we have

$$\begin{bmatrix} \nabla F(\bar{x}, \bar{y}) \\ 0 \end{bmatrix} + \nabla \psi(0,0,0)^\top (\alpha, \beta, \gamma, \lambda, \xi) = 0. \tag{3.125}$$

For all $(\alpha, \beta, \gamma, \lambda, \xi)$ in the normal cone $N_\Lambda(\psi(0,0,0))$, which can be written as

$$N_\Lambda(\psi(0,0,0)) = \{(\alpha, \beta, \gamma, \lambda, \xi) |\, \alpha_J \geq 0, \ \beta_{\theta \cup \nu} \geq 0, \ \lambda \geq 0, \ \xi_{\eta \cup \theta} \geq 0\}, \tag{3.126}$$

we have the following expression for the term $\nabla\psi(0,0,0)^\top(\alpha,\beta,\gamma,\lambda,\xi)$:

$$\nabla\psi(0,0,0)^\top(\alpha,\beta,\gamma,\lambda,\xi) = \begin{bmatrix} \sum\limits_{j\in J}\alpha_j\nabla G_j(\bar{x}) + \sum\limits_{i\in\theta\cup\nu}\beta_i\nabla_x g_i(\bar{x},\bar{y}) - \lambda\sum_{i=1}^{p}\bar{u}_i\nabla_x g_i(\bar{x},\bar{y}) + \sum_{l=1}^{m}\gamma_l\nabla_x\mathscr{L}(\bar{x},\bar{y},\bar{u}) \\ \sum\limits_{i\in\theta\cup\nu}\beta_i\nabla_y g_i(\bar{x},\bar{y}) - \lambda\sum_{i=1}^{p}\bar{u}_i\nabla_y g_i(\bar{x},\bar{y}) + \sum_{l=1}^{m}\gamma_l\nabla_y\mathscr{L}(\bar{x},\bar{y},\bar{u}) \\ \nabla_y g(\bar{x},\bar{y})\gamma - \lambda g(\bar{x},\bar{y}) - \varsigma \end{bmatrix} \tag{3.127}$$

where $\varsigma_{\eta\cup\theta} = \xi_{\eta\cup\theta}$ and $\varsigma_i = 0$ for all $i\in\nu$. If we also set $\alpha_j = 0$ for all $j\in J^c := \{1,\dots,k\}\setminus J$ and $\beta_i = 0$ for all $i\in\eta$, then combining (3.125)–(3.127) we obtain the relationships (3.16)–(3.17) and (3.33)–(3.37), while noting that conditions (3.35) and (3.36) are due to the feasibility of the point $(\bar{x},\bar{y},\bar{u})$. Hence by the proof of Theorem 3.1.9, we have the S-stationarity conditions (3.16), (3.23)–(3.25) and (3.30). For the bound on the multipliers, observe from the proof of the equivalence of the latter stationarity conditions with those in (3.16)–(3.17) and (3.33)–(3.37) (see Theorem 3.1.9) that the values of the multipliers α and γ remain unchanged in both directions. Thus we have $\|(\alpha,\gamma)\| \leq r$. □

We now provide an example of bilevel programming problem where the Guignard CQ is satisfied for the corresponding KKT reformulation.

Example 4 (validity of the Guignard CQ in bilevel programming)**.** *We consider the bilevel program*

$$\min_{x,y}\left\{x+y \mid x\geq 0,\ y\in S(x) := \arg\min\{xy \mid y\geq 0\}\right\}.$$

One can easily check that the KKT reformulation (3.8) *of this problem reduces to the mathematical program with complementarity constraints* $\min\limits_{x,y}\{x+y \mid x,y\geq 0,\ xy=0\}$ *and following [48], the Guignard CQ holds at the unique optimal solution point* $(0,0)$.

For some more details on the application of the Guignard CQ to more general frameworks of MPCCs, we refer the interested reader to [48, 49] and references therein. Further note that there are two other usual techniques to derive the S-type stationarity conditions in the MPCC theory that can be translated to our KKT reformulation (3.8). The first one is the use of the so-called MPEC-LICQ, which reduces in our case to the following CQ:

$$\left.\begin{array}{r} \sum_{j=1}^{k}\alpha_j\nabla G_j(\bar{x}) + \sum_{l=1}^{m}\gamma_l\nabla_x\mathscr{L}_l(\bar{x},\bar{y},\bar{u}) + \sum_{i=1}^{p}\beta_i\nabla_x g_i(\bar{x},\bar{y}) = 0 \\ \sum_{l=1}^{m}\gamma_l\nabla_y\mathscr{L}_l(\bar{x},\bar{y},\bar{u}) + \sum_{i=1}^{p}\beta_i\nabla_y g_i(\bar{x},\bar{y}) = 0 \\ \alpha_{J^c} = 0,\ \nabla_y g_\nu(\bar{x},\bar{y})\gamma = 0,\ \beta_\eta = 0 \end{array}\right\} \Longrightarrow \left\{\begin{array}{l} \alpha = 0, \\ \beta = 0, \\ \gamma = 0, \end{array}\right. \tag{3.128}$$

where $J^c := \{j \mid G_j(\bar{x}) < 0\}$. This approach was introduced by Scheel and Scholtes [110] where they proceed by defining an auxiliary problem to the corresponding MPCC problem. Ye [123] later introduced a weaker form of CQ (3.128) that she labeled as Partial MPEC-LICQ which can take the form

$$\left.\begin{array}{r} \sum_{j=1}^{k}\alpha_j\nabla G_j(\bar{x}) + \sum_{l=1}^{m}\gamma_l\nabla_x\mathscr{L}_l(\bar{x},\bar{y},\bar{u}) + \sum_{i=1}^{p}\beta_i\nabla_x g_i(\bar{x},\bar{y}) = 0 \\ \sum_{l=1}^{m}\gamma_l\nabla_y\mathscr{L}_l(\bar{x},\bar{y},\bar{u}) + \sum_{i=1}^{p}\beta_i\nabla_y g_i(\bar{x},\bar{y}) = 0 \\ \alpha_{J^c} = 0,\ \nabla_y g_\nu(\bar{x},\bar{y})\gamma = 0,\ \beta_\eta = 0 \end{array}\right\} \Longrightarrow \left\{\begin{array}{l} \beta_\theta = 0, \\ \nabla_y g_\theta(\bar{x},\bar{y})\gamma = 0, \end{array}\right. \tag{3.129}$$

for our problem (3.8). However, to apply the latter CQ, one should already have other weaker optimality conditions, like the M or C-ones derived above. We can summarize these two approaches in the following theorem:

Theorem 3.3.9 (S-stationarity conditions for the KKT reformulation (3.8) via the MPEC-LICQ and Partial MPEC-LICQ)**.** *Let* $(\bar{x},\bar{y},\bar{u})$ *be a local optimal solution of problem* (3.8) *and assume that* EITHER *CQ* (3.128) OR [*CQ* (3.97) $\wedge$ *CQ* (3.129)] *holds at* $(\bar{x},\bar{y},\bar{u})$. *Then the S-stationarity conditions* (3.16)*,* (3.23)–(3.25) *and* (3.30) *are satisfied.*

We now introduce a new approach to obtain the S-stationarity conditions for the KKT reformulation (3.8). This approach is based on a notion of partial calmness for the latter problem in the pattern of that of the LLVF reformulation (3.3) in the vein of its characterization in Theorem 3.2.2.

Definition 3.3.10 (partial calmness concept for the KKT reformulation (3.8))**.** *Let $(\bar{x},\bar{y},\bar{u})$ be a local optimal solution of problem* (3.8). *This problem is partially calm at $(\bar{x},\bar{y},\bar{u})$ if there exists a number $\lambda > 0$ such that $(\bar{x},\bar{y},\bar{u})$ is a local optimal solution of the partially penalized problem*

$$\min_{x,y,u} \{F(x,y) - \lambda \sum_{i=1}^{p} u_i g_i(x,y)|\ x \in X,\ \mathscr{L}(x,y,u) = 0,\ g(x,y) \le 0,\ u \ge 0\}. \tag{3.130}$$

Theorem 3.3.11 (sufficient conditions for the partial calmness condition for the KKT reformulation)**.** *Let $(\bar{x},\bar{y},\bar{u})$ be a local optimal solution of problem* (3.8). *Furthermore, assume that* EITHER *the family $\{(1.3)|\ x \in X\}$ has a uniform weak sharp minimum* OR *the set $\Omega := \{(x,y,u)|\, G(x) \le 0, \mathscr{L}(x,y,u) = 0, g(x,y) \le 0, u \ge 0\}$ is semismooth while inclusion $-(\bar{u}^\top \nabla g(\bar{x},\bar{y}), g(\bar{x},\bar{y})) \notin \mathrm{bd} N_\Omega(\bar{x},\bar{y},\bar{u})$ holds. Then problem* (3.8) *is partially calm at $(\bar{x},\bar{y},\bar{u})$ in the sense of Definition 3.3.10.*

Proof. Under the uniform weak sharp minimum, the result follows from [130]. In the second case, we obtain the partial calmness in the sense of Definition 3.3.10 by proceeding as in Theorem 3.2.8, where the counterpart of the mapping in (3.60) is chosen here as $\mathscr{M}(v) := \{(x,y,u) \in \Omega|\ -\sum_i^p u_i g_i(x,y) \le v\}$. □

The semismoothness is automatically satisfied for our set Ω here provided $G_j,\ j = 1,\dots,k$ is convex, and $(x,y) \mapsto \nabla_y f(x,y)$ and g are affine linear. Next we derive S-stationarity conditions for the KKT reformulation via the partial calmness concept in Definition 3.3.10.

Theorem 3.3.12 (S-stationarity conditions for the KKT reformulation via the partial calmness, I)**.** *Let $(\bar{x},\bar{y},\bar{u})$ be a local optimal solution of problem* (3.8)*, where the partial calmness in the sense of Definition 3.3.10 holds. Furthermore, assume that the following implication is satisfied at $(\bar{x},\bar{y},\bar{u})$:*

$$\left.\begin{array}{r} \sum_{j=1}^k \alpha_j \nabla G_j(\bar{x}) + \sum_{l=1}^m \gamma_l \nabla_x \mathscr{L}_l(\bar{x},\bar{y},\bar{u}) + \sum_{i=1}^p \beta_i \nabla_x g_i(\bar{x},\bar{y}) = 0 \\ \sum_{l=1}^m \gamma_l \nabla_y \mathscr{L}_l(\bar{x},\bar{y},\bar{u}) + \sum_{i=1}^p \beta_i \nabla_y g_i(\bar{x},\bar{y}) = 0 \\ \forall j = 1,\dots,k,\ \alpha_j \ge 0,\ \alpha_j G_j(\bar{x}) = 0 \\ \forall i = 1,\dots,p,\ \beta_i \ge 0,\ \beta_i g(\bar{x},\bar{y}) = 0 \\ \forall i = 1,\dots,p,\ \sum_{l=1}^m \gamma_l \nabla_{y_l} g_i(\bar{x},\bar{y}) \ge 0,\ \ \bar{u}_i \sum_{l=1}^m \gamma_l \nabla_{y_l} g_i(\bar{x},\bar{y}) = 0 \end{array}\right\} \Longrightarrow \left\{\begin{array}{l} \alpha = 0, \\ \beta = 0, \\ \gamma = 0. \end{array}\right. \tag{3.131}$$

Then the S-stationarity conditions (3.16), (3.23)–(3.25) *and* (3.30) *hold with $\|(\alpha,\gamma)\| \le r$ for some $r > 0$.*

Proof. Considering the definition of partial calmness in Definition 3.3.10, there exists $\lambda > 0$ such that $(\bar{x},\bar{y},\bar{u})$ is also a local optimal solution of the problem

$$\min_{x,y,u} \{F(x,y) - \lambda \sum_{i=1}^{p} u_i g_i(x,y)|\ (x,y,u) \in \Omega \cap \psi^{-1}(\Lambda)\}, \tag{3.132}$$

where $\Omega := \mathbb{R}^{n+m} \times \mathbb{R}^p_+$, $\Lambda := \mathbb{R}^k_- \times \mathbb{R}^p_- \times \{0_m\}$ and $\psi(x,y,u) := [G(x), g(x,y), \mathscr{L}(x,y,u)]$. We have $N_\Lambda(\psi(\bar{x},\bar{y},\bar{u})) = N_{\mathbb{R}^k_- \times \mathbb{R}^p_-}(G(\bar{x}), g(\bar{x},\bar{y})) \times \mathbb{R}^m$ with the formula of $N_{\mathbb{R}^k_- \times \mathbb{R}^p_-}(G(\bar{x}), g(\bar{x},\bar{y}))$ given in the right-hand-side of equation (3.80). Furthermore, the normal cone to Ω and the gradient of ψ are given by

$$N_\Omega(\bar{x},\bar{y},\bar{u}) = \{0_{n+m}\} \times \{\zeta \in \mathbb{R}^p | \zeta \le 0, \zeta^\top \bar{u} = 0\}, \tag{3.133}$$

$$\nabla \psi(x,y,u)^\top (\alpha,\beta,\gamma) = \triangle(\alpha,\beta,\gamma), \tag{3.134}$$

where $\triangle(\alpha,\beta,\gamma)$ is the matrix in (3.102). It follows from equalities (3.80) and (3.133)–(3.134) that the basic CQ applied to problem (3.132) at $(\bar{x},\bar{y},\bar{u})$ can equivalently be formulated as: there is no nonzero vector $(\alpha, \beta, \gamma) \in \mathbb{R}^{k+p+m}$ and a vector $\zeta \in \mathbb{R}^p$ (dummy multiplier) such that the first four lines of the left-hand-side of implication (3.131) and the system

$$\nabla_y g(\bar{x}, \bar{y})\gamma + \zeta = 0,\ \zeta \leq 0,\ \zeta^\top \bar{u} = 0 \tag{3.135}$$

are satisfied, respectively. It clearly follows that CQ (3.131) corresponds to the basic CQ (2.41) applied to problem (3.132), where the last line of the system in the left-hand-side of implication (3.131) is recovered from (3.135) by setting $\zeta = -\nabla_y g(\bar{x}, \bar{y})\gamma$. Hence, applying Proposition 2.2.5 to problem (3.132), it also follows from (3.80) and (3.133)–(3.134) that there exists $(\alpha, \beta, \gamma) \in \mathbb{R}^{k+p+m}$, with $\|(\alpha, \beta, \gamma)\| \leq r$ (for some $r > 0$), $\zeta \in \mathbb{R}^p$ and $\lambda > 0$ such that relationships (3.16)–(3.17) and (3.33)–(3.34) together with the following conditions are satisfied:

$$-\lambda g(\bar{x}, \bar{y}) + \nabla_y g(\bar{x}, \bar{y})\gamma + \zeta = 0,\ \zeta \leq 0,\ \zeta^\top \bar{u} = 0. \tag{3.136}$$

Thus condition (3.37) is regained from system (3.136) while noting that the feasibility of $(\bar{x}, \bar{y}, \bar{u})$ implies that we have $\sum_{i=1}^p \bar{u}_i g_i(\bar{x}, \bar{y}) = 0$.

We have shown that there exists $(\alpha, \beta, \gamma, \lambda)$ with $\lambda > 0$, such that we have (3.16)-(3.17) and (3.33)–(3.37) (where (3.35) and (3.36) are due to the feasibility of $(\bar{x}, \bar{y}, \bar{u})$). The S-stationarity conditions (3.16), (3.23)–(3.25) and (3.30) including the bound on (α, γ) are then derived from the latter conditions by proceeding as in the conclusion of Theorem 3.3.8. □

In the next result, we show that the qualification condition (3.131) can be weakened provided the perturbation map of the joint upper and lower-level feasible sets is calm.

Theorem 3.3.13 (S-stationarity conditions for the KKT reformulation via the partial calmness, II). *Let $(\bar{x}, \bar{y}, \bar{u})$ be a local optimal solution of problem* (3.8)*, where the partial calmness in the sense of Definition 3.3.10 holds. Furthermore, assume that the following CQs are satisfied at $(\bar{x}, \bar{y}, \bar{u})$:*

$$\text{The map } \mathscr{M}_2(t_1, t_2) := \{(x, y) |\, G(x) + t_1 \leq 0,\ g(x, y) + t_2 \leq 0\} \text{ is calm at } (0, 0, \bar{x}, \bar{y}), \tag{3.137}$$

$$\left.\begin{array}{r} \sum_{j=1}^k \alpha_j \nabla G_j(\bar{x}) + \sum_{l=1}^m \gamma_l \nabla_x \mathscr{L}_l(\bar{x}, \bar{y}, \bar{u}) + \sum_{i=1}^p \beta_i \nabla_x g_i(\bar{x}, \bar{y}) = 0 \\ \sum_{l=1}^m \gamma_l \nabla_y \mathscr{L}_l(\bar{x}, \bar{y}, \bar{u}) + \sum_{i=1}^p \beta_i \nabla_y g_i(\bar{x}, \bar{y}) = 0 \\ \forall j = 1, \ldots, k,\ \alpha_j \geq 0,\ \alpha_j G_j(\bar{x}) = 0 \\ \forall i = 1, \ldots, p,\ \beta_i \geq 0,\ \beta_i g(\bar{x}, \bar{y}) = 0 \\ \forall i = 1, \ldots, p,\ \sum_{l=1}^m \gamma_l \nabla_{y_l} g_i(\bar{x}, \bar{y}) \geq 0,\ \ \bar{u}_i \sum_{l=1}^m \gamma_l \nabla_{y_l} g_i(\bar{x}, \bar{y}) = 0 \end{array}\right\} \Longrightarrow \gamma = 0. \tag{3.138}$$

Then the S-stationarity conditions (3.16)*,* (3.23)–(3.25)*,* (3.30) *hold with $\|\gamma\| \leq r$ for some $r > 0$.*

Proof. Set $\Omega := \{(x, y, u) |\, u \geq 0,\ G(x) \leq 0,\ g(x, y) \leq 0\}$, then under the partial calmness assumption, we have from Definition 3.3.10, that there exists $\lambda > 0$ such that $(\bar{x}, \bar{y}, \bar{u})$ is also a local optimal solution of

$$\min_{x,y,u} \{F(x, y) - \lambda \sum_{i=1}^p u_i g_i(x, y) |\ (x, y, u) \in \Omega \cap \mathscr{L}^{-1}(0)\}.$$

Hence, it follows from Proposition 2.2.5 that if the following basic-type CQ holds at $(\bar{x}, \bar{y}, \bar{u})$

$$\left[0 \in \sum_{l=1}^m \gamma_l \nabla \mathscr{L}_l(\bar{x}, \bar{y}, \bar{u}) + N_\Omega(\bar{x}, \bar{y}, \bar{u}),\ \gamma \in \mathbb{R}^m\right] \Longrightarrow \gamma = 0, \tag{3.139}$$

then there exists $\gamma \in \mathbb{R}^m$ with $\|\gamma\| \leq r$ (for some $r > 0$) such that we have the optimality condition

$$0 \in \begin{bmatrix} \nabla F(\bar{x}, \bar{y}) \\ -\lambda g(\bar{x}, \bar{y}) \end{bmatrix} + \sum_{l=1}^m \gamma_l \nabla \mathscr{L}_l(\bar{x}, \bar{y}, \bar{u}) + N_\Omega(\bar{x}, \bar{y}, \bar{u}). \tag{3.140}$$

Now, observe that $\Omega = \Omega' \times \mathbb{R}^p_+$ with $\Omega' := \{(x,y)|\, G(x) \leq 0,\, g(x,y) \leq 0\}$. From Theorem 2.2.7 we have the expression of the normal cone to Ω' in the right-hand-side of equality (3.75) under CQ (3.137). Hence, we regain CQ (3.138) and the desired optimality conditions by inserting the aforementioned expression of $N_{\Omega'}(\bar{x},\bar{y})$ in (3.139) and (3.140)), while noting that $\zeta \in N_{\mathbb{R}^p_+}(\bar{u})$ if and only if $\zeta \leq 0$ and $\bar{u}^\top \zeta = 0$. Note that as in the proofs of Theorems 3.3.8 and 3.3.12, the first step gives the necessary optimality conditions of the type in (3.16)-(3.17) and (3.33)–(3.37), thus leading similarly to the result. □

Proceeding as in Remark 3.3.4, one can easily check that CQ (3.131) implies the fulfilment of both CQs (3.137) and (3.138). Also note that we have: CQ(3.128) $\Longrightarrow$ CQ(3.113) $\Longrightarrow$ CQ(3.97). However, the chain of implications that relates the resulting S, C and M-stationarity conditions, respectively, is completely different, cf. diagram in (3.31). Finally, the strong S-stationarity conditions for the bilevel program (P) can be obtained via its classical KKT reformulation as follows.

Corollary 3.3.14 (SP-S-stationarity via the KKT reformulation). *Let* $(\bar{x},\bar{y})$ *be a local optimal solution of problem (P), where the lower-level problem* (1.3) *is convex. Assume that the Slater CQ* (3.6) *holds at* $\bar{x}$. *Then* $(\bar{x},\bar{y})$ *is SP-S-stationary provided one of the following assertions holds at* $(\bar{x},\bar{y},\bar{u})$ *for all* $\bar{u} \in \Lambda(\bar{x},\bar{y})$*:*
(i) *Guignard CQ* (3.121)*,*
(ii) *MPEC-LICQ* (3.128)*,*
(iii) *Partial MPEC-LICQ* (3.97) *and CQ* (3.129)*,*
(iv) *Partial calmness in the sense of Definition 3.3.10 and CQ* (3.131)*,*
(v) *Partial calmness in the sense of Definition 3.3.10 and CQs* (3.137) *and* (3.138)*.*

Proof. In view of Theorem 3.1.3 (i), we have the results from Theorem 3.3.8, Theorem 3.3.9, Theorem 3.3.12 and Theorem 3.3.13, respectively. □

As usual, if (i), (ii), (iii), (iv) or (v) is satisfied at the point $(\bar{x},\bar{y},\bar{u})$, only for a single $\bar{u} \in \Lambda(\bar{x},\bar{y})$, we obtain the P-S-stationarity for $(\bar{x},\bar{y})$.

3.4 OPEC reformulation

In this section we are interested in deriving necessary optimality conditions for the classical optimistic bilevel optimization problem (P) via its OPEC reformulation (3.5). In order to apply the basic CQ (2.41) to the latter problem, we need to make sure that the normal cone mapping $Q : \mathbb{R}^n \times \mathbb{R}^m \rightrightarrows \mathbb{R}^m$ defined for $y \in K(x)$ by

$$Q(x,y) := N_{K(x)}(y) \tag{3.141}$$

while $Q(x,y) := \emptyset$, otherwise, is closed. Furthermore, to obtain detailed conditions, one should also be able to estimate the coderivative of this mapping. These issues will be studied in the next subsection. In Subsection 3.4.2, P-M-stationarity conditions are investigated via problem (3.5) and from various view points, while attempting to provide a clear comparison between the process to derive stationarity conditions for (P) via the KKT and OPEC/primal KKT reformulations.

3.4.1 The normal cone mapping

Recall the following result from [108, Proposition 3.3] stating that the mapping Q (3.141) is automatically closed, provided K is independent of x. To proceed first observe that in the latter case, the mapping $Q : \mathbb{R}^m \rightrightarrows \mathbb{R}^m$ is such that $\operatorname{gph} Q = \{(y,z)|\, z \in N_K(y)\}$.

Proposition 3.4.1 (closedness of the normal cone mapping when the lower-level feasible set is nonparametric). *Assume that* $K(x) := K$ *for all* $x \in X$*, then* $\operatorname{gph} Q = \{(y,z)|\, z \in N_K(y)\}$ *is closed as a subset of* $K \times \mathbb{R}^m$ *in the sense that: If* $x^k \to x$ $(x^k \in K)$ *and* $z^k \to z$ *with* $z^k \in N_K(x^k)$*, then* $z \in N_K(x)$.

Next we extend this result to the case where K is a moving set, i.e. a set-valued map from $\mathbb{R}^n$ to $\mathbb{R}^m$.

Proposition 3.4.2 (closedness of the normal cone mapping when the lower-level feasible set is parametric)**.** *Consider the normal cone mapping Q (3.141) and assume that $K : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ is convex-valued and inner semicontinuous on its graph. Then $\text{gph}\, Q := \{(x,y,z) | \, z \in N_{K(x)}(y)\}$ is closed as a subset of $\text{gph}\, K \times \mathbb{R}^m$ in the sense that: If $(x^k,y^k) \to (x,y)$ $((x^k,y^k) \in \text{gph}\, K)$ and $z^k \to z$ with $z^k \in Q(x^k,y^k)$, then we have $z \in Q(x,y)$.*

Proof. Let $(x^k,y^k) \to (x,y)$ $((x^k,y^k) \in \text{gph}K)$ and $z^k \to z$ with $z^k \in Q(x^k,y^k)$. Since $K(x^k)$ is assumed to be convex for all k, we have from the expression of the normal cone in (2.20) that

$$\langle z^k, u^k - y^k \rangle \leq 0, \; \forall u^k \in K(x^k), \; k \in \mathbb{N}. \tag{3.142}$$

K being inner semicontinuous on its graph, it follows that for an arbitrary $v \in K(x)$, there exists $v^k \in K(x^k)$ such that $v^k \to v$. Thus from (3.142), we have $\langle z^k, v^k - y^k \rangle \leq 0$ for all $k \in \mathbb{N}$. By passing to the limits, the latter implies that we have $\langle z, v - y \rangle \leq 0$ for all $v \in K(x)$; which completes the proof. □

Considering the case where the set-valued mapping K is defined as in (1.4), recall (cf. Subsection 2.1.4) that the concept of inner semicontinuity can be brought to usual terms through the following well-known result, see e.g. [89].

Lemma 3.4.3 (inner semicontinuity of K via the lower-level regularity)**.** *Let $(\bar{x},\bar{y}) \in \text{gph}\, K$ and assume that $(\bar{x},\bar{y})$ is lower-level regular* (3.65). *Then K is inner semicontinuous around $(\bar{x},\bar{y})$.*

In order to provide a detailed upper estimate of the coderivative of Q (3.141), we will use the following representation of the normal cone to the graph of $N_{\mathbb{R}^p_-}$ in terms of the normal cone to Π (3.95).

Lemma 3.4.4 (normal cone to the graph of $\mathbb{R}^p_-$ in terms of the normal cone to Π)**.** *Consider the point $(\bar{u},\bar{v}) \in \text{gph} N_{\mathbb{R}^p_-}$, then we have*

$$N_{\text{gph} N_{\mathbb{R}^p_-}}(\bar{u},\bar{v}) = \{(-u,v) \in \mathbb{R}^{2p} | (u,v) \in N_\Pi(-\bar{u},\bar{v})\}. \tag{3.143}$$

Proof. We start by noting that

$$\begin{aligned} \text{gph} N_{\mathbb{R}^p_-} &= \{(u,v) \in \mathbb{R}^{2p} | u \leq 0, \, v \geq 0, \, u^\top v = 0\} \\ &= \{(u,v) \in \mathbb{R}^{2p} | (-u,v) \in \Pi\}. \end{aligned}$$

This means that $\text{gph} N_{\mathbb{R}^p_-} = \vartheta^{-1}(\Pi)$, where $\vartheta(u,v) := (-u,v)$ and for $(\bar{u},\bar{v}) \in \mathbb{R}^{2p}$, one obviously has

$$\nabla \vartheta(\bar{u},\bar{v}) = \begin{bmatrix} -I_p & O \\ O & I_p \end{bmatrix}$$

with I_p and O denoting the $p \times p$ identity and zero matrix, respectively. Hence, the Jacobian matrix $\nabla \vartheta(\bar{u},\bar{v})$ is quadratic and nonsingular and it follows from [83, Corollary 2.12] that

$$N_{\text{gph} N_{\mathbb{R}^p_-}}(\bar{u},\bar{v}) = \nabla \vartheta(\bar{u},\bar{v})^\top N_\Pi(\vartheta(\bar{u},\bar{v})),$$

given that $\text{gph} N_{\mathbb{R}^p_-}$ and Π are closed sets. The result then follows. □

Based on Theorem 3.1 in [93] and the previous lemma, we now give an upper estimate of the coderivative of Q (3.141) that best suits our desire to write M-stationarity conditions for problem (3.5) in the sense of Definition 3.3.1. This will allow us to establish a clear relationship between the OPEC and KKT approaches to derive stationarity conditions for (P).

Theorem 3.4.5 (M-type representation of the upper estimate of the coderivative of the normal cone mapping)**.** *Let the lower-level problem* (1.3) *be convex and consider* $(\bar{x},\bar{y}) \in \mathrm{gph}S$ *such that the Slater CQ* (3.6) *is satisfied at* $\bar{x}$*, while the following set-valued mapping is calm at* $(0,\bar{x},\bar{y},\bar{u})$ *for all* $\bar{u} \in \Lambda(\bar{x},\bar{y})$*:*

$$M(v) := \big\{(x,y,u) \big|\ [g(x,y),\ u]+v \in \mathrm{gph}N_{\mathbb{R}^p_-}\big\}. \tag{3.144}$$

Then, for any $\gamma \in \mathbb{R}^m$*, the coderivative of* Q *at* $((\bar{x},\bar{y})|-\nabla_y f(\bar{x},\bar{y}))$ *can be estimated as*

$$\begin{aligned} D^*Q((\bar{x},\bar{y})|-\nabla_y f(\bar{x},\bar{y}))(\gamma) \subseteq \bigcup_{\bar{u}\in\Lambda(\bar{x},\bar{y})} \Bigg\{ \left[\begin{array}{c} \sum_{i=1}^p \bar{u}_i \nabla^2_{xy} g_i(\bar{x},\bar{y})^\top \gamma + \sum_{i=1}^p \beta_i \nabla_x g_i(\bar{x},\bar{y}) \\ \sum_{i=1}^p \bar{u}_i \nabla^2_{yy} g_i(\bar{x},\bar{y})^\top \gamma + \sum_{i=1}^p \beta_i \nabla_y g_i(\bar{x},\bar{y}) \end{array}\right] \Bigg| \\ \nabla_y g_\nu(\bar{x},\bar{y})\gamma = 0,\ \beta_\eta = 0 \\ \forall i \in \theta,\ \big(\beta_i > 0 \wedge \textstyle\sum_{l=1}^m \gamma_l \nabla_{y_l} g_i(\bar{x},\bar{y}) > 0\big) \vee \beta_i \textstyle\sum_{l=1}^m \gamma_l \nabla_{y_l} g_i(\bar{x},\bar{y}) = 0 \Bigg\}. \end{aligned} \tag{3.145}$$

Proof. Since the lower-level problem is convex, then the inclusion $(\bar{x},\bar{y}) \in \mathrm{gph}S$ implies that $-\nabla_y f(\bar{x},\bar{y}) \in Q(\bar{x},\bar{y})$. Hence, the combination of the Slater CQ at $\bar{x}$ and the calmness of the set-valued mapping M (3.144) at $(0,\bar{x},\bar{y},\bar{u})$ for all $\bar{u} \in \Lambda(\bar{x},\bar{y})$ give the following upper estimate of the coderivative of Q at $((\bar{x},\bar{y})|-\nabla_y f(\bar{x},\bar{y}))$

$$\begin{aligned} D^*Q((\bar{x},\bar{y})|-\nabla_y f(\bar{x},\bar{y}))(\gamma) \quad \subseteq \bigcup_{\bar{u}\in\Lambda(\bar{x},\bar{y})} \Big\{ & \textstyle\sum_{i=1}^p \bar{u}_i [\nabla(\nabla_y g_i)(\bar{x},\bar{y})]^\top \gamma + \\ & \nabla g(\bar{x},\bar{y})^\top D^*N_{\mathbb{R}^p_-}(g(\bar{x},\bar{y}),\bar{u})(\nabla_y g(\bar{x},\bar{y})\gamma) \Big\}, \end{aligned} \tag{3.146}$$

cf. [93, Theorem 3.1]. Now let (a,b) be an element of the set in the right-hand-side set of (3.146), then there exist $\bar{u} \in \Lambda(\bar{x},\bar{y})$ and $\beta \in D^*N_{\mathbb{R}^p_-}(g(\bar{x},\bar{y}),\bar{u})(\nabla_y g(\bar{x},\bar{y})\gamma)$ such that we have

$$\left[\begin{array}{c} a \\ b \end{array}\right] = \left[\begin{array}{c} \sum_{i=1}^p \bar{u}_i \nabla^2_{xy} g_i(\bar{x},\bar{y})^\top \gamma + \sum_{i=1}^p \beta_i \nabla_x g_i(\bar{x},\bar{y}) \\ \sum_{i=1}^p \bar{u}_i \nabla^2_{yy} g_i(\bar{x},\bar{y})^\top \gamma + \sum_{i=1}^p \beta_i \nabla_y g_i(\bar{x},\bar{y}) \end{array}\right]. \tag{3.147}$$

On the other hand, note that

$$\begin{aligned} \beta \in D^*N_{\mathbb{R}^p_-}(g(\bar{x},\bar{y}),\bar{u})(\nabla_y g(\bar{x},\bar{y})\gamma) \iff & \quad (\beta,\ -\nabla_y g(\bar{x},\bar{y})\gamma) \in N_{\mathrm{gph}N_{\mathbb{R}^p_-}}(g(\bar{x},\bar{y}),\bar{u}) \\ \overset{(3.143)}{\iff} & \quad (-\beta,\ -\nabla_y g(\bar{x},\bar{y})\gamma) \in N_\Pi(-g(\bar{x},\bar{y}),\bar{u}) \\ \overset{(3.96)}{\iff} & \quad \begin{cases} \nabla_y g_\nu(\bar{x},\bar{y})\gamma = 0,\ \beta_\eta = 0 \\ \forall i \in \theta,\ \big(\beta_i > 0 \wedge \nabla_y g_i(\bar{x},\bar{y})\gamma > 0\big) \vee \beta_i(\nabla_y g_i(\bar{x},\bar{y})\gamma) = 0. \end{cases} \end{aligned}$$

Combining the latter equivalence with equality (3.147), we have the desired result. □

3.4.2 Necessary optimality conditions

Considering the case where K is given in (1.4), we first present a slightly modified version of Theorem 6.1 in [93], while providing a sketch of the proof for completeness.

Theorem 3.4.6 (P-M-stationarity via the OPEC reformulation)**.** *Let* $(\bar{x},\bar{y})$ *be a local optimal solution of the OPEC reformulation* (3.5)*, i.e. of problem* (P)*, where the lower-level problem is convex. Assume that* $\bar{x}$ *satisfies the Slater CQ* (3.6)*, while for all* $\bar{u} \in \Lambda(\bar{x},\bar{y})$*, the mappings* M (3.144) *and* P *defined by*

$$P(v_1,v_2) := \big\{(x,y,u) \in M(v_1) |\ [G(x),\ \mathscr{L}(x,y,u)] + v_2 \in \mathbb{R}^k_- \times \{0_m\}\big\}, \tag{3.148}$$

are calm at $(0,\bar{x},\bar{y},\bar{u})$ *and* $(0,0,\bar{x},\bar{y},\bar{u})$*, respectively. Then* $(\bar{x},\bar{y})$ *is P-M-stationary with* $\|(\alpha,\gamma)\| \le r$*, for some number* $r > 0$*.*

Proof. Consider the following values for ψ and Λ, respectively:

$$\psi(x,y) := [G(x), x, y, -\nabla_y f(x,y)],\ \Lambda := \mathbb{R}^k_- \times \operatorname{gph} Q \ \text{ and } \Omega := \mathbb{R}^n \times \mathbb{R}^m. \tag{3.149}$$

Applying Proposition 2.2.5 to the corresponding operator constraint reformulation (2.40) of problem (3.5), one can easily check that there exists $(\alpha,\gamma) \in \mathbb{R}^{k+m}$ with $\|(\alpha,\gamma)\| \leq r$ (for some $r > 0$) such that we have

$$0 \in \nabla F(\bar{x},\bar{y}) + \begin{bmatrix} \nabla G(\bar{x})^\top \alpha + \nabla^2_{xy} f(\bar{x},\bar{y})^\top \gamma \\ \nabla^2_{yy} f(\bar{x},\bar{y})^\top \gamma \end{bmatrix} + D^* Q((\bar{x},\bar{y})| - \nabla_y f(\bar{x},\bar{y}))(\gamma) \tag{3.150}$$

provided Λ is closed and the set-valued mapping Ψ in (2.42) (with ψ, Λ and Ω given in (3.149)) is calm at $(0,\bar{x},\bar{y})$. Obviously, the closedness of Λ is ensured by the Slater CQ, by a combination of Proposition 3.4.2 and Lemma 3.4.3. As for the calmness of Ψ, it is obtained by the calmness of the set-valued mapping P (3.148) at $(0,0,\bar{x},\bar{y},\bar{u})$ for all $\bar{u} \in \Lambda(\bar{x},\bar{y})$. In fact the proof of the latter claim can be adapted from the proof of Theorem 4.3 in [93]. The result follows from the combination of inclusions (3.145) and (3.150). □

In our case, the constraint $G(x) \leq 0$ is included in ψ whereas it is part of Ω in [93]. The reason for this is to get a close link between CQ (3.97) and the qualification conditions in Theorem 3.4.6, cf. Remark 3.4.8. Also, here $\Lambda(\bar{x},\bar{y})$ is not a singleton as in [93]. To obtain this, at the place of the Slater CQ, it is required in [93] that $\nabla_y g(\bar{x},\bar{y})$ has full rank. Finally, in contrary to Theorem 6.1 of [93], the multipliers α and γ are bounded in Theorem 3.4.6 by a known number, something which can be useful when constructing an algorithm for the bilevel program.

Remark 3.4.7 (relationship with previous work on M-stationarity conditions for OPECs)**.** *There are two equivalent ways to interpret the coderivative term in the right-hand-side of inclusion* (3.146)*. The first one used in [114] consist in writing it directly in terms of $g(\bar{x},\bar{y})$, $\bar{u}$ and $\nabla_y g(\bar{x},\bar{y})\gamma$. It should however be mentioned that in [114], g does not depend on the parameter x. The second one that we have used here consist in first computing the normal cone to the graph of $N_{\mathbb{R}^p_-}$ at $(g(\bar{x},\bar{y}),\bar{u})$. Then, translating inclusion $(\beta, -\nabla_y g(\bar{x},\bar{y})\gamma) \in N_{\operatorname{gph} N_{\mathbb{R}^p_-}}(g(\bar{x},\bar{y}),\bar{u})$ by means of equality* (3.143) *directly leads to the M-stationarity conditions in the sense of Definition 3.1.6, cf. Theorems 3.4.5 and 3.4.6. At first view, it is not apparent that the optimality conditions in [114] are in fact equivalent to the corresponding M-stationarity conditions in* (3.16) *and* (3.23)–(3.26)*. Moreover, a condition was later suggested in [64], in order to obtain S-type optimality conditions for an OPEC from the M-ones. By the way, let us mention that in the case of our problem, the S-stationarity conditions defined in [64] correspond to those of Definition 3.1.8. Hence, in view of Theorem 3.3.9 the corresponding version of the Partial MPEC-LICQ* (3.129) *can also lead from M to S-type optimality conditions in the framework of an OPEC, something which was not mentioned in reference [64].*

Remark 3.4.8 (on the KKT and OPEC/primal KKT approaches to derive stationarity conditions for the classical bilevel program)**.** *At first recall that the OPEC and KKT reformulations of problem (P) are closely related, in the sense that if we insert the formula* (3.7) *of the normal cone to $K(x)$ in the former reformulation, we obtain the latter one. However it appears that the process to derive necessary optimality conditions for both problems are far from each other. In fact we have seen in Section 3.3 that most of the techniques for the more general MPCC/MPEC theory can be translated to the KKT reformulation, while for the OPEC one, one needs a more advanced machinery, namely the computation of the coderivative of the normal cone mapping, cf. Theorem 3.4.5. Thus it seems important to look closely at the two processes of deriving necessary optimality conditions for (P). To do so, we first start by noting the following link between the CQs used for the two approaches: Proceeding as in Remark 3.3.4 it is a simple exercise to check that if CQ* (3.97) *holds at $(\bar{x},\bar{y},\bar{u})$, then the set-valued mappings M* (3.144) *and P* (3.148) *are calm at $(0,\bar{x},\bar{y},\bar{u})$ and $(0,0,\bar{x},\bar{y},\bar{u})$, respectively. Thus if we neglect the bounds on the multipliers, it appears* *that under the same CQs, the KKT reformulation leads to stationarity conditions (SP-M-stationarity) which are stronger than those obtained via the OPEC reformulation (P-M-stationarity), see the diagram in* (3.31)*. It can not be otherwise considering the union in inclusion* (3.145)*. Moreover, the classical KKT reformulation provides a much bigger flexibility in designing surrogates for the other optimality conditions known for MPCCs, see for example the S-type stationarity conditions obtained in Subsection 3.3.3. However, it is very difficult to directly derive the S-type optimality conditions for an OPEC given that the Fréchet normal cone (which induces the S-stationarity) does not have as good calculus rules as that of Mordukhovich. Nevertheless, we should not lost sight on the fact that the OPEC reformulation is completely equivalent to the initial problem (cf. Theorem 3.1.2) while there is an unbalanced relationship between (P) and its KKT reformulation, cf. Theorem 3.1.3, due to the new variable u which is* a priori *part of problem* (3.8) *while u appears* a posteriori *in the process of estimating the coderivative of the normal cone set-valued mapping, see* (3.145)*.*

We now introduce a *partial calmness*-like concept for the OPEC reformulation (3.5), which is a weaker CQ that will allow us to still get the optimality conditions in Theorem 3.4.6. To bring the concept of partial calmness in the LLVF reformulation to the OPEC reformulation, one possibility is to observe that

$$\psi(x,y) \in \Lambda \Longleftrightarrow \rho(x,y) := d_\Lambda o \psi(x,y) = 0, \tag{3.151}$$

$$\text{where } \Lambda := \operatorname{gph} Q,\ \psi(x,y) := (x,y,-\nabla_y f(x,y)). \tag{3.152}$$

Here Q is given by (3.141) and d_Λ denotes the distance function. It follows that the OPEC reformulation (3.5) can equivalently be written as

$$\min_{x,y} \{F(x,y) |\, x \in X,\ \rho(x,y) = 0\}. \tag{3.153}$$

Such a transformation for an OPEC has already been suggested in [61], but with no further details on the process to necessary optimality conditions. Next, we first show that the basic CQ (2.41), which reduces for problem (3.153) to

$$\partial\rho(\bar{x},\bar{y}) \cap (-N_{X\times\mathbb{R}^m}(\bar{x},\bar{y})) = \emptyset \tag{3.154}$$

may often not be satisfied. To proceed, first observe that applying the chain rule (2.13) to ρ, it follows that we have

$$\partial\rho(\bar{x},\bar{y}) \subseteq \bigcup \left\{\partial\langle u,\psi\rangle(\bar{x},\bar{y}),\, u \in \mathbb{B} \cap N_\Lambda(\psi(\bar{x},\bar{y}))\right\}, \tag{3.155}$$

provided Λ is (locally) closed and hence, $\partial d_\Lambda(a) = \mathbb{B} \cap N_\Lambda(a)$ for any $a \in \Lambda$, cf. [109, Example 8.53].

Proposition 3.4.9 (failure of the basic CQ for problem (3.153))**.** *Let* $(\bar{x},\bar{y})$ *be a feasible point of the OPEC reformulation* (3.5)*, i.e. of problem* (P)*, where the lower-level problem is convex. Assume that* $\bar{x}$ *satisfies the Slater CQ* (3.6)*, while equality holds in* (3.155)*, then CQ* (3.154) *fails at* $(\bar{x},\bar{y})$*.*

Proof. Note that under the convexity assumption, problem (3.5) is well-defined as an equivalent reformulation of (P). Secondly, the Slater CQ (3.6) at $\bar{x}$ ensures the local closedness of Λ in (3.152), cf. combination of Proposition 3.4.2 and Lemma 3.4.3. This implies that inclusion (3.155) is satisfied. Now observe that if equality holds in (3.155), then $0 \in \partial\rho(\bar{x},\bar{y})$ since as a normal cone, $N_\Lambda(\psi(\bar{x},\bar{y}))$ always contains the origin point. For the latter reason, we also have $0 \in N_{X\times\mathbb{R}^m}(\bar{x},\bar{y})$. Hence, the result. □

This behavior of CQ (3.154) is close to that of the similar CQ (3.40) in the framework of the LLVF reformulation (3.3), where instead of $\rho(x,y)$ one has $f(x,y) - \varphi(x)$. In contrary to the LLVF case, where the weak basic CQ (3.42) was shown to work, in particular for the simple convex bilevel program (cf. Example 1), we show in what follows that its OPEC counterpart fails. In fact, let the upper-level feasible set X be convex, then, passing to the boundary of $N_{X\times\mathbb{R}^m}$ generates a CQ, analogous to (3.42) and expected to have more chances to be satisfied:

$$\partial\rho(\bar{x},\bar{y}) \cap (-\operatorname{bd} N_{X\times\mathbb{R}^m}(\bar{x},\bar{y})) = \emptyset. \tag{3.156}$$

Unfortunately, it is shown in the next proposition that this qualification condition fails under the assumptions of Proposition 3.4.9. In the proof, we consider that the space $\mathbb{R}^n \times \mathbb{R}^m$ is endowed with the product topology.

Proposition 3.4.10 (failure of the weak basic CQ for problem (3.153)). *Let $(\bar{x}, \bar{y})$ be a feasible point of the OPEC reformulation* (3.5)*, where the upper-level feasible set X is convex. Furthermore, suppose that all the assumptions of Proposition 3.4.9 are satisfied. Then the weak basic CQ* (3.156) *fails at* $(\bar{x}, \bar{y})$.

Proof. Recall that the convexity of X ensures that $X \times \mathbb{R}^m$ is a semismooth set, which implies that condition (3.156) is well-defined as a CQ in the sense of [61]. As already noticed in the proof of Proposition 3.4.9, we automatically have $0 \in \partial\rho(\bar{x}, \bar{y})$ under the assumptions made in the current result. Hence, it suffices now to show that we also have $0 \in \mathrm{bd} N_{X\times\mathbb{R}^m}(\bar{x}, \bar{y})$. Suppose that the latter does not hold, and observe that $\mathrm{bd} N_{X\times\mathbb{R}^m}(\bar{x}, \bar{y}) = N_{X\times\mathbb{R}^m}(\bar{x}, \bar{y}) \setminus \mathrm{int} N_{X\times\mathbb{R}^m}(\bar{x}, \bar{y})$, given that $N_{X\times\mathbb{R}^m}(\bar{x}, \bar{y})$ is a closed set, while taking into account the closeness of $X \times \mathbb{R}^m$. Thus we have $0 \in \mathrm{int} N_{X\times\mathbb{R}^m}(\bar{x}, \bar{y}) = \mathrm{int}\,[N_X(\bar{x}) \times \{0\}] = \emptyset$ (cf. product topology on $\mathbb{R}^n \times \mathbb{R}^m$), which is absurd. This completes the proof. □

Even though one can easily construct examples where equality holds in (3.155), the generalization of this fact would generally require the set Λ to be normally regular at $\psi(\bar{x}, \bar{y})$, which may not be easy to have given that Λ is the graph of a normal cone mapping. In the vein of Theorem 3.2.2 and similarly to Definition 3.3.10, we now introduce a partial calmness concept for the OPEC reformulation.

Definition 3.4.11 (partial calmness concept for the OPEC reformulation (3.5)). *Let $(\bar{x}, \bar{y})$ be a local optimal solution of problem* (3.5)*. This problem is partially calm at $(\bar{x}, \bar{y})$ if there exists a number $\lambda > 0$ such that $(\bar{x}, \bar{y})$ is a local optimal solution of the partially penalized problem*

$$\min_{x,y} \{F(x,y) + \lambda\rho(x,y) |\ x \in X\}. \tag{3.157}$$

In the next theorem we provide a sufficient condition ensuring the satisfaction of the above partial calmness condition.

Theorem 3.4.12 (sufficient condition for the partial calmness of the OPEC reformulation). *Assume that the following set-valued mapping is calm at $(0, 0, \bar{x}, \bar{y}, \bar{u})$ for all $\bar{u} \in \Lambda(\bar{x}, \bar{y})$:*

$$\overline{P}(v_1, v_2) := \{(x,y,u) \in X \times \mathbb{R}^m \times \mathbb{R}^p | \mathscr{L}(x,y,u) + v_1 = 0\} \cap M(v_2) \tag{3.158}$$

with M defined in (3.144)*. Then problem* (3.153) *is partially calm in the sense of Definition 3.4.11.*

Proof. Start by noting that the calmness of the mapping $\overline{P}$ (3.158) at $(0, 0, \bar{x}, \bar{y}, \bar{u})$ for all $\bar{u} \in \Lambda(\bar{x}, \bar{y})$ implies the calmness of the following set-valued mapping at $(0, \bar{x}, \bar{y})$

$$\Psi(v) := \{(x,y) \in X \times \mathbb{R}^m |\ \psi(x,y) + v \in \Lambda\}, \tag{3.159}$$

where ψ and Λ are given in (3.152), cf. [93, Proof of Theorem 4.3]. On the other hand, Ψ (3.159) is calm at $(0, \bar{x}, \bar{y})$ if and only if the following set-valued mapping is calm at $(0, \bar{x}, \bar{y})$

$$\widetilde{\Psi}(t) := \{(x,y) \in X \times \mathbb{R}^m | \rho(x,y) \le t\}. \tag{3.160}$$

This result established in [62], for the case where $X \times \mathbb{R}^m$ corresponds to a normed space, remains valid in our setting where X is defined in (1.4). Additionally, one can easily check that the calmness of $\widetilde{\Psi}$ (3.160) is equivalent to the calmness of a set-valued mapping obtained by replacing $\rho(x,y) \le t$ in (3.160) by $\rho(x,y) + t \le 0$. To conclude the proof, one can easily check, cf. Theorem 3.2.8, that the calmness of the mapping $\widetilde{\Psi}$ (3.160) at $(0, \bar{x}, \bar{y})$ implies the partial calmness of problem (3.153) at $(\bar{x}, \bar{y})$, in the sense of Definition 3.4.11. □

In the following result we show that the stationarity conditions obtained via the partial exact penalization in Definition 3.4.11 are identical to those derived in Theorem 3.4.6. Observe however from Theorem 3.4.12 that the slight adjustment of the mapping P in (3.148) to $\overline{P}$ (3.158) clearly show that the partial calmness concept in the sense of Definition 3.4.11 is a weaker qualification condition.

Theorem 3.4.13 (P-M-stationarity via the OPEC reformulation under the partial calmness)**.** *Let* $(\bar{x},\bar{y})$ *be a local optimal solution of the OPEC reformulation* (3.5)*, i.e. of problem* (P)*, where the lower-level problem is convex. Assume that* $\bar{x}$ *satisfies the Slater CQ* (3.6)*, while for all* $\bar{u} \in \Lambda(\bar{x},\bar{y})$*, the mapping* M (3.144) *is calm at* $(0,\bar{x},\bar{y},\bar{u})$*. Furthermore, assume that problem* (3.5) *is partially calm in the sense of Definition 3.4.11. Then the point* $(\bar{x},\bar{y})$ *is P-M-stationary with* $\|(\alpha,\gamma)\| \leq r$*, for some* $r > 0$*.*

Proof. Under the partial calmness assumption, it follows from Definition 3.4.11, that there exists $r_1 > 0$ such that $(\bar{x},\bar{y})$ is a local optimal solution of problem

$$\min_{x,y} \{F(x,y) + r_1\rho(x,y) | \, G(x) \leq 0\}.$$

Applying Proposition 2.2.5 to the latter problem, it follows that there exists α with $\|\alpha\| \leq r_2$ for some $r_2 > 0$, such that relationship (3.16) together with the following inclusion hold

$$0 \in \nabla F(\bar{x},\bar{y}) + \begin{bmatrix} \sum_{j=1}^{p} \alpha_j \nabla G_j(\bar{x}) \\ 0 \end{bmatrix} + r_1\partial\rho(\bar{x},\bar{y}). \tag{3.161}$$

Recall that with the Slater CQ satisfied at $\bar{x}$, inclusion (3.155) is satisfied. Hence, the combination of (3.155) and (3.161) and implies that there exists $\gamma \in \mathbb{R}^m$, with $\|\gamma\| \leq r_1$, such that inclusion (3.150) holds. The rest of the proof then follows as that of Theorem 3.4.6. In this case, r can be chosen as $r = \min\{r_1, r_2\}$. □

Remark 3.4.14 (the distance function as exact penalty term)**.** *The need of the partial calmness (in the sense of Definition 3.4.11) in order to have* ρ (3.151) *as an exact penalization term can be avoided by considering a classical result of Clarke [16, Proposition 2.4.3] which amounts to saying that the distance function is automatically an exact penalty term provided the objective function (F in our case) is Lipschitz continuous. To proceed one should observe that* $\psi(x,y) \in \Lambda$ *is also equivalent to* $d_{\psi^{-1}(\Lambda)}(x,y) = 0$*. Hence* $(\bar{x},\bar{y})$ *being a local optimal optimal solution of problem* (3.5) *implies that there exists* $\lambda > 0$ *such that* $(\bar{x},\bar{y})$ *is also a local optimal solution of the problem*

$$\min_{x,y} \{F(x,y) + \lambda \, d_{\psi^{-1}(\Lambda)}(x,y) | \, x \in X\},$$

without any qualification condition. In exchange though, computing the basic subdifferential of the distance function $d_{\psi^{-1}(\Lambda)}$ *would then require an assumption closely related to the partial calmness in the sense of Definition 3.4.11, i.e. the calmness of a certain set-valued mapping.*

To conclude this section we consider the following version of the OPEC reformulation (3.5) of problem (P), where the feasible set is nonparametric:

$$\min_{x,y} \{F(x,y) | 0 \in \nabla_y f(x,y) + N_K(y)\}. \tag{3.162}$$

Next we will give a new class of dual necessary optimality conditions based on the following primal optimality conditions of problem (3.162) derived in [25, Theorem 2.6]. For this result, it was assumed that the corresponding lower-level problem be amply parameterized in the sense of Dontchev and Rockafellar [41], that is

$$\text{rank } (\nabla^2_{xy} f(\bar{x},\bar{y})) = m, \tag{3.163}$$

where rank$A$ stands for the rank of the matrix A.

Theorem 3.4.15 (primal necessary optimality conditions of problem (3.162))**.** *Let* $(\bar{x},\bar{y})$ *be a local optimal solution of problem* (3.162) *and assume that CQ* (3.163) *is satisfied at this point. Further assume that K is a polyhedral set, then we have*

$$\nabla F(\bar{x},\bar{y})^\top(u,v) \geq 0,\ \forall (u,v):\ 0 \in \nabla^2_{xy} f(\bar{x},\bar{y})u + \nabla^2_{yy} f(\bar{x},\bar{y})v + N_{K(\bar{x},\bar{y})}(v), \tag{3.164}$$

where $K(\bar{x},\bar{y}) := T_K(\bar{y}) \cap (\nabla_y f(\bar{x},\bar{y}))^\perp$ *is the critical cone to K at* $(\bar{y}, \nabla_y f(\bar{x},\bar{y}))$*.*

If we give an explicit value to K, a complete description of the primal optimality conditions of the above result is possible. However our main concern though is to derive the dual form of these optimality conditions and compare them with those of the previous sections. In the next result, we show that the first step is possible without any additional assumption.

Theorem 3.4.16 (dual form of the optimality conditions in Theorem 3.4.15)**.** *Let* $(\bar{x},\bar{y})$ *be a local optimal solution of problem (3.162) and assume that CQ* (3.163) *is satisfied at this point. Further assume that K is a polyhedral set, then there exists* $\gamma \in \mathbb{R}^m$ *with* $\|\gamma\| \leq r$ *(for some* $r > 0$*) such that we have*

$$\nabla_x F(\bar{x},\bar{y}) + \nabla^2_{xy} f(\bar{x},\bar{y})^\top \gamma = 0, \tag{3.165}$$

$$0 \in \nabla_y F(\bar{x},\bar{y}) + \nabla^2_{yy} f(\bar{x},\bar{y})^\top \gamma + D^* N_{K(\bar{x},\bar{y})}(0,0)(\gamma). \tag{3.166}$$

Proof. Under the assumptions of this theorem, it follows from Theorem 3.4.15 that $(\bar{x},\bar{y})$ satisfies the primal optimality conditions (3.164). They imply that $(0,0)$ is an optimal solution of the problem

$$\min_{u,v} \{\nabla_x F(\bar{x},\bar{y})^\top u + \nabla_y F(\bar{x},\bar{y})^\top v |\ \psi(u,v) \in \Lambda\} \tag{3.167}$$

where the function ψ and the set Λ are respectively defined by

$$\psi(u,v) := \left[v,\ -\nabla^2_{xy} f(\bar{x},\bar{y})u - \nabla^2_{yy} f(\bar{x},\bar{y})v\right] \text{ and } \Lambda := \operatorname{gph} N_{K(\bar{x},\bar{y})}. \tag{3.168}$$

It is clear, in view of Proposition 3.4.1, that Λ is a closed set. Furthermore, since the critical normal critical cone $K(\bar{x},\bar{y})$ is a polyhedral set, then Λ is also a polyhedral set. In addition to the linearity of the function ψ (3.168), the set-valued mapping

$$\Psi(\vartheta) := \{(u,v)|\ \psi(u,v) + \vartheta \in \Lambda\}$$

is calm at $(0,0,0)$. Hence, by Proposition 2.2.5, applied to problem (3.167), there exists $(\xi,\gamma) \in N_\Lambda(\psi(0,0))$ with $\|(\xi,\gamma)\| \leq r$ for some $r > 0$, such that we have

$$\nabla F(\bar{x},\bar{y}) + \nabla \psi(0,0)^\top(\xi,\gamma) = 0, \tag{3.169}$$

One can easily observe that we respectively have

$$(\xi,\gamma) \in N_\Lambda(\psi(0,0)) \overset{(2.24)}{\Longleftrightarrow} \xi \in D^* N_{K(\bar{x},\bar{y})}(0,0)(-\gamma), \tag{3.170}$$

$$\text{and } \nabla\psi(0,0)^\top(\xi,\gamma) = \begin{bmatrix} -\nabla^2_{xy} f(\bar{x},\bar{y})^\top \gamma \\ \xi - \nabla^2_{yy} f(\bar{x},\bar{y})^\top \gamma \end{bmatrix}. \tag{3.171}$$

Combining relationships (3.169)–(3.171), we have the result. □

The stationarity conditions in Theorem 3.4.16 were labeled in [114] as *CM* (Critical-Mordukhovich)-stationarity conditions because of the presence of the normal cone to the *critical cone* $K(\bar{x},\bar{y})$. The result in [114] is tailored to the general setting with the lower-level feasible set K defined by functional inequalities (possibly nonlinear). Therefore, the assumptions there are stronger than the ones used here, precisely, they are the counterparts of those used in Corollary 3.2.14, i.e. the SSOC (3.88) and CRCQ (3.89). Using the same approach as in Theorem 3.4.5 to estimate the coderivative of the normal cone mapping, one can derive the P-M-stationarity conditions provided K is explicitly defined and possibly under additional assumptions. More details and discussions on estimating the generalized derivatives/coderivatives of normal cone mappings can be found in [3, 60, 63, 93, 114] and references therein.

3.5 Concluding comments to Chapter 3

Apart from Corollary 3.2.14, all the necessary optimality conditions derived in this chapter for problem (P) are tailored to the case where the lower-level problem may have more than one optimal solution. The KM-stationarity conditions were first derived in [128]. These conditions were then obtained in [26] from a different perspective and under various settings of the functions involved. The CQ used in both papers is the partial calmness condition introduced in the former one.

To the best of our knowledge, the KN-stationarity conditions were first derived in [125] under the convexity of the lower-level value function φ (3.2) while applying nonsmooth extensions of some well-known CQs which are not discussed in this thesis. Later in [120], these conditions are obtained while replacing the convexity of φ by the concavity of the function $f(x,y)-\varphi(x)$. Dempe et al. [26] also derive these conditions under the convexity of φ via the partial calmness condition. However, in the latter paper, it is additionally required that the solution map S (1.2) be inner semicompact, which is in fact not necessary as observed in [35] (see Theorem 3.2.12 (ii)). Considering a bilevel infinite program (i.e. where the lower-level has infinitely many inequality constraints), Dinh et al. [40] suggest necessary optimality conditions which are closely related to the KN-stationarity conditions. Here, the partial calmness is also used as CQ while convexity is assumed for the corresponding lower-level value function. It should however be mentioned that upper-level constraints are not considered in the latter paper. In the absence of convexity of φ, the KN-stationarity conditions were derived for the first time in [91] under the inner semicontinuity of S while using the partial calmness as CQ.

Note that in Subsection 3.2.3 (also see [33]), we also use the weak basic CQ (3.42) to derive both the KM and KN-stationarity conditions. Not only our proofs are different from those in aforementioned works, but we provide exact bounds (which can be computed via Proposition 2.2.5, cf. the proof) on some multipliers, which makes a really difference, especially in the framework of the partial calmness condition, cf. Theorem 3.2.12. Further discussions about necessary optimality conditions for (P) via the LLVF reformulation can be found in the above mentioned papers and their references.

As far as the P-M-stationarity is concerned, it was first obtained in [127] from a perspective different from that in Subsection 3.3.1. However, a clear connection between M-stationarity from both the KKT and OPEC reformulations was established in [34]. The other stationarity concepts discussed in Subsection 3.3 and tailored to (P), were introduced in the latter paper. Further details about the more general MPCC/MPEC can be found e.g. in [51, 100, 110, 126]. Considering the necessary optimality conditions of (P) obtained via the LLVF reformulation (KM and KN-stationarity) and those derived from the primal KKT/OPEC and KKT reformulations (M, C and S-stationarity) it was noted in Subsection 3.1.2 that they differ considerably from each other. On the one hand, the stationarity conditions obtained via the latter reformulations contain second order terms from the lower-level problem which may be expensive for algorithms. On the other hand however, the framework to derive stationarity conditions for (P) via the LLVF reformulation may be rather restrictive in some particular settings as recently shown by Henrion and Surowiec in the paper [65] where a set of conditions is given ensuring that the set-valued mapping $\Psi(v) := \{(x,y) \in \Omega |\ \psi(x,y)+v \in \Lambda\}$ (with $\Omega,\ \psi$ and Λ in (3.149)) is automatically calm, while its LLVF reformulation counterpart in (3.60) fails to be calm.

Let us now say some word about the lower-level regularity (3.65) widely used in Subsection 3.2.3. Note that in most cases, it can be replaced by the combination of CQ (3.77) and the following one:

$$\Big[\sum_{i=1}^{p} \beta_i \nabla g_i(\bar{x},\bar{y}) = 0,\ \ \beta_i \geq 0,\ \beta_i g_i(\bar{x},\bar{y}) = 0,\ \ i=1,\dots,p\Big] \Longrightarrow \beta_i = 0,\ i=1,\dots,p. \tag{3.172}$$

Observe here that in contrary to the lower-level regularity condition (3.65), the derivative of the functions $g_i,\ i=1,\dots,p$ in condition (3.172) are w.r.t. (x,y). The lower-level regularity (3.65) is strong enough to ensure that conditions (3.77) and (3.172) are both satisfied. But in general, the latter conditions are not related; moreover, it may happen that (3.77) holds while (3.172) fails. Thus the former can be replaced by a weaker CQ, in particular the calmness condition of a certain set-valued mapping. In the resulting framework, one still secure the Lipschitz continuity of the lower-level value function (3.2) needed in the corresponding results. This strategic choice to consider (3.77) and (3.172) simultaneously instead of (3.65) will be better understood in Chapter 5 when we consider the LLVF reformulation of the two-level value function, where the CQ of type (3.172) does note hold.

Finally, it is worth mentioning that if we neglect the bounds on the multipliers, all the stationarity conditions in this chapter can be generated via the coderivative of the lower-level solution set-valued mapping S (1.2). To see this, observe that if $(\bar{x},\bar{y})$ is a local optimal solution of problem (P), then

$$-\nabla_x F(\bar{x},\bar{y}) - \sum_{j=1}^{k} \alpha_j \nabla G_j(\bar{x}) \in D^* S(\bar{x},\bar{y})(\nabla_y F(\bar{x},\bar{y})), \tag{3.173}$$

provided some conditions are satisfied. It is a simple exercise to check this while using the estimates of the coderivative of S that we provide in Chapters 4 and 5. It would however be clear in Chapters 5 and 6 that to obtain the necessary optimality for the original optimistic bilevel program (P_o) and those of the pessimistic bilevel program (P_p), we may additionally need S to be Lipschitz-like.

4 Sensitivity analysis of OPCC and OPEC value functions

Considering the structure of the pessimistic and original optimistic bilevel programs, we provide general results on sensitivity analysis of value functions of optimization problems with complementarity constraints (OPCC) and those with generalized equation constraints (OPEC). They will then be applied in the subsequent chapters to obtained stationarity conditions for the aforementioned problems. To proceed, we first consider a general “abstract" framework of the value/marginal functions

$$\mu(x) := \min_y \{\psi(x,y)|\, y \in \Psi(x)\} \tag{4.1}$$

with $\psi\colon \mathbb{R}^n \times \mathbb{R}^m \to \overline{\mathbb{R}}$ and $\Psi\colon \mathbb{R}^n \rightrightarrows \mathbb{R}^m$. Denoting the argminimum mapping in (4.1) by

$$\Psi_o(x) := \arg\min_y \{\psi(x,y)|\, y \in \Psi(x)\} = \{y \in \Psi(x)|\, \psi(x,y) \le \mu(x)\},$$

we summarize in the next theorem some known results on general marginal functions needed in this chapter; see [89, Corollary 1.109] and [90, Theorem 5.2].

Theorem 4.0.1 (properties of general marginal functions)**.** *Let the marginal function μ be given in* (4.1)*, where the graph of Ψ is locally closed around $(\bar{x},\bar{y}) \in \operatorname{gph}\Psi$, and where ψ is strictly differentiable at this point. The following assertions hold:*

(i) *Let Ψ_o be inner semicontinuous at $(\bar{x},\bar{y})$. Then μ is lower semicontinuous at $\bar{x}$ and we have the following upper bound for its basic subdifferential*

$$\partial\mu(\bar{x}) \subseteq \nabla_x\psi(\bar{x},\bar{y}) + D^*\Psi(\bar{x},\bar{y})(\nabla_y\psi(\bar{x},\bar{y})).$$

If in addition Ψ is Lipschitz-like around $(\bar{x},\bar{y})$, then we also have the Lipschitz continuity of μ around $\bar{x}$.

(ii) *Let Ψ_o be inner semicompact at $\bar{x}$. Then μ is lower semicontinuous at $\bar{x}$ and*

$$\partial\mu(\bar{x}) \subseteq \bigcup_{\bar{y}\in\Psi_o(\bar{x})} \{\nabla_x\psi(\bar{x},\bar{y}) + D^*\Psi(\bar{x},\bar{y})(\nabla_y\psi(\bar{x},\bar{y}))\}.$$

If in addition Ψ is Lipschitz-like around $(\bar{x},\bar{y})$ for all vectors $\bar{y} \in \Psi_o(\bar{x})$, then μ is Lipschitz continuous around $\bar{x}$.

Depending on specific structures of the set-valued mapping Ψ, our aim in Sections 4.1 and 4.2 (also see Subsection 5.1.2) is to give detailed upper bounds for $D^*\Psi(\bar{x},\bar{y})$ in terms of problem data. Verifiable rules for Ψ to be Lipschitz-like will also be provided. Thus, implying explicit upper bounds for $\partial\mu(\bar{x})$ and the local Lipschitz continuity of μ. More discussions on the inner semicompactness and inner semicontinuity of argminimum mappings can be found in Remark 3.2.15 and the references therein.

It is worth mentioning that this chapter also provides the background to generate all the necessary optimality conditions of the previous chapter via the estimates of the coderivative of lower-level solution map (1.2) (see (3.173)). Together with its consequences in the next chapter, it bridges the gab between the classical optimistic bilevel program (P) and the two-level value function approach for (P_o) and (P_p) considered in Chapter 5 and Chapter 6, respectively.

An also important point to make here is that the notations used in the next section to designate functions, multipliers and index sets are independent from those of the previous chapter. Notations referring to those in Chapter 3 will however resume in Section 3.4 and follow similarly in Chapters 5 and 6.

4.1 Sensitivity analysis of OPCC value functions

In this section we consider the parametric optimization problem belonging to the class of *optimization problems with complementarity constraints* (OPCC) also commonly denoted as MPCC or MPEC:

$$\begin{aligned}\min_y\{F(x,y)|\quad & g(x,y)\leq 0,\, h(x,y)=0,\\ & G(x,y)\geq 0,\, H(x,y)\geq 0,\, G(x,y)^\top H(x,y)=0\},\end{aligned} \tag{4.2}$$

where $F:\mathbb{R}^n\times\mathbb{R}^m\to\mathbb{R}$, $g:\mathbb{R}^n\times\mathbb{R}^m\to\mathbb{R}^a$, $h:\mathbb{R}^n\times\mathbb{R}^m\to\mathbb{R}^b$ and $G,H:\mathbb{R}^n\times\mathbb{R}^m\to\mathbb{R}^d$ are all continuously differentiable functions. Denoting by

$$\begin{aligned}S^c(x):=\{y\in\mathbb{R}^m|\quad & g(x,y)\leq 0,\, h(x,y)=0,\\ & G(x,y)\geq 0,\, H(x,y)\geq 0,\, G(x,y)^\top H(x,y)=0\}\end{aligned} \tag{4.3}$$

the sets of *feasible solutions* to (4.2), associate with (4.2) the optimal value function

$$\mu^c(x):=\min_y\{F(x,y)|y\in S^c(x)\}. \tag{4.4}$$

The main goal of this section is to conduct a local sensitivity analysis for the OPCC problem (4.2) around the given optimal solution. By this we understand deriving efficient subdifferential estimates for the optimal value function (4.4), verifiable conditions for its local Lipschitz continuity and for the Lipschitz-like property of the feasible solution map (4.3) entirely in terms of the initial data of (4.2). According to the variational analysis results discussed above (see Theorem 4.0.1), this relates to evaluating the *coderivative* (2.24) of the solution map (4.3). Adopting the terminology already used in the previous chapter, the sensitivity analysis results established below and the associated constraint qualifications are expressed via the sets of M, C and S-type multipliers used in the corresponding M(ordukhovich), C(larke) and S(trong) stationarity conditions for our pessimistic and original optimistic problems; cf. Chapters 5 and 6.

Furthermore, fix a pair $(\bar{x},\bar{y})\in \operatorname{gph} S^c$ and associate with it the following partition of the indices for the functions involved in the complementarity system of (4.3):

$$\begin{aligned}\eta:=\quad & \eta(\bar{x},\bar{y}):=\{i=1,\ldots,d|\quad G_i(\bar{x},\bar{y})=0,\, H_i(\bar{x},\bar{y})>0\},\\ \theta:=\quad & \theta(\bar{x},\bar{y}):=\{i=1,\ldots,d|\quad G_i(\bar{x},\bar{y})=0,\, H_i(\bar{x},\bar{y})=0\},\\ \nu:=\quad & \nu(\bar{x},\bar{y}):=\{i=1,\ldots,d|\quad G_i(\bar{x},\bar{y})>0,\, H_i(\bar{x},\bar{y})=0\}.\end{aligned} \tag{4.5}$$

Recall that the difference between the various types of multiplier sets depends on the structure of the components corresponding to the biactive set θ. Now consider a vector $v\in\mathbb{R}^{n+m}$ and define the set of *M-type multipliers* associated with problem (4.2) by

$$\begin{aligned}\Lambda^{cm}(\bar{x},\bar{y},v):=\{(\alpha,\beta,\gamma,\zeta)|\quad & \alpha\geq 0,\, \alpha^\top g(\bar{x},\bar{y})=0,\\ & \gamma_\nu=0,\, \zeta_\eta=0,\\ & \forall i\in\theta,\, (\gamma_i<0\wedge\zeta_i<0)\vee\gamma_i\zeta_i=0,\\ & v+\nabla g(\bar{x},\bar{y})^\top\alpha+\nabla h(\bar{x},\bar{y})^\top\beta+\nabla G(\bar{x},\bar{y})^\top\gamma+\nabla H(\bar{x},\bar{y})^\top\zeta=0\}.\end{aligned} \tag{4.6}$$

Similarly we define the set $\Lambda^{cm}_y(\bar{x},\bar{y},v)$, with $v\in\mathbb{R}^m$, obtained by replacing the gradients of g,h,G, and H in equation $v+\nabla g(\bar{x},\bar{y})^\top\alpha+\nabla h(\bar{x},\bar{y})^\top\beta+\nabla G(\bar{x},\bar{y})^\top\gamma+\nabla H(\bar{x},\bar{y})^\top\zeta=0$ by their partial derivatives with respect to y. In the case where $v:=\nabla_y F(\bar{x},\bar{y})$ we denote $\Lambda^{cm}_y(\bar{x},\bar{y}):=\Lambda^{cm}_y(\bar{x},\bar{y},\nabla_y F(\bar{x},\bar{y}))$.

The corresponding sets of *C-type multipliers* denoted by $\Lambda^{cc}(\bar{x},\bar{y},v)$, $\Lambda^{cc}_y(\bar{x},\bar{y},v)$ and $\Lambda^{cc}_y(\bar{x},\bar{y})$ are defined similarly to $\Lambda^{cm}(\bar{x},\bar{y},v)$, $\Lambda^{cm}_y(\bar{x},\bar{y},v)$ and $\Lambda^{cm}_y(\bar{x},\bar{y})$, respectively, with the replacement of condition $(\gamma_i<0\wedge\zeta_i<0)\vee\gamma_i\zeta_i=0$ on θ by $\gamma_i\zeta_i\geq 0$ on the same set. For the case of *S-type multipliers* we need to define only $\Lambda_y^{cs}(\bar{x},\bar{y})$, which is a analog of $\Lambda_y^{cm}(\bar{x},\bar{y})$ with the replacement of the aforementioned condition therein by $\gamma_i \le 0 \wedge \zeta_i \le 0$ for all $i \in \theta$. The following links between the sets $\Lambda_y^{cs}(\bar{x},\bar{y})$, $\Lambda_y^{cm}(\bar{x},\bar{y})$, and $\Lambda_y^{cc}(\bar{x},\bar{y})$ is obvious:

$$\Lambda_y^{cs}(\bar{x},\bar{y}) \subseteq \Lambda_y^{cm}(\bar{x},\bar{y}) \subseteq \Lambda_y^{cc}(\bar{x},\bar{y}). \tag{4.7}$$

To further simplify the presentation of this section, we introduce the following *Lagrange-type* and *singular Lagrange-type functions*, respectively, associated with problem (4.2):

$$\begin{aligned} L(x,y,\alpha,\beta,\gamma,\zeta) &:= F(x,y) + \textstyle\sum_{i=1}^a \alpha_i g_i(x,y) + \sum_{j=1}^b \beta_j h_j(x,y) + \sum_{k=1}^d \gamma_k G_k(x,y) + \sum_{l=1}^d \zeta_l H_l(x,y), \\ L_o(x,y,\alpha,\beta,\gamma,\zeta) &:= \textstyle\sum_{i=1}^a \alpha_i g_i(x,y) + \sum_{j=1}^b \beta_j h_j(x,y) + \sum_{k=1}^d \gamma_k G_k(x,y) + \sum_{l=1}^d \zeta_l H_l(x,y). \end{aligned}$$

In the sequel the derivative of L_o with respect to (x,y) is often needed and is denoted by

$$\nabla L_o(x,y,\alpha,\beta,\gamma,\zeta) := \sum_{i=1}^a \alpha_i \nabla g_i(x,y) + \sum_{j=1}^b \beta_j \nabla h_j(x,y) + \sum_{k=1}^d \gamma_k \nabla G_k(x,y) + \sum_{l=1}^d \zeta_l \nabla H_l(x,y).$$

The following *optimal solution/argminimum* map for the OPCC problem (4.2) given by

$$S_o^c(x) := \{y \in S^c(x) | F(x,y) \le \mu^c(x)\} \tag{4.8}$$

(with S^c given by (4.3)) plays a significant role in our subsequent sensitivity analysis in this section.

4.1.1 Sensitivity analysis via M-type multipliers

To proceed in this subsection, we define the *M-qualification conditions* at $(\bar{x},\bar{y})$:

$$\begin{array}{ll} (A_1^1) & (\alpha,\beta,\gamma,\zeta) \in \Lambda^{cm}(\bar{x},\bar{y},0) \Longrightarrow \alpha = 0,\ \beta = 0,\ \gamma = 0,\ \zeta = 0, \\ (A_2^1) & (\alpha,\beta,\gamma,\zeta) \in \Lambda_y^{cm}(\bar{x},\bar{y},0) \Longrightarrow \nabla_x L_o(\bar{x},\bar{y},\alpha,\beta,\gamma,\zeta) = 0, \\ (A_3^1) & (\alpha,\beta,\gamma,\zeta) \in \Lambda_y^{cm}(\bar{x},\bar{y},0) \Longrightarrow \alpha = 0,\ \beta = 0,\ \gamma = 0,\ \zeta = 0, \end{array} \tag{4.9}$$

and observe the obvious links between them:

$$(A_2^1) \Longleftarrow (A_3^1) \Longrightarrow (A_1^1). \tag{4.10}$$

The next theorem provides a constructive upper estimate of the coderivative (2.24) of the OPCC feasible solution map (4.3) and gives a verifiable condition for its *robust Lipschitzian stability*, i.e., the validity of the Lipschitz-like property.

Theorem 4.1.1 (coderivative estimate and Lipschitz-like property of OPCC feasible solutions via M–multipliers)**.** *Let $(\bar{x},\bar{y}) \in gph\, S^c$, and let (A_1^1) hold at $(\bar{x},\bar{y})$. Then we have for all $v \in \mathbb{R}^m$*

$$D^*S^c(\bar{x},\bar{y})(v) \subseteq \{\nabla_x L_o(\bar{x},\bar{y},\alpha,\beta,\gamma,\zeta) | (\alpha,\beta,\gamma,\zeta) \in \Lambda_y^{cm}(\bar{x},\bar{y},v)\}. \tag{4.11}$$

If in addition (A_2^1) is satisfied at $(\bar{x},\bar{y})$, then S^c is Lipschitz-like around this point.

Proof. We start by recalling (cf. (3.106)) that the complementarity system in the feasible set of (4.2) is equivalent to the following inclusion $(G_i(x,y),H_i(x,y)) \in \{(u,v) \in \mathbb{R}^2 | u \ge 0, v \ge 0, u^\top v = 0\} := \Lambda_i$ for $i = 1,\dots,d$, and the graph of S^c can be rewritten in the form $\mathrm{gph}\, S^c = \{(x,y) | \psi(x,y) \in \Lambda\}$ via the vector-valued function ψ and the polyhedral set Λ defined by

$$\psi(x,y) := \left[g(x,y), h(x,y), (G_i(x,y),H_i(x,y))_{i=1}^d\right] \text{ and } \Lambda := \mathbb{R}_-^a \times \{0_b\} \times \prod_{i=1}^d \Lambda_i.$$

Applying Theorem 2.2.7, we have the following inclusion

$$N_{\operatorname{gph} S^c}(\bar{x},\bar{y}) \subseteq \nabla\psi(\bar{x},\bar{y})^\top N_\Lambda(\psi(\bar{x},\bar{y})) \tag{4.12}$$

provided the validity of the qualification condition

$$\left[\nabla\psi(\bar{x},\bar{y})^\top(\alpha,\beta,(\gamma_i,\zeta_i)_{i=1}^d)=0,\ (\alpha,\beta,(\gamma_i,\zeta_i)_{i=1}^d)\in N_\Lambda(\psi(\bar{x},\bar{y}))\right] \Longrightarrow \begin{cases} \alpha=0,\ \beta=0, \\ \gamma=0,\ \zeta=0. \end{cases} \tag{4.13}$$

It is easy to check that we have equality $\nabla\psi(\bar{x},\bar{y})^\top(\alpha,\beta,(\gamma_i,\zeta_i)_{i=1}^d)=\nabla L_o(\bar{x},\bar{y},\alpha,\beta,\gamma,\zeta)$ for any quadruple $(\alpha,\beta,\gamma,\zeta)$ and that, by the product formula for limiting normals, we have

$$N_\Lambda(\psi(\bar{x},\bar{y}))=N_{\mathbb{R}^a_-}(g(\bar{x},\bar{y}))\times N_{\{0_b\}}(h(\bar{x},\bar{y}))\times\prod_{i=1}^d N_{\Lambda_i}(G_i(\bar{x},\bar{y}),H_i(\bar{x},\bar{y})).$$

Using the expression of the normal cone to the sets Λ_i, $i=1,\ldots,d$, from Lemma 3.3.1 (cf. 3.96), we get

$$N_{\Lambda_i}(G_i(\bar{x},\bar{y}),H_i(\bar{x},\bar{y}))=\{(\gamma_i,\zeta_i)|\quad \begin{array}{l}\gamma_i=0 \text{ if } i\in\nu,\ \zeta_i=0 \text{ if } i\in\eta \\ (\gamma_i<0,\ \zeta_i<0)\vee(\gamma_i\zeta_i=0),\ \text{if } i\in\theta\},\end{array}$$

which implies that the qualification condition (4.13) reduces to (A_1^1) in this case and that inclusion (4.11) in the theorem results from (4.12) and the coderivative definition (2.24). Finally, the Lipschitz-like property of S^c around $(\bar{x},\bar{y})$ under the additional M-qualification condition (A_2^1) follows from (4.11) due to the coderivative criterion (2.30). □

Now we can readily get efficient estimates of the limiting subdifferential of the value function (4.4) and verifiable conditions for its local Lipschitz continuity.

Theorem 4.1.2 (M-type sensitivity analysis for OPCC value functions)**.** *The following assertions hold for the value function μ^c in* (4.4)*:*

(i) *Let the argminimum mapping S_o^c from* (4.8) *be inner semicontinuous at $(\bar{x},\bar{y})$, and let (A_1^1) hold at $(\bar{x},\bar{y})$. Then we have the subdifferential upper estimate*

$$\partial\mu^c(\bar{x})\subseteq\{\nabla_x L(\bar{x},\bar{y},\alpha,\beta,\gamma,\zeta)|\,(\alpha,\beta,\gamma,\zeta)\in\Lambda_y^{cm}(\bar{x},\bar{y})\}.$$

If in addition (A_2^1) is satisfied at $(\bar{x},\bar{y})$, then μ^c is Lipschitz continuous around $\bar{x}$.

(ii) *Assume that S_o^c is inner semicompact at $\bar{x}$ and that (A_1^1) holds at $(\bar{x},\bar{y})$ for all $\bar{y}\in S_o^c(x)$. Then we have the subdifferential upper estimate*

$$\partial\mu^c(\bar{x})\subseteq\{\nabla_x L(\bar{x},\bar{y},\alpha,\beta,\gamma,\zeta)|\,\bar{y}\in S_o^c(\bar{x}),\,(\alpha,\beta,\gamma,\zeta)\in\Lambda_y^{cm}(\bar{x},\bar{y})\}.$$

If in addition (A_2^1) is satisfied at $(\bar{x},\bar{y})$ for all $\bar{y}\in S_o^c(\bar{x})$, then the value function μ^c is Lipschitz continuous around $\bar{x}$.

Proof. It follows from the results of Theorem 4.1.1 and Theorem 4.0.1. □

Note that a subdifferential upper estimate similar to assertion (ii) Theorem 4.1.2 was obtained in [76] in the case of $G(x,y):=y$ under a certain growth hypothesis implying the inner semicompactness of the optimal solution map S_o^c from (4.8).

We do not pay any special attention to the lower semicontinuity of the value function (4.4) in Theorem 4.1.2 and subsequent results on value functions. By Theorem 4.0.1 this easily follows from the proof under the inner semicontinuity or the weaker inner semicompactness of the solution map S_o^c.

Remark 4.1.3 (On the qualification conditions)**.** *There are various sufficient conditions for the validity of the qualification condition* (A_1^1)*; see, e.g., Section 3.3 for related developments. Furthermore,* (A_1^1) *can be replaced by the weaker* calmness *assumption on the mapping*

$$\Psi(v) := \{(x,y) |\ \psi(x,y) + v \in \Lambda\}, \tag{4.14}$$

where ψ *and* Λ *are defined in the proof of Theorem 4.1.1, cf. Theorem 2.2.7. Note that the latter calmness assumption automatically holds when the mappings* g, h, G *and* H *are linear. Observe finally that due to the relationships* (4.10) *both assumptions* (A_1^1) *and* (A_2^1) *can be replaced by the fulfillment of the single condition* (A_3^1).

Remark 4.1.4 (on the difference between the upper estimates of the subdifferential of OPCC value functions)**.** *Following the pattern of Theorem 4.0.1, the basic difference between the upper estimate of* $\partial\mu^c$ *in assertions (i) and (ii) of Theorem 4.1.2 resides in the fact that in the first case we have to compute the gradient of the Lagrange-type function L associated with the OPCC* (4.2) *only at the point* $(\bar{x},\bar{y})$ *where* S_o^c *is inner semicontinuous. In the second case though this should be done at all* $(\bar{x}, y)$ *with* $y \in S_o^c(\bar{x})$*. Thus the upper bound of* $\partial\mu^c$ *obtained under the inner semicontinuity is obviously much tighter since it is always a subset of the one in (ii). As already mentioned in the previous chapter (see Remark 3.2.15), the inner semicontinuity of* S_o^c *is automatically satisfied if this mapping is Lipschitz-like around the point in question. Moreover, if* $S_o^c(\bar{x}) = \{\bar{y}\}$*, then* S_o^c *is inner semicontinuous at* $(\bar{x},\bar{y})$*.*

Remark 4.1.5 (more on relationships to previous work on the topic)**.** *Recall here that the technique employed in Theorem 4.1.1 that transforms the complementarity system in* (4.2) *into an inclusion in the pattern of* (3.106) *is rather common in the field of OPCCs to study some issues different from those considered here, cf. Section 3.3 (also see [97, 123, 127]) for related developments. Note however that some differences occur in constructing the set* Λ *corresponding, in the proof of Theorem 4.1.1, to* $\{(u,v) \in \mathbb{R}^d |\ u \geq 0,\ v \geq 0,\ u^\top v = 0\}$ *while in the aforementioned papers* $\Lambda = gph N_{R_+^d}$ *if often chosen. Note also in [66] a transformation in the vein of* (3.106) *is employed to derive an exact penalty result and then optimality conditions for the so-called* mathematical programs with vanishing constraints. *Having in mind this transformation, the methods developed here (cf., in particular, the proofs of Theorem 4.1.1 and Theorem 4.1.2) can readily be applied to conduct a local sensitivity analysis for the latter class of programs.*

4.1.2 Sensitivity analysis via C-type multipliers

Similarly to Subsection 3.1 we introduce the following *C-qualification conditions* at $(\bar{x},\bar{y})$:

$$\begin{array}{ll} (A_1^2) & (\alpha,\beta,\gamma,\zeta) \in \Lambda^{cc}(\bar{x},\bar{y},0) \Longrightarrow \alpha = 0,\ \beta = 0,\ \gamma = 0,\ \zeta = 0, \\ (A_2^2) & (\alpha,\beta,\gamma,\zeta) \in \Lambda_y^{cc}(\bar{x},\bar{y},0) \Longrightarrow \nabla_x L_o(\bar{x},\bar{y},\alpha,\beta,\gamma,\zeta) = 0, \\ (A_3^2) & (\alpha,\beta,\gamma,\zeta) \in \Lambda_y^{cc}(\bar{x},\bar{y},0) \Longrightarrow \alpha = 0,\ \beta = 0,\ \gamma = 0,\ \zeta = 0, \end{array} \tag{4.15}$$

with the similar relationships between them: $(A_2^2) \Longleftarrow (A_3^2) \Longrightarrow (A_1^2)$. To proceed, we use the nonsmooth transformation of the feasible set to the OPCC introduced in [110] and that we also considered in the proof of Theorem 3.3.6:

$$S^c(x) := \{y \in \mathbb{R}^m |\ g(x,y) \leq 0,\ h(x,y) = 0,\ \min\{G_i(x,y), H_i(x,y)\} = 0,\ i = 1,\dots,d\}. \tag{4.16}$$

Employing this transformation, a C-counterpart of Theorem 4.1.2 can be derived with a different proof and a larger estimate for the coderivative of S^c under the C-qualification conditions. In this proof, more detail is provided for a better understanding of the special case in the proof of Theorem 3.3.6.

Theorem 4.1.6 (coderivative estimate and Lipschitz-like property of OPCC feasible solutions via C-multipliers)**.** *Let $(\bar{x},\bar{y}) \in gph\, S^c$, and let (A_1^2) hold at $(\bar{x},\bar{y})$. Then we have for all $v \in \mathbb{R}^m$*

$$D^*S^c(\bar{x},\bar{y})(v) \subseteq \{\nabla_x L_o(\bar{x},\bar{y},\alpha,\beta,\gamma,\zeta) \,|\, (\alpha,\beta,\gamma,\zeta) \in \Lambda_y^{cc}(\bar{x},\bar{y},v)\}.$$

If in addition (A_2^2) is satisfied at $(\bar{x},\bar{y})$, then S^c is Lipschitz-like around this point.

Proof. From the expression of S^c in (4.16) we get $\text{gph}\, S^c = \{(x,y) \,|\, \psi(x,y) \in \Lambda\}$, where ψ and Λ are defined by

$$\psi(x,y) := [g(x,y),\, h(x,y),\, V(x,y)] \text{ and } \Lambda := \mathbb{R}^a_- \times \{0_b\} \times \{0_d\} \tag{4.17}$$

with $V_i(x,y) := \min\{G_i(x,y), H_i(x,y)\} = 0$ for $i = 1,\dots,d$. Since ψ is locally Lipschitzian around $(\bar{x},\bar{y})$, it follows from Theorem 2.2.7 that

$$N_{\text{gph}\, S^c}(\bar{x},\bar{y}) \subseteq \{\partial\langle u,\psi\rangle(\bar{x},\bar{y}) \,|\, u \in N_\Lambda(\psi(\bar{x},\bar{y}))\} \tag{4.18}$$

provided that the qualification condition

$$\Big[0 \in \partial\langle u,\psi\rangle(\bar{x},\bar{y}),\;\; u \in N_\Lambda(\psi(\bar{x},\bar{y}))\Big] \Longrightarrow u = 0 \tag{4.19}$$

is satisfied. Furthermore, we have the normal cone representation

$$N_\Lambda(\psi(\bar{x},\bar{y})) = N_{\mathbb{R}^a_-}(g(\bar{x},\bar{y})) \times N_{\{0_b\}}(h(\bar{x},\bar{y})) \times N_{\{0_d\}}(V(\bar{x},\bar{y})) \tag{4.20}$$

and calculate the subdifferential of the scalarization in (4.18) by

$$\partial\langle(\alpha,\beta,\chi),\psi\rangle(\bar{x},\bar{y}) = \nabla g(\bar{x},\bar{y})^\top \alpha + \nabla h(\bar{x},\bar{y})^\top \beta + \partial\langle\chi,V\rangle(\bar{x},\bar{y}) \tag{4.21}$$

for $(\alpha,\beta,\chi) \in N_\Lambda(\psi(\bar{x},\bar{y}))$. Since the function V is nondifferentiable and χ may contain negative components by (4.20), we apply the convex hull "co" to our basic subdifferential (2.3) in (4.21) in order to instate the plus/minus symmetry

$$\partial\langle\chi,V\rangle(\bar{x},\bar{y}) \subseteq \text{co}\,\partial\langle\chi,V\rangle(\bar{x},\bar{y}) \subseteq \sum_{i=1}^{d} \chi_i \bar{\partial} V_i(\bar{x},\bar{y})$$

via Clarke's generalized gradient $\bar{\partial} V_i$. Considering the partition of the index set $\{1,\dots,d\}$ in (4.5), we arrive by [16] at the following calculations:

$$\bar{\partial} V_i(\bar{x},\bar{y}) = \begin{cases} \nabla G_i(\bar{x},\bar{y}) & \text{if } i \in \eta, \\ \nabla H_i(\bar{x},\bar{y}) & \text{if } i \in \nu, \\ \text{co}\,\{\nabla G_i(\bar{x},\bar{y}), \nabla H_i(\bar{x},\bar{y})\} & \text{if } i \in \theta. \end{cases}$$

Invoking the classical Carathéodory theorem gives us

$$\text{co}\,\{\nabla G_i(\bar{x},\bar{y}), \nabla H_i(\bar{x},\bar{y})\} = \{t_i \nabla G_i(\bar{x},\bar{y}) + (1-t_i)\nabla H_i(\bar{x},\bar{y}) \,|\, t_i \in [0,1]\},$$

and hence we obtain from (4.21) the inclusions

$$\begin{aligned} \partial\langle(\alpha,\beta,\chi),\psi\rangle(\bar{x},\bar{y}) \subseteq\;& \big\{\nabla L_o(\bar{x},\bar{y},\alpha,\beta,\gamma,\zeta) \,|\, \gamma_\eta = 0, \zeta_\nu = 0 \\ & \quad \forall i \in \theta, \exists t_i \in [0,1], r_i \in \mathbb{R} \text{ s.t. } \gamma_i = r_i t_i,\, \zeta_i = r_i(1-t_i)\big\} \\ \subseteq\;& \big\{\nabla L_o(\bar{x},\bar{y},\alpha,\beta,\gamma,\zeta) \,|\, \gamma_\eta = 0, \zeta_\nu = 0 \\ & \quad \forall i \in \theta,\, \gamma_i\zeta_i \geq 0\big\}. \end{aligned} \tag{4.22}$$

Since the qualification condition (4.19) is equivalent to

$$\{(\alpha,\beta,\chi) \,|\, 0 \in \partial\langle(\alpha,\beta,\chi),\psi\rangle(\bar{x},\bar{y}),\, (\alpha,\beta,\chi) \in N_\Lambda(\psi(\bar{x},\bar{y}))\} = \{(0,0,0)\},$$

the second inclusion in (4.22) shows that (A_1^2) is sufficient for this to hold. Furthermore, by (4.18) the second inclusion of (4.22) leads to an upper estimate of $N_{\text{gph}\, S^c}$, which allows us via the coderivative definition (2.24) to recover the upper bound of D^*S^c in the theorem. The latter implies the Lipschitz-like property of S^c under (A_2^2) as in Theorem 4.1.1. □

As in the previous subsection, we arrive at the following sensitivity results for the OPCC value function (4.4) via C-multipliers.

Theorem 4.1.7 (C-type sensitivity analysis for OPCC value functions)**.** *The following assertions hold for the value function μ^c in* (4.4)*:*

(i) *Let the optimal solution map S_o^c is inner semicontinuous at $(\bar{x},\bar{y})$, and let (A_1^2) holds at $(\bar{x},\bar{y})$. Then we have the subdifferential upper estimate*

$$\partial\mu^c(\bar{x}) \subseteq \{\nabla_x L(\bar{x},\bar{y},\alpha,\beta,\gamma,\zeta) \,|\, (\alpha,\beta,\gamma,\zeta) \in \Lambda_y^{cc}(\bar{x},\bar{y})\}.$$

If in addition (A_2^2) holds at $(\bar{x},\bar{y})$, then μ^c is Lipschitz continuous around $\bar{x}$.

(ii) *Assume that S_o^c is inner semicompact at $\bar{x}$ and that (A_1^2) holds at $(\bar{x},\bar{y})$ for all $\bar{y} \in S_o^c(x)$. Then we have the subdifferential upper estimate*

$$\partial\mu^c(\bar{x}) \subseteq \{\nabla_x L(\bar{x},\bar{y},\alpha,\beta,\gamma,\zeta) \,|\, \bar{y} \in S_o^c(\bar{x}),\, (\alpha,\beta,\gamma,\zeta) \in \Lambda_y^{cc}(\bar{x},\bar{y})\}.$$

If in addition (A_2^2) also holds at $(\bar{x},\bar{y})$ for all $\bar{y} \in S_o^c(\bar{x})$, then μ^c is Lipschitz continuous around $\bar{x}$.

Proof. It follows from the results of Theorem 4.1.7 and Theorem 4.0.1. □

Note that assertion (ii) of Theorem 4.1.7 can be found in [76] for $G(x,y) = y$ under the following assumption corresponding to the replacement of the set $\Lambda^{cc}(\bar{x},\bar{y},0)$ in (A_1^2) by

$$\begin{aligned}\{(\alpha,\beta,\gamma,\zeta)| \quad & \alpha \geq 0,\, \alpha^\top g(\bar{x},\bar{y}) = 0,\\ & \gamma_\eta = 0,\, \zeta_\nu = 0,\\ & \forall i \in \theta,\, \exists t_i \in [0,1], r_i \in \mathbb{R} \text{ s.t. } \gamma_i = r_i t_i,\, \zeta_i = r_i(1-t_i),\\ & \nabla g(\bar{x},\bar{y})^\top \alpha + \nabla h(\bar{x},\bar{y})^\top \beta + \nabla G(\bar{x},\bar{y})^\top \gamma + \nabla H(\bar{x},\bar{y})^\top \zeta = 0\}.\end{aligned}$$

The latter assumption is weaker than (A_1^2), but in our assumption we simply need to check that the components of γ and ζ are of the same sign on θ rather than constructing them as in the above set. It is also important to mention that all the points made in Remark 4.1.3 can be restated here accordingly. In particular, (A_1^2) can be substituted by the weaker calmness of the set-valued mapping Ψ from (4.14) with ψ and Λ given in (4.17). This is obviously satisfied if the functions g,h,G, and H are linear, because $V_i(x,y) = \min\{G_i(x,y), H_i(x,y)\}$ is piecewise linear provided the linearity of G_i and H_i.

4.1.3 Sensitivity analysis via S-type multipliers

The need for S-type stationarity conditions in the context of OPCCs is the best one would want to have since these conditions are *equivalent* to the KKT type optimality conditions whenever the OPCC is treated as an ordinary nonlinear optimization problem.

Having this in mind, we attempt here to suggest a tighter upper bound for the basic subdifferential of the OPCC value function μ^c (4.4). In order to obtain an upper bound for $\partial\mu^c$ containing $\Lambda_y^{cs}(\bar{x},\bar{y})$ rather than $\Lambda_y^{cm}(\bar{x},\bar{y})$ or $\Lambda_y^{cc}(\bar{x},\bar{y})$, we impose the following *S-qualification condition* (with the index set I defined by $I := I(\bar{x},\bar{y}) := \{i = 1,\dots,a|\, g_i(\bar{x},\bar{y}) < 0\}$)

$$(A_1^3) \qquad \left.\begin{aligned} \nabla L_o(\bar{x},\bar{y},\alpha,\beta,\gamma,\zeta) = 0\\ \alpha_I = 0,\, \gamma_\nu = 0,\, \zeta_\eta = 0 \end{aligned}\right\} \Longrightarrow \gamma_\theta = 0,\, \zeta_\theta = 0 \tag{4.23}$$

introduced by Ye [123] and later named in [126] as *Partial MPEC-LICQ* (linear independence constraint qualification). This condition and another close while weaker one have also been used by Flegel et al. [51] to recover the S-stationarity conditions of a OPCC from the M-ones. In the next theorem we obtain a new S-type upper bound for $\partial\mu^c$ by a similar methodology, i.e., going from the M-type bound provided above. Note that assumption (A_1^1) is the one introduced in Subsection 4.1.1.

Theorem 4.1.8 (S-type sensitivity analysis for OPCC value functions). *The following assertions hold for the value function μ^c from* (4.4)*:*

(i) *Let the optimal solution map S_o^c be inner semicontinuous at $(\bar{x},\bar{y})$, and let assumptions (A_1^1) and (A_1^3) be satisfied at $(\bar{x},\bar{y})$. Then we have*

$$\partial\mu^c(\bar{x}) \subseteq \{\nabla_x L(\bar{x},\bar{y},\alpha,\beta,\gamma,\zeta) | (\alpha,\beta,\gamma,\zeta) \in \Lambda_y^{cs}(\bar{x},\bar{y})\}.$$

(ii) *Let S_o^c be inner semicompact at $\bar{x}$ with (A_1^1) and (A_1^3) being satisfied at $(\bar{x},\bar{y})$ for all $\bar{y} \in S_o^c(\bar{x})$. Then we have*

$$\partial\mu^c(\bar{x}) \subseteq \{\nabla_x L(\bar{x},\bar{y},\alpha,\beta,\gamma,\zeta) | \bar{y} \in S_o^c(\bar{x}),\, (\alpha,\beta,\gamma,\zeta) \in \Lambda_y^{cs}(\bar{x},\bar{y})\}.$$

Proof. We provide the proof only for assertion(i); the other case can be proved similarly.

Assuming (A_1^1) and the inner semicontinuity of S_o^c, we have the upper estimate of $\partial\mu^c$ from Theorem 4.1.2 (i). Further, denote by $A(\bar{x},\bar{y})$ (resp. $B(\bar{x},\bar{y})$) the right-hand side of the inclusion in Theorem 4.1.2 (i) (resp. Theorem 4.1.8 (i)). It remains to show that $A(\bar{x},\bar{y}) = B(\bar{x},\bar{y})$, under the S-qualification condition (A_1^3).

We obviously have $A(\bar{x},\bar{y}) \supseteq B(\bar{x},\bar{y})$. To justify the opposite inclusion, pick any $a(\alpha,\beta,\gamma,\zeta) \in A(\bar{x},\bar{y})$ and search for $b(\alpha^o,\beta^o,\gamma^o,\zeta^o) \in B(\bar{x},\bar{y})$ such that $a(\alpha,\beta,\gamma,\zeta) = b(\alpha^o,\beta^o,\gamma^o,\zeta^o)$. If the latter equality were to hold, we would get

$$\begin{cases} \nabla L_o(\bar{x},\bar{y},\alpha-\alpha^o,\beta-\beta^o,\gamma-\gamma^o,\zeta-\zeta^o) = 0, \\ \alpha_I^o - \alpha_I^o = 0,\, \gamma_\nu - \gamma_\nu^o = 0,\, \zeta_\eta - \zeta_\eta^o = 0. \end{cases}$$

Thus it follows from (A_1^3) that $\gamma_\theta^o = \gamma_\theta$ and $\zeta_\theta^o = \zeta_\theta$. To conclude the proof, choose $\alpha^o := \alpha$, $\beta^o := \beta$, $\gamma_{\theta^c}^o := \gamma_{\theta^c}$ and $\zeta_{\theta^c}^o := \zeta_{\theta^c}$ with $\theta^c := \{i = 1,\ldots,d\} \setminus \theta$. □

We can see from the proof that it can be repeated while using the C-type upper bound in Subsection 4.1.2 instead of the M-one. This shows that under the assumption (A_1^3) all the S-type, M-type, and C-type upper bounds for $\partial\mu^c$ are the same. It places us in the situation similar to that already recognized in the context of the various types of stationarity concepts known for OPCCs: they agree with each other under an appropriate assumption.

We also mention two possibilities for the local Lipschitz continuity of μ^c in the framework of Theorem 4.1.8. The first one is either to replace (A_1^1) by (A_3^1) or to add (A_2^1) to the assumptions; cf. (4.10) and Theorem 4.1.2. The second possibility is to replace (A_1^3) by the following stronger qualification condition:

$$(A_2^3) \qquad \left.\begin{array}{r} \nabla_y L_o(\bar{x},\bar{y},\alpha,\beta,\gamma,\zeta) = 0 \\ \alpha_I = 0,\, \gamma_\nu = 0,\, \zeta_\eta = 0 \end{array}\right\} \Longrightarrow \alpha = 0,\, \beta = 0,\, \gamma = 0,\, \zeta = 0. \tag{4.24}$$

The latter condition corresponding to the well-known MPEC-LICQ for the parametric OPCC (4.2) has the advantage, in the framework of Theorem 4.1.8 (i), to ensure even more than the Lipschitz continuity of μ^c; namely, its *strict differentiability* as stated in the next corollary.

Corollary 4.1.9 (S-type sensitivity analysis for OPCC value functions under the MPEC-LICQ). *Assume that S_o^c is inner semicontinuous at the point $(\bar{x},\bar{y})$, where the qualification condition (A_2^3) is also satisfied. Then the value function μ^c is strictly differentiable at $\bar{x}$ with*

$$\nabla\mu^c(\bar{x}) = \nabla_x L(\bar{x},\bar{y},\alpha,\beta,\gamma,\zeta),$$

where $(\alpha,\beta,\gamma,\zeta)$ is the unique multiplier of the set $\Lambda_y^{cs}(\bar{x},\bar{y})$.

Proof. We can see that the set on the right-hand-side of the inclusion in Theorem 4.1.8 (i) is a singleton; hence $\partial\mu^c(\bar{x})$ is a singleton as well. Since the value function μ^c is surely locally Lipschitzian around $\bar{x}$ under the MPEC-LICQ (4.24), the latter uniqueness ensures its strict differentiability at this point; see Subsection 2.1.1. □

In case of (ii) we additionally need $S_o^c(\bar{x})$ to be a singleton to ensure the strict differentiability of μ^c at $\bar{x}$. The latter corresponds to the framework provided by Hu and Ralph [67], and hence it shows (see Remark 4.1.4) that the assumptions imposed in [67] imply the inner semicontinuity of the set-valued mapping S_o^c at the solution point. Note also that assertion (ii) of Theorem 4.1.8 closely relates to the corresponding result of [76] obtained in a particular case from a different perspective. Finally, we mention that the S-qualification condition (A_1^3) does *not imply* the equalities between the multiplier sets in (4.7); for this we need the stronger assumption consisting in replacing the gradients of g, h, G, and H in (A_1^3) by their partial gradients with respect to the y-variable.

4.2 Sensitivity analysis of OPEC value functions

This section is devoted to the study of the following parametric *optimization problem with generalized equation constraints* (OPEC):

$$\min_y \{F(x,y)|\, 0 \in h(x,y) + N_{K(x)}(y)\}, \tag{4.25}$$

where $F : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$, $h : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^p$ are continuously differentiable functions, and K denotes the set-valued mapping (moving set) defined as in (1.4), with g also continuously differentiable and $g(x,.)$ convex for all $x \in \mathbb{R}^n$. Note that model (4.25) is written in the form of *quasi-variational inequalities* described by the normal cone to moving sets; see, e.g., [93] and the references therein. On the other hand, problem (4.25) is closely related to the MPCC considered in the previous section. Indeed, it has been well recognized that the complementarity system defining the feasible set of problem (4.2) can equivalently be written as

$$0 \in -G(x,y) + N_{\mathbb{R}^d_-}(-H(x,y)),$$

which is in the form of OPEC constraints in (4.25) with the normal cone to the constant nonpositive orthant. In the other direction, by replacing the normal cone in (4.25) by its expression in (3.7), under a certain CQ, we get a particular case of problem (4.2). Despite this close link, sensitivity analysis of the OPEC optimal value function

$$\mu^e(x) := \min_y \{F(x,y)|\, 0 \in h(x,y) + N_{K(x)}(y)\} \tag{4.26}$$

associated with problem (4.25) in its given form is of independent interest. Indeed, in this way we obtain different estimates for the limiting subdifferential of the two-level value function φ_o from (1.7), which is of our main attention in the next chapter, where this issue will be comprehensively discussed.

To present our main result in this section on the generalized differentiation and Lipschitz continuity of the value function μ^e, we proceed similarly to Section 4.1 and consider first the *feasible solution map* of the parametric generalized equation in (4.25) defined by

$$S^e(x) := \{y \in \mathbb{R}^m |\, 0 \in h(x,y) + N_{K(x)}(y)\}. \tag{4.27}$$

A detailed study of the robust Lipschitzian stability of (4.27) based on the coderivative analysis has been carried out by Mordukhovich and Outrata [93]. Note that the work in [93] heavily relies on an estimate of the coderivative of normal cone mappings $(x,y) \rightrightarrows N_{K(x)}(y)$ given therein. Before introducing the rules to be used here (which emerged from [93]), some notation is necessary to simplify the presentation. Provided there is no confusion with (3.9) in this section we define $\mathscr{L}(x,y,u) := h(x,y) + \nabla_y g(x,y)^\top u$ and consider the corresponding set of *lower-level Lagrange multipliers* similarly to (3.11) with the same notation $\Lambda(\bar{x},\bar{y})$. Consider also the following *special M-type* multiplier set that plays an important role in the sequel:

$$\begin{aligned}\Lambda^{em}(\bar{x},\bar{y},\bar{u},v) := \{(\beta,\gamma)|\quad & v + \textstyle\sum_{i=1}^p \beta_i \nabla g_i(\bar{x},\bar{y}) + \sum_{l=1}^m \gamma_l \nabla_{x,y}\mathscr{L}_l(\bar{x},\bar{y},\bar{u}) = 0,\\ & \nabla_y g_\nu(\bar{x},\bar{y})\gamma = 0,\ \beta_\eta = 0,\\ & \forall i \in \theta,\ (\beta_i > 0 \wedge \nabla_y g_i(\bar{x},\bar{y})\gamma > 0) \vee \beta_i(\nabla_y g_i(\bar{x},\bar{y})\gamma) = 0\},\end{aligned} \tag{4.28}$$

where the index sets η, θ and ν are defined in (3.22). Similarly to Section 4.1, we further define $\Lambda_y^{em}(\bar{x},\bar{y},\bar{u},v)$, with $v \in \mathbb{R}^m$, by replacing the first equation in the right-hand-side of (4.28) with $v + \sum_{i=1}^p \beta_i \nabla_y g_i(\bar{x},\bar{y}) + \sum_{l=1}^m \gamma_l \nabla_y \mathscr{L}_l(\bar{x},\bar{y},\bar{u}) = 0$ and then set $\Lambda_y^{em}(\bar{x},\bar{y},\bar{u}) := \Lambda_y^{em}(\bar{x},\bar{y},\bar{u},\nabla_y F(\bar{x},\bar{y}))$. The following *EM-qualification conditions* deduced from [93] can be formulated as:

$$\begin{array}{ll} (A_1^4) & \forall \bar{u} \in \Lambda(\bar{x},\bar{y}) : [\sum_{i=1}^p \beta_i \nabla g_i(\bar{x},\bar{y}) = 0,\, \beta_\eta = 0] \Longrightarrow \beta = 0, \\ (A_2^4) & \forall \bar{u} \in \Lambda(\bar{x},\bar{y}) : (\beta,\gamma) \in \Lambda^{em}(\bar{x},\bar{y},\bar{u},0) \Longrightarrow \beta = 0,\, \gamma = 0, \\ (A_3^4) & [\bar{u} \in \Lambda(\bar{x},\bar{y}),\, (\beta,\gamma) \in \Lambda_y^{em}(\bar{x},\bar{y},\bar{u},0)] \Longrightarrow \sum_{i=1}^p \beta_i \nabla_x g_i(\bar{x},\bar{y}) + \sum_{l=1}^m \gamma_l \nabla_x \mathscr{L}_l(\bar{x},\bar{y},\bar{u}) = 0, \\ (A_4^4) & [\bar{u} \in \Lambda(\bar{x},\bar{y}),\, (\beta,\gamma) \in \Lambda_y^{em}(\bar{x},\bar{y},\bar{u},0)] \Longrightarrow \beta = 0,\, \gamma = 0. \end{array} \tag{4.29}$$

Note that in (A_1^4), the index set η depends on $\bar{u}$ by means of its definition in (3.22). It is easy to observe the relationships between these qualification conditions:

$$(A_3^4) \Longleftarrow (A_4^4) \Longrightarrow (A_2^4),\, (A_1^4).$$

We are now ready to establish the main result of this section, where S_o^e denotes the *optimal solution map* to the parametric optimization problem (4.25) given by

$$S_o^e(x) := \{y \in S^e(x) | \, F(x,y) - \mu^e(x) \le 0\}. \tag{4.30}$$

Theorem 4.2.1 (M-type sensitivity analysis for OPEC value functions)**.** *The following assertions hold for the value function μ^e from* (4.27)*:*

(i) *Let the optimal solution map S_o^e* (4.30) *be inner semicontinuous at the point $(\bar{x},\bar{y})$, where the lower-level regularity* (3.65) *and the conditions (A_1^4)–(A_2^4) hold. Then we have the subdifferential estimate*

$$\partial \mu^e(\bar{x}) \subseteq \bigcup_{\bar{u} \in \Lambda(\bar{x},\bar{y})} \bigcup_{(\beta,\gamma) \in \Lambda_y^{em}(\bar{x},\bar{y},\bar{u})} \Big\{ \nabla_x F(\bar{x},\bar{y}) + \sum_{i=1}^p \beta_i \nabla_x g(\bar{x},\bar{y}) + \sum_{l=1}^p \gamma_l \nabla_x \mathscr{L}_l(\bar{x},\bar{y},\bar{u}) \Big\}.$$

If in addition (A_3^4) is satisfied at $(\bar{x},\bar{y})$, then μ^e is Lipschitz continuous around $\bar{x}$.

(ii) *Let S_o^e* (4.30) *be inner semicompact at $\bar{x}$, and let (A_1^4)–(A_2^4) and lower-level regularity* (3.65) *be satisfied at $(\bar{x},\bar{y})$ for all $\bar{y} \in S_o^e(\bar{x})$. Then we have the subdifferential estimate*

$$\partial \mu^e(\bar{x}) \subseteq \bigcup_{\bar{y} \in S_o(\bar{x})} \bigcup_{\bar{u} \in \Lambda(\bar{x},\bar{y})} \bigcup_{(\beta,\gamma) \in \Lambda_y^{em}(\bar{x},\bar{y},\bar{u})} \Big\{ \nabla_x F(\bar{x},\bar{y}) + \textstyle\sum_{i=1}^p \beta_i \nabla_x g(\bar{x},\bar{y}) + \sum_{l=1}^p \gamma_l \nabla_x \mathscr{L}_l(\bar{x},\bar{y},\bar{u}) \Big\}.$$

If (A_3^4) is also satisfied at $(\bar{x},\bar{y})$ for all $\bar{y} \in S_o^e(\bar{x})$, then μ^e is Lipschitz continuous around $\bar{x}$.

Proof. We justify only assertion (i); the one in (ii) can be proved similarly. Since F is continuously differentiable and S_o^e is inner semicontinuous, it follows from Theorem 4.0.1 (i) that

$$\partial \mu^e(\bar{x}) \subseteq \nabla_x F(\bar{x},\bar{y}) + D^* S^e(\bar{x},\bar{y})(\nabla_y F(\bar{x},\bar{y})). \tag{4.31}$$

Applying further [93, Theorem 4.3] to the solution map S^e (4.27) and taking into account that the *EM*-qualification conditions (A_1^4)–(A_2^4) together with the lower-level regularity (3.65) are satisfied, we get the coderivative estimate

$$D^* S^e(\bar{x},\bar{y})(\nabla_y F(\bar{x},\bar{y})) \subseteq \bigcup_{\bar{u} \in \Lambda(\bar{x},\bar{y})} \bigcup_{(\beta,\gamma) \in \Lambda_y^{em}(\bar{x},\bar{y},\bar{u})} \Big\{ \sum_{i=1}^p \beta_i \nabla_x g_i(\bar{x},\bar{y}) + \sum_{l=1}^m \gamma_l \nabla_x \mathscr{L}_l(\bar{x},\bar{y},\bar{u}) \Big\}. \tag{4.32}$$

Then the upper estimate of the basic subdifferential of μ^e in the theorem follows by combining (4.31) and (4.32). The local Lipschitz continuity of μ^e around $\bar{x}$ also follows from Theorem 4.0.1 (i) by recalling [93] that S^e is Lipschitz-like around $(\bar{x},\bar{y})$ provided we add (A_3^4) to the previous assumptions. □

To the best of our knowledge, the first result in the direction of Theorem 4.2.1(ii) goes back to Lucet and Ye [76], where a similar subdifferential estimate was obtained under a growth hypothesis (implying the inner semicompactness of S_o^e) for a particular case of the problem under consideration. Note however that their result deals only with the case where K is *independent* of x. Assertion (i) of Theorem 4.2.1 clearly provides a tighter subdifferential upper bound under the inner semicontinuity assumption. We also mention the work by Mordukhovich et al. [92] in the framework where the regular and limiting subdifferentials of μ^e are estimated in the case of

$$S^e(x) := \{y|\, 0 \in h(x,y) + Q(x,y)\}$$

in (4.26) with a general set-valued mapping $Q(x,y)$ not specified to our setting $Q(x,y) := N_{K(x)}(y)$ in terms of the initial data of (3.7).

Finally, note that according to [93], the qualification condition (A_1^4) in Theorem 4.2.1 can be replaced by the weaker *calmness property* of the set-valued mapping M (3.144) at $(0,\bar{x},\bar{y},\bar{u})$, for all $\bar{u} \in \Lambda(\bar{x},\bar{y})$. Similarly, condition (A_3^4) can be replaced by the calmness property of the mapping $P(v_1,v_2) := \{(x,y,u) \in M(v_1)|\, \mathscr{L}(x,y,u) + v_2 = 0\}$ at $(0,0,\bar{x},\bar{y},\bar{u})$, for all $\bar{u} \in \Lambda(\bar{x},\bar{y})$. Both calmness assumptions are automatical when the mappings g and $(x,y) \mapsto \nabla_y f(x,y)$ are linear.

5 Original optimistic bilevel programming problem

5.1 Sensitivity analysis of two-level value functions

Our main concern in this section is to conduct a local sensitivity analysis of the two-level optimal value function

$$\varphi_o(x) := \min_y \{F(x,y)|y \in S(x)\}, \tag{5.1}$$

where the lower-level solution set-valued mapping S is defined in (1.2). In the next subsection, we explore all the three approaches (M, C and S) to this issue discussed in the previous chapter. The lower-level value function representation of the mapping S is considered in Subsection 6.2.4. Throughout the section, the following solution set-valued mapping of the parametric problem associated to (5.1) will play an important role:

$$S_o(x) = \{y \in S(x)|\ F(x,y) \le \varphi_o(x)\}. \tag{5.2}$$

In order for the two-level value function in (5.1) to be well-defined, it is assumed throughout this chapter that $S_o(x) \neq \emptyset$ for all $x \in X$.

5.1.1 KKT reformulation of the two-level value function

Here, a KKT/OPCC reformulation of the two-level value function (5.1) is essential for the analysis. Note that from now on we come back to the expression of the lower-level Lagrange function $\mathscr{L}$ in (3.9) which corresponds to the one in Section 4.2 while setting $h(x,y) := \nabla_y f(x,y)$. Hence, for this reason the set of Lagrange multipliers $\Lambda(\bar{x},\bar{y})$ of the lower-level problem will further on be the one in (3.11).

Lemma 5.1.1 (KKT reformulation of the two-level value function)**.** *Let $x \in X$ and assume that $f(x,.)$ and $g(x,.)$ are convex. Furthermore, we suppose that the Slater CQ* (3.6) *is satisfied at x. Then we have*

$$\varphi_o(x) = \min_{y,u}\{F(x,y)|\ \mathscr{L}(x,y,u) = 0,\ u \ge 0,\ g(x,y) \le 0,\ u^\top g(x,y) = 0\}. \tag{5.3}$$

Proof. Fix $\bar{x} \in X$ where the lower-level problem (1.3) is convex and the Slater CQ (3.6) holds. Further let $\bar{y}$ be a global optimal solution to the problem $\min_y\{F(\bar{x},y)|\,y \in S(\bar{x})\}$. Then we have the relationships

$$\begin{aligned}\varphi_o(\bar{x}) = \ & F(\bar{x},\bar{y}),\\ \le \ & F(\bar{x},y) : \forall y \in S(\bar{x}),\\ \le \ & F(\bar{x},y) : \forall y \text{ with } 0 \in \nabla_y f(\bar{x},y) + N_{K(\bar{x})}(y) \text{ (by convexity of } f(\bar{x},.) \text{ and } g(\bar{x},.)),\\ \le \ & F(\bar{x},y) : \forall (y,u) \text{ with } \mathscr{L}(\bar{x},y,u),\ u \ge 0, g(\bar{x},y) \le 0,\ u^\top g(\bar{x},y) = 0,\end{aligned}$$

where the last inequality is due to the normal cone representation (3.7) while taking into account that the the Slater CQ (3.6) holds at $\bar{x}$. □

Having this transformation of the two-level value function φ_o, at least two observations can be made. First we note that for each $x \in X$ the value of $\varphi_o(x)$ is obtained from a *global solution* to the parametric problem

$$\min_{y,u}\{F(x,y)|\ \mathscr{L}(x,y,u) = 0,\ u \ge 0,\ g(x,y) \le 0,\ u^\top g(x,y) = 0\}. \tag{5.4}$$

Thus the major difficulty arising when establishing the link between local solutions of the auxiliary problem (P) and its KKT reformulation (see [24] for details) does *not* appear here. Secondly, the presence of the complementarity constraints $u \geq 0$, $g(x,y) \leq 0$, $u^\top g(x,y) = 0$ in (5.4) leads to the violation of the MFCQ as already observed in Chapter 3 (Section 3.3). However the results of Chapter 4 can be applied. To proceed, consider the feasible solution map associated with (5.4) defined by

$$S^h(x) := \{(y,u) | \mathscr{L}(x,y,u) = 0,\, u \geq 0,\, g(x,y) \leq 0,\, u^\top g(x,y) = 0\} \tag{5.5}$$

and the optimal solution map of problem (5.4) given by

$$S^h_o(x) := \{(y,u) \in S^h(x) | F(x,y) \leq \varphi_o(x)\}. \tag{5.6}$$

Furthermore, consider the multiplier sets $\Lambda^{em}(\bar{x},\bar{y},\bar{u},0)$, $\Lambda^{em}_y(\bar{x},\bar{y},\bar{u},0)$ and $\Lambda^{em}_y(\bar{x},\bar{y},\bar{u})$ defined in Section 4.2 of Chapter 4, cf. (4.28). In the vein of the rules in (4.9), the following related M-qualification conditions are used in the next theorem:

$$\begin{array}{ll} (A^m_1) & (\beta,\gamma) \in \Lambda^{em}(\bar{x},\bar{y},\bar{u},0) \Longrightarrow \beta = 0,\, \gamma = 0,\\ (A^m_2) & (\beta,\gamma) \in \Lambda^{em}_y(\bar{x},\bar{y},\bar{u},0) \Longrightarrow \sum_{i=1}^p \beta_i \nabla_x g_i(\bar{x},\bar{y}) + \sum_{l=1}^m \gamma_l \nabla_x \mathscr{L}_l(\bar{x},\bar{y},\bar{u}) = 0,\\ (A^m_3) & (\beta,\gamma) \in \Lambda^{em}_y(\bar{x},\bar{y},\bar{u},0) \Longrightarrow \beta = 0,\, \gamma = 0. \end{array} \tag{5.7}$$

Theorem 5.1.2 (M-type sensitivity analysis for the two-level value function)**.** *Let the lower-level problem* (1.3) *be convex, and assume that the Slater CQ* (3.6) *holds at* $\bar{x}$*. Then the following assertions are satisfied:*

(i) *Assume that the optimal solution map* S^h_o (5.6) *is inner semicontinuous at* $(\bar{x},\bar{y},\bar{u})$*, where condition* (A^m_1) *also holds. Then the limiting subdifferential of* φ_o *is estimated by*

$$\partial\varphi_o(\bar{x}) \subseteq \bigcup_{(\beta,\gamma)\in\Lambda^{em}_y(\bar{x},\bar{y},\bar{u})} \Big\{\nabla_x F(\bar{x},\bar{y}) + \sum_{i=1}^p \beta_i \nabla_x g_i(\bar{x},\bar{y}) + \sum_{l=1}^m \gamma_l \nabla_x \mathscr{L}_l(\bar{x},\bar{y},\bar{u})\Big\}. \tag{5.8}$$

Furthermore, φ_o *is Lipschitz continuous around* $\bar{x}$ *provided that* (A^m_2) *is also satisfied at* $(\bar{x},\bar{y},\bar{u})$*.*

(ii) *Let the set-valued mapping* S^h_o (5.6) *be inner semicompact at* $\bar{x}$*, while condition* (A^m_1) *holds at* $(\bar{x},\bar{y},\bar{u})$ *for all* $(\bar{y},\bar{u}) \in S^h_o(\bar{x})$*, then*

$$\partial\varphi_o(\bar{x}) \subseteq \bigcup_{(\bar{y},\bar{u})\in S^h_o(\bar{x})}\ \bigcup_{(\beta,\gamma)\in\Lambda^{em}_y(\bar{x},\bar{y},\bar{u})} \Big\{\nabla_x F(\bar{x},\bar{y}) + \sum_{i=1}^p \beta_i \nabla_x g_i(\bar{x},\bar{y}) + \sum_{l=1}^m \gamma_l \nabla_x \mathscr{L}_l(\bar{x},\bar{y},\bar{u})\Big\}. \tag{5.9}$$

If in addition, condition (A^m_2) *is satisfied at* $(\bar{x},\bar{y},\bar{u})$ *for all* $(\bar{y},\bar{u}) \in S^h_o(\bar{x})$*, then* φ_o *is Lipschitz continuous around* $\bar{x}$*.*

Proof. We provide the proof for (i) since (ii) can be obtained similarly. By setting $y := (y,u)$ in the framework of Theorem 4.1.2, we simply need to specify the various multiplier sets therein to our setting. It follows from Lemma 5.1.1 that

$$\begin{aligned} \varphi_o(x) = \min_{y,u}\{F(x,y) |\quad & h(x,y,u) = 0,\\ & G(x,y,u) \geq 0,\, H(x,y,u) \geq 0,\, G(x,y,u)^\top H(x,y,u) = 0\}, \end{aligned} \tag{5.10}$$

where $h(x,y,u) := \mathscr{L}(x,y,u)$, $G(x,y,u) := u$, and $H(x,y,u) := -g(x,y)$. Thus we have from Section 4.1 that

$$\begin{aligned} \Lambda^{cm}(\bar{x},\bar{y},\bar{u},v_1,v_2,v_3) = \{(\beta,\gamma,\zeta) |\quad & \zeta_\nu = 0, \beta_\eta = 0,\\ & (\zeta_i < 0,\, \beta_i < 0) \vee (\beta_i\zeta_i = 0), \forall i \in \theta,\\ & v_1 + \textstyle\sum_{l=1}^m \gamma_l \nabla_x \mathscr{L}(\bar{x},\bar{y},\bar{u}) - \sum_{i=1}^p \beta_i \nabla_x g_i(\bar{x},\bar{y}) = 0,\\ & v_2 + \textstyle\sum_{l=1}^m \gamma_l \nabla_y \mathscr{L}(\bar{x},\bar{y},\bar{u}) - \sum_{i=1}^p \beta_i \nabla_y g_i(\bar{x},\bar{y}) = 0,\\ & v_3 + \nabla_y g(\bar{x},\bar{y})\gamma + \zeta = 0\}. \end{aligned} \tag{5.11}$$

Normally, considering the notational conventions in Section 4.1, we should have the multipliers β, γ and ζ attached to the functions h, G and H, respectively, in (5.10). But in order to have a clear picture on the relationships between the developments here and those in Sections 3.3, 3.4 and 4.2, we adopt the notations above.

Now note that from (5.11) we have $\zeta = -\nabla_y g(\bar{x},\bar{y})\gamma$ by setting $v_3 := 0$ in the relationship $v_3 + \nabla_y g(\bar{x},\bar{y})\gamma + \zeta = 0$. Thus

$$\begin{aligned}\Lambda^{cm}(\bar{x},\bar{y},\bar{u},0) &= \{(\beta,\gamma,-\nabla_y g(\bar{x},\bar{y})\gamma)|\ \nabla_y g_\nu(\bar{x},\bar{y})\gamma = 0,\ \beta_\eta = 0,\\ &\qquad (\nabla_y g_i(\bar{x},\bar{y})\gamma > 0,\ \beta_i < 0) \vee \beta_i(\nabla_y g_i(\bar{x},\bar{y})\gamma) = 0,\ \forall i \in \theta,\\ &\qquad \textstyle\sum_{l=1}^m \gamma_l \nabla_{x,y}\mathcal{L}(\bar{x},\bar{y},\bar{u}) - \sum_{i=1}^p \beta_i \nabla g_i(\bar{x},\bar{y}) = 0\}\\ &= \{(\beta,\gamma,-\nabla_y g(\bar{x},\bar{y})\gamma)|\ (-\beta,\gamma) \in \Lambda^{em}(\bar{x},\bar{y},\bar{u},0)\}.\end{aligned}$$

From the latter equality, one can easily check that (A_1^m) is a sufficient condition for the following implication to hold:

$$(\beta,\gamma,\zeta) \in \Lambda^{cm}(\bar{x},\bar{y},\bar{u},0) \Longrightarrow \beta = 0,\ \gamma = 0,\ \zeta = 0. \tag{5.12}$$

Under the qualification condition (5.12), we have from Theorem 4.1.2 (i) that the basic subdifferential of the two-level value function φ_o (5.1) can be estimated by

$$\partial\varphi_o(\bar{x}) \subseteq \bigcup_{(\beta,\gamma,\zeta)\in\Lambda_y^{cm}(\bar{x},\bar{y},\bar{u})} \left\{\nabla_x F(\bar{x},\bar{y}) - \sum_{i=1}^p \beta_i \nabla_x g_i(\bar{x},\bar{y}) + \sum_{l=1}^m \gamma_l \nabla_x \mathcal{L}_l(\bar{x},\bar{y},\bar{u})\right\}, \tag{5.13}$$

where, by proceeding as above, the multiplier set $\Lambda_y^{cm}(\bar{x},\bar{y},\bar{u})$ can be written as follows

$$\Lambda_y^{cm}(\bar{x},\bar{y},\bar{u}) := \Lambda_y^{cm}(\bar{x},\bar{y},\bar{u},\nabla_y F(\bar{x},\bar{y}),0) = \{(\beta,\gamma,-\nabla_y g(\bar{x},\bar{y})\gamma)|\ (-\beta,\gamma) \in \Lambda_y^{em}(\bar{x},\bar{y},\bar{u})\}.$$

Thus we have $(\beta,\gamma,\zeta) \in \Lambda_y^{cm}(\bar{x},\bar{y},\bar{u})$ if and only if $\zeta = -\nabla_y g(\bar{x},\bar{y})\gamma$ and $(-\beta,\gamma) \in \Lambda_y^{em}(\bar{x},\bar{y},\bar{u})$. Inclusion (5.8) then follows while noting that the term $\nabla_x F(\bar{x},\bar{y}) - \sum_{i=1}^p \beta_i \nabla_x g_i(\bar{x},\bar{y}) + \sum_{l=1}^m \gamma_l \nabla_x \mathcal{L}_l(\bar{x},\bar{y},\bar{u})$ does not depend on the multiplier ζ.

Also proceeding as above for condition (A_1^m), it is a simple exercise to check that (A_2^m) implies that we have

$$(\beta,\gamma,\zeta) \in \Lambda^{cm}(\bar{x},\bar{y},\bar{u},0) \Longrightarrow -\sum_{i=1}^p \beta_i \nabla_x g_i(\bar{x},\bar{y}) + \sum_{l=1}^m \gamma_l \nabla_x \mathcal{L}_l(\bar{x},\bar{y},\bar{u}) = 0,$$

which on its turn implies that the two-level value function φ_o is Lipschitz continuous around $\bar{x}$, while applying by Theorem 4.1.2 (i) via the reformulation of φ_o in (5.10). In a similar way, the qualification condition (A_3^m) implies the fulfilment of the counterpart of (A_3^1) in (4.9). □

We can similarly consider the *C-type* multiplier sets $\Lambda^{ec}(\bar{x},\bar{y},\bar{u},0)$, $\Lambda_y^{ec}(\bar{x},\bar{y},\bar{u},0)$ and $\Lambda_y^{ec}(\bar{x},\bar{y},\bar{u})$, which are obtained by replacing condition (3.26) in $\Lambda^{em}(\bar{x},\bar{y},\bar{u},0)$, $\Lambda^{em}(\bar{x},\bar{y},\bar{u},0)$ and $\Lambda_y^{em}(\bar{x},\bar{y},\bar{u})$ by (3.29). Then the upper bounds of the limiting subdifferential via C-type multipliers and the local Lipschitz continuity of the two-level value function φ_o under the C-type conditions can be derived as in Theorem 5.1.2 with $\Lambda^{em}(\bar{x},\bar{y},\bar{u},0)$, $\Lambda_y^{em}(\bar{x},\bar{y},\bar{u},0)$ and $\Lambda_y^{em}(\bar{x},\bar{y},\bar{u})$ replaced by $\Lambda^{ec}(\bar{x},\bar{y},\bar{u},0)$, $\Lambda_y^{ec}(\bar{x},\bar{y},\bar{u},0)$ and $\Lambda_y^{ec}(\bar{x},\bar{y},\bar{u})$, respectively. Similarly to (5.7), we also used the C-qualification conditions:

$$\begin{array}{ll}(A_1^c) & (\beta,\gamma) \in \Lambda^{ec}(\bar{x},\bar{y},\bar{u},0) \Longrightarrow \beta = 0,\ \gamma = 0,\\ (A_2^c) & (\beta,\gamma) \in \Lambda_y^{ec}(\bar{x},\bar{y},\bar{u},0) \Longrightarrow \sum_{i=1}^p \beta_i \nabla_x g_i(\bar{x},\bar{y}) + \sum_{l=1}^m \gamma_l \nabla_x \mathcal{L}_l(\bar{x},\bar{y},\bar{u}) = 0,\\ (A_3^c) & (\beta,\gamma) \in \Lambda_y^{ec}(\bar{x},\bar{y},\bar{u},0) \Longrightarrow \beta = 0,\ \gamma = 0.\end{array} \tag{5.14}$$

Theorem 5.1.3 (C-type sensitivity analysis of the two-level value function)**.** *Let the lower-level problem* (1.3) *be convex, while the Slater CQ* (3.6) *holds at* $\bar{x}$. *Then the following assertions are satisfied:*

(i) *Assume that the optimal solution map* S_o^h (5.6) *is inner semicontinuous at* $(\bar{x},\bar{y},\bar{u})$*, where condition* (A_1^c) *also holds. Then the limiting subdifferential of* φ_o *is estimated by*

$$\partial\varphi_o(\bar{x}) \subseteq \bigcup_{(\beta,\gamma)\in\Lambda_y^{ec}(\bar{x},\bar{y},\bar{u})} \Big\{\nabla_x F(\bar{x},\bar{y}) + \sum_{i=1}^{p}\beta_i\nabla_x g_i(\bar{x},\bar{y}) + \sum_{l=1}^{m}\gamma_l\nabla_x\mathscr{L}_l(\bar{x},\bar{y},\bar{u})\Big\}. \tag{5.15}$$

Furthermore, φ_o *is Lipschitz continuous around* $\bar{x}$ *provided that* (A_2^c) *is also satisfied at* $(\bar{x},\bar{y},\bar{u})$.

(ii) *Let the set-valued mapping* S_o^h (5.6) *be inner semicompact at* $\bar{x}$*, while condition* (A_1^c) *holds at* $(\bar{x},\bar{y},\bar{u})$ *for all* $(\bar{y},\bar{u}) \in S_o^h(\bar{x})$*, then*

$$\partial\varphi_o(\bar{x}) \subseteq \bigcup_{(\bar{y},\bar{u})\in S_o^h(\bar{x})}\ \bigcup_{(\beta,\gamma)\in\Lambda_y^{ec}(\bar{x},\bar{y},\bar{u})} \Big\{\nabla_x F(\bar{x},\bar{y}) + \sum_{i=1}^{p}\beta_i\nabla_x g_i(\bar{x},\bar{y}) + \sum_{l=1}^{m}\gamma_l\nabla_x\mathscr{L}_l(\bar{x},\bar{y},\bar{u})\Big\}. \tag{5.16}$$

If in addition, condition (A_2^c) *is satisfied at* $(\bar{x},\bar{y},\bar{u})$ *for all* $(\bar{y},\bar{u}) \in S_o^h(\bar{x})$*, then* φ_o *is Lipschitz continuous around* $\bar{x}$.

Proof. Similarly to the proof of Theorem 5.1.2, it follows from Theorem 4.1.7 while considering the representation of the two-level value function in Lemma 5.1.1. □

To consider *S-type* upper bounds for the subdifferential of the two-level value function φ_o, define the set $\Lambda_y^{es}(\bar{x},\bar{y},\bar{u})$ similarly to $\Lambda_y^{em}(\bar{x},\bar{y},\bar{u})$ with replacing condition (3.26) by (3.30) and arrive at the following sensitivity result. The following S-qualification conditions are needed:

$$\begin{array}{ll} (A_1^s) & \left.\begin{array}{r}\sum_{i=1}^{p}\beta_i\nabla g_i(\bar{x},\bar{y}) + \sum_{l=1}^{m}\gamma_l\nabla_{x,y}\mathscr{L}_l(\bar{x},\bar{y},\bar{u}) = 0\\ \nabla_y g_\nu(\bar{x},\bar{y})\gamma = 0,\ \beta_\eta = 0\end{array}\right\} \Longrightarrow \beta_\theta = 0,\ \nabla_y g_\theta(\bar{x},\bar{y})\gamma = 0,\\ (A_2^s) & \left.\begin{array}{r}\sum_{i=1}^{p}\beta_i\nabla_y g_i(\bar{x},\bar{y}) + \sum_{l=1}^{m}\gamma_l\nabla_y\mathscr{L}_l(\bar{x},\bar{y},\bar{u}) = 0\\ \nabla_y g_\nu(\bar{x},\bar{y})\gamma = 0,\ \beta_\eta = 0\end{array}\right\} \Longrightarrow \beta = 0,\ \gamma = 0.\end{array} \tag{5.17}$$

Theorem 5.1.4 (S-type sensitivity analysis for the two-level value function φ_o)**.** *Let the lower-level problem* (1.3) *be convex, while the Slater CQ* (3.6) *holds at* $\bar{x}$*. Then the following assertions hold:*

(i) *Assume that the optimal solution map* S_o^h (5.6) *is inner semicontinuous at* $(\bar{x},\bar{y},\bar{u})$*, where conditions* (A_1^m) *and* (A_1^s) *are also satisfied. Then the limiting subdifferential of* φ_o *is estimated by*

$$\partial\varphi_o(\bar{x}) \subseteq \bigcup_{(\beta,\gamma)\in\Lambda_y^{es}(\bar{x},\bar{y},\bar{u})} \Big\{\nabla_x F(\bar{x},\bar{y}) + \sum_{i=1}^{p}\beta_i\nabla_x g_i(\bar{x},\bar{y}) + \sum_{l=1}^{m}\gamma_l\nabla_x\mathscr{L}_l(\bar{x},\bar{y},\bar{u})\Big\}. \tag{5.18}$$

Furthermore, if conditions (A_1^m) *and* (A_1^s) *are replaced by* (A_2^s)*, then the two-level value function* φ_o *is strictly differentiable at* $\bar{x}$*, and its derivative is given by*

$$\nabla\varphi_o(\bar{x}) = \nabla_x F(\bar{x},\bar{y}) + \sum_{i=1}^{p}\beta_i\nabla_x g_i(\bar{x},\bar{y}) + \sum_{l=1}^{m}\gamma_l\nabla_x\mathscr{L}_l(\bar{x},\bar{y},\bar{u}), \tag{5.19}$$

where (β,γ) *is the unique multiplier vector of the set* $\Lambda_y^{es}(\bar{x},\bar{y},\bar{u})$.

(ii) *Let the set-valued mapping* S_o^h (5.6) *be inner semicompact at* $\bar{x}$*, while conditions* (A_1^m) *and* (A_1^s) *hold at* $(\bar{x},\bar{y},\bar{u})$ *for all* $(\bar{y},\bar{u}) \in S_o^h(\bar{x})$*, then we have*

$$\partial\varphi_o(\bar{x}) \subseteq \bigcup_{(\bar{y},\bar{u})\in S_o^h(\bar{x})}\ \bigcup_{(\beta,\gamma)\in\Lambda_y^{es}(\bar{x},\bar{y},\bar{u})} \Big\{\nabla_x F(\bar{x},\bar{y}) + \sum_{i=1}^{p}\beta_i\nabla_x g_i(\bar{x},\bar{y}) + \sum_{l=1}^{m}\gamma_l\nabla_x\mathscr{L}_l(\bar{x},\bar{y},\bar{u})\Big\}. \tag{5.20}$$

Proof. Considering the representation of the two-level value function φ_o in (5.3) (also see (5.10)), it is obvious that (A_1^s) corresponds to the qualification condition (A_1^3) in (4.23). Thus inclusion (5.18) follows from Theorem 4.1.8 (i) while taking into account the estimate of $\partial\varphi_o$ in (5.8). As for the strict differentiability of φ_o, and the derivative in (5.19), they are obtained from Corollary 4.1.9 while noting that assumption (A_2^s) is the counterpart of (A_2^3) (see (4.24)) for the KKT reformulation of φ_o in (5.3).

The estimate of the basic subdifferential in (5.9) is obtained similarly to the one in (5.18) while using Theorem 4.1.8 (ii). □

Although the M-type sensitivity analysis of the two-level value function is self-evident while considering Theorem 4.2.1, we nevertheless state it here in order to have a clear picture of the differences with the sensitivity analysis result in Theorem 5.1.2. To proceed, recall that we now have $h(x,y) := \nabla_y f(x,y)$, then further set $(A_1^e) := (A_1^4)$, $(A_2^e) := (A_2^4)$, $(A_3^e) := (A_3^4)$ and $(A_4^e) := (A_4^4)$. Then we have:

Corollary 5.1.5 (M-type sensitivity analysis of the two-level value function via the OPEC reformulation)**.** *Let the lower-level problem* (1.3) *be convex, then the following assertions hold:*

(i) *Let the optimal solution map S_o* (5.2) *be inner semicontinuous at the point $(\bar{x},\bar{y})$, where the lower-level regularity* (3.65) *and the conditions (A_1^e)–(A_2^e) hold. Then we have the subdifferential estimate*

$$\partial\varphi_o(\bar{x}) \subseteq \bigcup_{\bar{u}\in\Lambda(\bar{x},\bar{y})} \bigcup_{(\beta,\gamma)\in\Lambda_y^{em}(\bar{x},\bar{y},\bar{u})} \Big\{\nabla_x F(\bar{x},\bar{y}) + \sum_{i=1}^{p} \beta_i \nabla_x g(\bar{x},\bar{y}) + \sum_{l=1}^{p} \gamma_l \nabla_x \mathscr{L}_l(\bar{x},\bar{y},\bar{u})\Big\}.$$

If in addition (A_3^e) is satisfied at $(\bar{x},\bar{y})$, then φ_o is Lipschitz continuous around $\bar{x}$.

(ii) *Let S_o* (5.2) *be inner semicompact at $\bar{x}$, and let (A_1^e)–(A_2^e) and the lower-level regularity* (3.65) *be satisfied at $(\bar{x},\bar{y})$ for all $\bar{y}\in S_o(\bar{x})$. Then we have the subdifferential estimate*

$$\partial\varphi_o(\bar{x}) \subseteq \bigcup_{\bar{y}\in S_o(\bar{x})} \bigcup_{\bar{u}\in\Lambda(\bar{x},\bar{y})} \bigcup_{(\beta,\gamma)\in\Lambda_y^{em}(\bar{x},\bar{y},\bar{u})} \Big\{\nabla_x F(\bar{x},\bar{y}) + \textstyle\sum_{i=1}^{p} \beta_i \nabla_x g(\bar{x},\bar{y}) + \sum_{l=1}^{p} \gamma_l \nabla_x \mathscr{L}_l(\bar{x},\bar{y},\bar{u})\Big\}.$$

If (A_3^e) is also satisfied at $(\bar{x},\bar{y})$ for all $\bar{y}\in S_o(\bar{x})$, then φ_o is Lipschitz continuous around $\bar{x}$.

Proof. Under the convexity of the lower-level problem (1.3), we have the following OPEC representation of the two-level value function φ_o (5.1):

$$\varphi_o(x) = \min_y \{F(x,y)|\ 0 \in \nabla_y f(x,y) + N_{K(x)}(y)\} \tag{5.21}$$

while considering the OPEC reformulation of the lower-level solution set-valued mapping in (3.4). The result follows trivially by applying Theorem 4.2.1. □

Remark 5.1.6 (On the differences between the M-type sensitivity analysis of the two-level value function via the KKT and primal KKT/OPEC reformulations)**.** *The difference between the upper estimate of $\partial\varphi_o$ in Theorem 5.1.2 (i) and the one in Corollary 5.1.5 (i) is clear. The latter contains the union over $\Lambda(\bar{x},\bar{y})$, which makes it larger than the one obtained in the former result. For the estimates in Theorem 5.1.2 (ii) and Corollary 5.1.5 (ii), they seem to be more closer in the sense that $\bar{y}\in S_o(\bar{x})$ if and only if there exists $\bar{u}$ such that $(\bar{y},\bar{u}) \in S_o^h(\bar{x}) := \{(y,u)|\ u\in\Lambda(\bar{x},y),\ F(\bar{x},y)\le\varphi_o(\bar{x})\}$. On the qualification conditions, note that there are many similarities, but they are not identical. However, if we consider only the strongest ones (A_3^m) and (A_4^e), we do have the same CQs, but as already noted, the results are not the same in terms of the basic subdifferentials of the two-level value function φ_o. This observation is similar to the one made in Section 3.4 (cf. Remark 3.4.8) on the relationship between the stationarity conditions obtained via the KKT and primal KKT/OPEC approaches. Thus the fundamental reason for this remains the fact that the appearance of the lower-level multipliers $\bar{u}$ in Corollary 5.1.5 is a posteriori while it is a priori in Theorem 5.1.2.*

To conclude this subsection, some comments on the qualifications conditions used here are in order. Following Remark 4.1.3, we conclude that condition (A_1^m) can be replaced by the weaker *calmness* property of the set-valued mapping

$$\Psi(z,\vartheta) := \{(x,y,u)|\mathscr{L}(x,y,u)+z=0,\ (-g_i(x,y),u_i)+\vartheta_i \in \Lambda_i,\ i:=1,\dots,p\}, \tag{5.22}$$

where $\Lambda_i := \{(a,b)\in\mathbb{R}^2|\, a\geq 0,\, b\geq 0,\, ab=0\}$. The latter assumption is automatically satisfied when the mappings $g\colon \mathbb{R}^n\times\mathbb{R}^m\to\mathbb{R}^m$ and $(x,y)\mapsto\nabla_y f(x,y)$ are linear. Condition (A_3^m) can replace the two other assumptions (A_1^m) and (A_2^m). Note however that in the case of the LLVF reformulation considered in the next subsection, the latter sequence of implication between the CQs will not be available. Also observe that comments similar to the above ones can be stated for the C-qualification conditions in (5.14).

5.1.2 LLVF reformulation of the two-level value function

In this subsection we develop the *lower-level value function* (LLVF) approach to sensitivity analysis of the two-level value function φ_o from (5.1). Considering the value function representation of the lower-level solution set-valued mapping S in (3.1), we have the following LLVF reformulation of the two-level value function φ_o:

$$\varphi_o(x) := \min_y\{F(x,y)|\, g(x,y)\leq 0,\ f(x,y)-\varphi(x)\leq 0\}, \tag{5.23}$$

where φ is defined in (3.2). It is worth recalling that calculating the *coderivative* of the optimal solution map S (1.2) is highly significant in our approach. In the current situation, this means computing the limiting normal cone to the graph of S:

$$\operatorname{gph} S = \{(x,y)\in\Omega|\, f(x,y)-\varphi(x)\leq 0\}\ \text{ with }\ \Omega := \{(x,y)\mid g(x,y)\leq 0\} \tag{5.24}$$

in terms of the initial data. To proceed in this way by using the conventional results of the generalized differential calculus [89] requires the fulfillment of the *basic qualification condition*, which reads in this case as (3.40) where the value of Ω in (3.38) is replaced the one in (5.24). However, it follows from Theorem 3.2.1 that the corresponding condition does not hold. The following weaker assumption helps circumventing this difficulty:

(A_1^v) The mapping $\Xi(v) := \{(x,y)\in\Omega|\ f(x,y)-\varphi(x)\leq v\}$ is calm at $(0,\bar{x},\bar{y})$. (5.25)

By applying the concept of stability regions known in linear programming (see, e.g., [21]), to the optimal value function φ it is possible to show, by means of Robinson's theorem [106] on the upper-Lipschitz continuity of a polyhedral set-valued mapping, that (A_1^v) is automatically satisfied if f and g are linear. The corresponding weak basic CQ (3.42) is a sufficient condition for the fulfilment of (A_1^v). Sufficient conditions for the uniform weak sharp minimum discussed in Chapter 3 also imply the satisfaction of (A_1^v) in the corresponding setting.

For the Lipschitz-like property of the solution map S in (3.1), our additional qualification condition is formulated as follows:

(A_2^v) $$\big[(\lambda,\beta)\in\Lambda_y^o(\bar{x},\bar{y},0),\ u\in\partial(-\varphi)(\bar{x})\big] \Longrightarrow \lambda u = -\lambda\nabla_x f(\bar{x},\bar{y}) - \sum_{i=1}^p \beta_i\nabla_x g_i(\bar{x},\bar{y}), \tag{5.26}$$

where $\Lambda_y^o(\bar{x},\bar{y},v)$ for $v\in\mathbb{R}^m$ denotes a particular set of multipliers that plays an important role in the rest of the section:

$$\Lambda_y^o(\bar{x},\bar{y},v) := \{(\lambda,\beta)|\quad \begin{aligned} & v+\lambda\nabla_y f(\bar{x},\bar{y})+\textstyle\sum_{i=1}^p \beta_i\nabla_y g_i(\bar{x},\bar{y})=0,\\ & \lambda\geq 0,\ \beta_i\geq 0,\ \beta_i g_i(\bar{x},\bar{y})=0,\ i=1,\dots,p\}. \end{aligned} \tag{5.27}$$

The next proposition describes a setting where assumption (A_2^v) is automatically satisfied.

Proposition 5.1.7 (validity of assumption (A_2^v))**.** *Let $f : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ and $g : \mathbb{R}^m \to \mathbb{R}^p$ be two convex and continuously differentiable functions. Consider the value function*

$$\varphi(x) := \min_y \{f(x,y)|\, g(y) \leq 0\}$$

and the corresponding solution map $S(x) = \min_y \{f(x,y)|\, g(y) \leq 0\}$. Take $(\bar{x},\bar{y}) \in \operatorname{gph} S$ with $\varphi(\bar{x}) < \infty$, and assume that there exists $\hat{y}$ with $g_i(\hat{y}) < 0$ for $i = 1,\ldots,p$. Then (A_2^v) holds at $(\bar{x},\bar{y})$.

Proof. Under the setting of this proposition, it follows from the convex case of Theorem 3.2.9 (ii) that the function $-\varphi$ is strictly differentiable at $\bar{x}$ and $\partial(-\varphi)(\bar{x}) = \{-\nabla_x f(\bar{x},\bar{y})\}$, which therefore justifies our conclusion. □

The LLVF reformulation of the solution set-valued mapping of the parametric problem corresponding to the two-level value function φ_o (5.23), useful in the next result, reads as

$$S_o(x) = \{y \in K(x)|\, f(x,y) \leq \varphi(x),\ F(x,y) \leq \varphi_o(x)\}. \tag{5.28}$$

Theorem 5.1.8 (coderivative estimate and Lipschitz-like property of lower-level solution maps)**.** *Let the solution map* (5.28) *be inner semicontinuous at $(\bar{x},\bar{y}) \in \operatorname{gph} S_o$, and let the condition (A_1^v) and the lower-level regularity* (3.65) *be satisfied at this point. Then we have for all $v \in \mathbb{R}^m$*

$$D^*S(\bar{x},\bar{y})(v) \subseteq \bigcup_{(\lambda,\beta)\in\Lambda_y^o(\bar{x},\bar{y},v)} \Big\{\lambda(\nabla_x f(\bar{x},\bar{y}) + \partial(-\varphi)(\bar{x})) + \sum_{i=1}^p \beta_i \nabla_x g_i(\bar{x},\bar{y})\Big\}. \tag{5.29}$$

If in addition (A_2^v) holds at $(\bar{x},\bar{y})$, then S is Lipschitz-like around this point.

Proof. It follows from Theorem 3.2.9 (ii) that the lower-level value function φ is Lipschitz continuous around $\bar{x}$ under the lower-level regularity and the inner semicontinuity assumptions. If we add the calmness property (A_1^v), then

$$N_{\operatorname{gph} S}(\bar{x},\bar{y}) \subseteq \bigcup_{\lambda\geq 0} \{\lambda(\nabla f(\bar{x},\bar{y}) + \partial(-\varphi)(\bar{x}) \times \{0\}) + N_\Omega(\bar{x},\bar{y})\}$$

by Theorem 2.2.7 while taking into account that the constraint $f(x,y) - \varphi(x) \leq 0$ is active at $(\bar{x},\bar{y})$. The coderivative estimate (5.29) of the theorem follows now from definition (2.24) and the well-known expression of the normal cone

$$N_\Omega(\bar{x},\bar{y}) = \Big\{\sum_{i=1}^p \beta_i \nabla g_i(\bar{x},\bar{y})|\ \beta_i \geq 0,\ \beta_i g_i(\bar{x},\bar{y}) = 0,\ i = 1,\ldots,p\Big\},$$

which holds under the validity of the lower-level regularity (3.65) at $(\bar{x},\bar{y})$. Further, by (5.29) and the coderivative criterion (2.30) for the Lipschitz-like property we get that the latter holds provided that

$$\left.\begin{array}{r} u \in \lambda(\nabla_x f(\bar{x},\bar{y}) + \partial(-\varphi)(\bar{x})) + \sum_{i=1}^p \beta_i \nabla_x g_i(\bar{x},\bar{y}) \\ (\lambda,\beta) \in \Lambda_y^o(\bar{x},\bar{y},0) \end{array}\right\} \Longrightarrow u = 0,$$

which is in fact equivalent to the assumed qualification condition (A_2^v). □

Remark 5.1.9 (sensitivity result of the solution map S under full convexity and inner semicompactness)**.** *It follows from the alternative statement in Theorem 3.2.9 (ii) that the inner semicontinuity of S can be dropped in the assumptions of Theorem 5.1.8 if the functions f and g are fully convex. As usual, the inner semicontinuity can be replaced by inner semicompactness with a larger inclusion in* (5.29).

To conduct a local sensitivity analysis of the two-level value function φ_o defined in (5.23), we associate with it the optimal solution map S_o (5.28). Having in mind the definition of the multiplier set $\Lambda^o_y(\bar{x},\bar{y},v)$ in (5.27), we put $\Lambda^o_y(\bar{x},\bar{y}) := \Lambda^o_y(\bar{x},\bar{y},\nabla_y F(\bar{x},\bar{y}))$. Then sensitivity results for φ_o are given next.

Theorem 5.1.10 (LLVF-type sensitivity analysis for the two-level value function)**.** *Consider the LLVF reformulation* (5.23) *of the two-level value function, then the following assertions are satisfied:*

(i) *Assume that S_o* (5.28) *is inner semicontinuous at $(\bar{x},\bar{y})$ and that condition (A^v_1) and the lower-level regularity* (3.65) *hold at this point. Then we have*

$$\partial\varphi_o(\bar{x}) \subseteq \bigcup_{(\lambda,\beta)\in\Lambda^o_y(\bar{x},\bar{y})} \ \bigcup_{\mu\in\Lambda(\bar{x},\bar{y})} \Big\{\nabla_x F(\bar{x},\bar{y}) + \sum_{i=1}^{p}(\beta_i - \lambda\mu_i)\nabla_x g_i(\bar{x},\bar{y})\Big\}. \tag{5.30}$$

If in addition (A^v_2) is satisfied at $(\bar{x},\bar{y})$, then φ_o is Lipschitz continuous around $\bar{x}$.

(ii) *Assume that S_o* (5.28) *is inner semicompact at $\bar{x}$, that the lower-level regularity* (3.65) *holds at $(\bar{x},y)$ for all $y\in S(\bar{x})$, while (A^v_1) holds at $(\bar{x},y)$ for all $y\in S_o(\bar{x})$. Then we have*

$$\partial\varphi_o(\bar{x}) \subseteq \bigcup_{y\in S_o(\bar{x})} \ \bigcup_{(\lambda,\beta)\in\Lambda^o_y(\bar{x},y)} \Big\{\nabla_x F(\bar{x},y) + \lambda\nabla_x f(\bar{x},y) + \lambda\partial(-\varphi)(\bar{x}) + \textstyle\sum_{i=1}^{p}\beta_i\nabla_x g_i(\bar{x},y)\Big\}, \tag{5.31}$$

where the subdifferential $\partial(-\varphi)(\bar{x})$ is estimated in Theorem 3.2.9 (i)*. If in addition (A^v_2) is satisfied at $(\bar{x},y)$ for all $y\in S_o(\bar{x})$, then φ_o is Lipschitz continuous around $\bar{x}$.*

Proof. To justify (i), observe by Theorem 4.0.1(i) that

$$\partial\varphi_o(\bar{x}) \subseteq \nabla_x F(\bar{x},\bar{y}) + D^*S(\bar{x},\bar{y})(\nabla_y F(\bar{x},\bar{y})).$$

under the inner semicontinuity assumption on S_o. Since we obviously have $S_o(x)\subseteq S(x)$ for all $x\in X$, the lower-level optimal solution map S in (1.2) is also inner semicontinuous at $(\bar{x},\bar{y})\in \operatorname{gph} S_o$. Thus the upper estimate of $\partial\varphi_o(\bar{x})$ in this theorem follows from those for the coderivative of S in Theorem 5.1.8 and for the subdifferential of the lower-level value function φ in Theorem 3.2.9 (ii). To justify the local Lipschitz continuity of φ_o in (i) under (A^v_2), recall that the latter condition implies the Lipschitz-like property of S around $(\bar{x},\bar{y})$ by Theorem 5.1.8. Thus we have the claimed result from Theorem 4.0.1 (i).

Assertion (ii) is proved similarly following the discussion in Remark 5.1.9. □

Observe that for the subdifferential estimate of φ_o in Theorem 5.1.10 (i), the upper bound of the basic subdifferential does not contain the partial derivative of the lower-level cost function f with respect to the upper-level variable x. This will induce in the context of necessary optimality conditions for the original optimistic formulation (P_o) in the next section a remarkable phenomenon first discovered in [26] in the framework concerning the auxiliary problem (P). Note that such a phenomenon is no longer true if the inner semicontinuity assumption on S_o is replaced by the inner semicompactness one in assertion (ii) of Theorem 5.1.10. Finally, we mention that the inner semicompactness of S_o in Theorem 5.1.10 (ii) can be replaced by the easier while more restrictive uniform boundedness assumption imposed on S_o or even on the lower-level solution map S.

5.2 Necessary optimality conditions

This section is devoted to applications of the above sensitivity results to deriving necessary optimality conditions for the original optimistic formulation (P_o) in bilevel programming. In fact we establish certain *stationarity conditions* of various types among which are of those types known for more conventional auxiliary optimistic formulation (P) together with stationarity conditions of the novel types for (P_o). Next we define the stationarity concepts tailored to (P_o).

Definition 5.2.1 (KM, KN and M-stationarity concepts for the original optimistic bilevel program (P_o)). *Consider a feasible point $\bar{x}$ of the original optimistic bilevel program (P_o), then it is said to be:*

(i) P_o-KM-STATIONARY *if there exist $\bar{y} \in S_o(\bar{x})$, $(\alpha, \beta) \in \mathbb{R}^{k+p}$, $\lambda \in \mathbb{R}_+$, $(\mu_s, \upsilon_s) \in \mathbb{R}^{k+1}$ and $y_s \in S(\bar{x})$ with $s = 1, \ldots, n+1$ such that the KM-stationarity conditions* (3.12)–(3.18) *are satisfied.*

(ii) SP_o-KN-STATIONARY (*resp.* P_o-KN-STATIONARY) *if for every $\bar{y} \in S_o(\bar{x})$ (resp. for some $\bar{y} \in S_o(\bar{x})$) we can find a triple $(\alpha, \beta, \mu) \in \mathbb{R}^{k+2p}$ and a number $\lambda \in \mathbb{R}_+$ such that the KN-stationarity conditions* (3.13), (3.16)–(3.17), *and* (3.19)–(3.21) *are satisfied.*

(iii) SP_o-M-STATIONARY (*resp.* P_o-M-STATIONARY) *if for every $(\bar{y}, \bar{u}) \in S_o^h(\bar{x})$ (resp. for some $(\bar{y}, \bar{u}) \in S_o^h(\bar{x})$) we can find a triple $(\alpha, \beta, \gamma) \in \mathbb{R}^{k+p+m}$ such that the M-stationarity conditions* (3.16) *and* (3.23)–(3.26) *are satisfied.*

Similarly to the M-stationarity, we define the C-stationarity (resp. S-stationarity) by replacing condition (3.26) with (3.29) and (3.30), respectively. Observe that a diagram similar to the one in (3.31) can be constructed for the above stationarity concepts of the original optimistic bilevel program (P_o). In the following results, we provide frameworks to obtain these conditions.

Theorem 5.2.2 (derivation of the KM-stationarity conditions for (P_o)). *Let $\bar{x}$ be an upper-level regular* (3.72) *local optimal solution of (P_o), and let S_o* (5.28) *be inner semicompact at $\bar{x}$ while the lower-level regularity* (3.65) *is satisfied at all $(\bar{x}, \bar{y})$ with $\bar{y} \in S(\bar{x})$. Suppose furthermore that (A_1^v) and (A_2^v) are satisfied at all $(\bar{x}, \bar{y})$ with $\bar{y} \in S_o(\bar{x})$. Then the point $\bar{x}$ is P_o-KM-stationary.*

Proof. Under the assumptions made, it follows from Theorem 5.1.10 (ii) that the two-level value function φ_o (5.23) is Lipschitz continuous around $\bar{x}$. Thus $\partial \varphi_o(\bar{x}) \neq \emptyset$ while $\partial^\infty \varphi_o(\bar{x}) = \{0\}$, and the qualification condition (2.38) in Theorem 2.2.1 holds at $\bar{x}$. Employing now the optimality condition (2.39) of the latter theorem with the well-known formula

$$N_X(\bar{x}) = \Big\{ \sum_{j=1}^{k} \alpha_j \nabla G_j(\bar{x}) |\ \alpha_j \geq 0,\ \alpha_j G_j(\bar{x}) = 0,\ j = 1, \ldots k \Big\} \tag{5.32}$$

valid under the assumed upper-level regularity of $\bar{x}$ and then taking into account that the set on the right-hand-side of inclusion (5.31) is nonempty, we arrive at the P_o-KM-stationarity. □

Theorem 5.2.3 (derivation of the KN-stationarity conditions for (P_o)). *Let $\bar{x}$ be an upper-level regular* (3.72) *local optimal solution to the bilevel program (P_o), then the following assertions are satisfied:*

(i) *Assume that for all $\bar{y} \in S_o(\bar{x})$, the inner semicontinuity of S_o* (5.28)*, the lower-level regularity* (3.65)*, and conditions (A_1^v) and (A_2^v) are all satisfied at $(\bar{x}, \bar{y})$. Then $\bar{x}$ is SP_o-KN-stationary.*

(ii) *If all the assumptions of (i) are satisfied at a certain point $(\bar{x}, \bar{y})$ with $\bar{y} \in S_o(\bar{x})$, then $\bar{x}$ is P_o-KN-stationary.*

Proof. For (i), note that inclusion (5.30) depends on the couple $(\bar{x}, \bar{y}) \in \operatorname{gph} S_o$, where S_o is inner semicontinuous. Thus, it is clear that if the latter is satisfied at every $(\bar{x}, \bar{y}) \in \operatorname{gph} S_o$ and all the other qualification conditions of Theorem 5.1.10 (i) hold at these points, then the result follows in the the lines of that of Theorem 5.2.2. On the other hand, (ii) is obtained from the combination of Theorem 2.2.1 (ii) and Theorem 5.1.10 (ii). □

Theorem 5.2.4 (derivation of the M-stationarity conditions for (P_o)). *Let $\bar{x}$ be an upper-level regular* (3.72) *local optimal solution to (P_o), where the lower-level problem* (1.3) *is convex. Assume that the Slater CQ holds at $\bar{x}$ while relationships (A_1^m) and (A_2^m) are satisfied at $(\bar{x}, \bar{y}, \bar{u})$ for all $(\bar{y}, \bar{u}) \in S_o^h(\bar{x})$. Then the following assertions are satisfied:*

(i) *If the solution set-valued mapping S_o^h* (5.6) *is inner semicontinuous at $(\bar{x}, \bar{y}, \bar{u})$ for all $(\bar{y}, \bar{u}) \in S_o^h(\bar{x})$, then $\bar{x}$ is SP_o-M-stationary.*

(ii) *If S_o^h* (5.6) *is inner semicompact at $\bar{x}$, then $\bar{x}$ is P_o-M-stationary.*

Proof. It follows similarly to that of the previous theorem while using Theorem 2.2.1 (ii) and Theorem 5.1.2 (i) and (ii), respectively. □

Another possibility to derive the P$_o$-M-stationarity is by using the upper estimate of $\partial\varphi_o(\bar{x})$ obtained via the OPEC reformulation of the two-level value function (5.21), cf. Corollary 5.1.5 (ii). Note also that if the inner semicontinuity and qualification conditions (A_1^m) and (A_2^m) are satisfied only at one point $(\bar{x},\bar{y},\bar{u})$ in Theorem 5.2.4, we still can derive the P$_o$-M-stationarity at the difference that the reference couple $(\bar{y},\bar{u}) \in S_o^h(\bar{x})$ is known a priori. A comment similar to the latter one can be made for the P$_o$-KN-stationarity.

Proceeding similarly to the proof of Theorem 5.2.4, the C-stationarity (resp. S-stationarity) for a local optimal solution to problem (P$_o$) can be derived by a combination of Theorem 2.2.1 (ii) and Theorem 5.1.3 (resp. Theorem 5.1.4). Next we provide an illustrative example for the assumptions made in Theorem 5.2.4.

Example 5 (validity of the qualification conditions to derive stationarity conditions for (P$_o$))**.** *Consider the original optimistic version of the bilevel program in Example 3:*

$$\min_{x\in\mathbb{R}_+}\{\min_{y\in\mathbb{R}}\{x^2+y^2|\, y\in S(x):=\arg\min\{xy+y|\, y\geq 0\}\}\}.$$

The KKT reformulation of the corresponding two-level value function is

$$\varphi_o(x):=\min_{y,u}\{x^2+y^2|\, x-u+1=0,\ u\geq 0, y\geq 0,\ uy=0\}=\begin{cases} x^2 & \text{if } x\geq -1,\\ \infty & \text{otherwise.}\end{cases}$$

It is obvious that $\bar{x}=0$ is the (unique) optimistic optimal solution of this program and that φ_o is continuously differentiable near $\bar{x}$. On the other hand, we have $S_o^h(x)=\{(0,x+1)\}$ if $x\geq -1$ and $S_o^h(x)=\emptyset$ otherwise, and hence S_o^h reduces to a single-valued and continuous mapping on its graph. Furthermore, $\Lambda_y^{em}(\bar{x},\bar{y},\bar{u},0)=\{0\}\times\mathbb{R}$ if $(\bar{x},\bar{y},\bar{u})=(-1,0,0)$ and $\Lambda_y^{em}(\bar{x},\bar{y},\bar{u},0)=\{(0,0)\}$ for all the other points of $\mathrm{gph}\, S_o^h$. *From the observations made in Sections 4.1-6.1, this implies that the corresponding CQs (A_1^m) and (A_2^m) are satisfied at all points of the graph of S_o^h except $(-1,0,0)$, which is not optimal.*

It is worth mentioning that the upper-level regularity (3.72) imposed in the results above can be replaced by the weaker *calmness* property of the mapping $v\rightrightarrows\{x\in\mathbb{R}^n|\, G(x)+v\leq 0\}$, which is automatically satisfied in if G is a linear function. Furthermore, as mentioned previously in Subsection 5.1.1, the qualification condition (A_1^m) can also be replaced by the weaker calmness property of the mapping Ψ in (5.22), which holds if both functions g and $(x,y)\mapsto\nabla_y f(x,y)$ are linear.

Remark 5.2.5 (on the need of the Lipschitz continuity of the two-level value function φ_o)**.** *We can see from the proof of Theorem 5.2.4 that the local Lipschitz continuity of φ_o was used twice: to ensure the nonemptiness of $\partial\varphi_o(\bar{x})$ and the application of the optimality condition* (2.39) *of Theorem 2.2.1. Observe to this end that the Lipschitz property of φ_o is not needed for bilevel programs without upper-level constraints (i.e., if $X:=\mathbb{R}^n$); in this case the qualification condition* (2.38) *holds automatically. The latter also allows us to drop assumption (A_2^m) in Theorem 5.2.4. However, we still have to make sure that $\partial\varphi_o(\bar{x})\neq\emptyset$, which happens in many non-Lipschitzian situations; see e.g., [89, 92, 109].*

5.3 Original versus classical optimistic bilevel programming

To begin this section, we first recall that the classical optimistic bilevel optimization problem (P) and its original optimistic formulation (P$_o$) are equivalent for global solutions. However, for local solutions, the relationship is clarified in the next proposition.

Proposition 5.3.1 (relationship between the local optimal solutions of the classical and original optimistic programs)**.** *Consider problems (P) and (P_o), then the following assertions are satisfied:*

(i) *Let $\bar{x}$ be a local optimal solution of problem (P_o), then $(\bar{x},\bar{y})$ with $\bar{y} \in S_o(\bar{x})$ is a local optimal solution of problem (P).*

(ii) *Let $(\bar{x},\bar{y})$ be a local optimal solution of (P), with $\bar{y} \in S_o(\bar{x})$. Assume that S_o* (5.2) *is inner semicontinuous at $(\bar{x},\bar{y})$, then $\bar{x}$ is a local optimal solution of problem (P_o).*

Proof. (i) By contradiction, assume that $(\bar{x},\bar{y})$ with $\bar{y} \in S_o(\bar{x})$ is not a local optimal solution of problem (P). Then, we can find a sequence (x^k,y^k) with $x^k \to \bar{x}$, $y^k \to \bar{y}$ and $x^k \in X$, $y^k \in S(x^k)$ such that we have $F(x^k,y^k) < F(\bar{x},\bar{y}) = \varphi_o(\bar{x})$ for all k. By the definition of φ_o, it follows that $\varphi_o(x^k) \leq F(x^k,y^k)$ for all k. Thus we have $\varphi_o(x^k) < \varphi_o(\bar{x})$, $x^k \in X$ for all k. This completes the proof of (i) given that $\bar{x} \in X$.

(ii) Assume that $\bar{x}$ is not a local optimal solution of (P_o) while all the other hypotheses of (ii) in the proposition are satisfied. We can find a sequence $x^k \to \bar{x}$ ($x^k \in X$) such that $\varphi_o(x^k) < \varphi_o(\bar{x})$ for all k. S_o being inner semicontinuous at $(\bar{x},\bar{y})$, there is a sequence $y^k \in S_o(x^k)$ such that $y^k \to \bar{y}$. Considering the definition of S_o (5.2), we have $y^k \in S(x^k)$.

On the other hand, with $y^k \in S_o(x^k)$, we have $F(x^k,y^k) = \varphi_o(x^k)$ for all k. Hence, it follows that $F(x^k,y^k) < F(\bar{x},\bar{y})$, $x^k \in X$, $y^k \in S(x^k)$ for all k, given that $\varphi_o(x^k) < \varphi_o(\bar{x})$ for all k. This contradicts the fact that $(\bar{x},\bar{y})$ is a local optimal solution of problem (P), given that $x^k \to \bar{x}$, $y^k \to \bar{y}$. □

Let us now remind that it was initially shown in [44] that (i) holds under the upper semicontinuity of the lower-level solution set-valued mapping S (1.2). This assumption was then weakened to the uniform boundedness in [26]. It appears from the above that no assumption is in fact needed. However, for the converse implication, the inner semicontinuity of the mapping S_o (5.2) seems to be indispensable as shown in the next example.

Example 6 (importance of the inner semicontinuity of S_o in Proposition 5.3.1 (ii))**.** *Consider the following bilevel program taken from* [21]*:*

$$\text{“}\min_x\text{”} \left\{x|\; x \in [-1,1],\; y \in S(x) := \arg\min_y \{xy|\; y \in [0,1]\}\right\}.$$

The lower-level solution set-valued mapping can be described as:

$$S(x) = \begin{cases} [0,1] & \text{if } \; x = 0, \\ \{0\} & \text{if } \; x > 0, \\ \{1\} & \text{if } \; x < 0. \end{cases}$$

$(0,0)$ *is a local optimal solution for the classical optimistic model (P), while* 0 *is not a local optimal solution for the corresponding original optimistic formulation (P_o). Note that in this case, we have $S_o(x) = S(x)$ for all $x \in X := [-1,1]$. Moreover, $-1/2^k \to 0$, but $S_o(-1/2^k) = S(-1/2^k) = \{1\}$ for all k, while $1 \neq 0$. Thus S_o is not inner semicontinuous at* $(0,0)$.

In the next result, we establish the link between the stationarity conditions for the original optimistic formulation (P_o) in the previous section and those of the *conventional/auxiliary optimistic* problem (P) obtained in Chapter 3.

Theorem 5.3.2 (relationship between the stationarity conditions for the classical and original optimistic bilevel programs)**.** *The following assertions hold:*

(i) *A point $\bar{x}$ is P_o-KM-stationary (resp. P_o-KN-stationary) if and only if there exists $\bar{y} \in S_o(\bar{x})$ such that $(\bar{x},\bar{y})$ is P-KM-stationary (resp. P-KN-stationary).*

(ii) *A point $\bar{x}$ is P_o-M-stationary (resp. P_o-C, P_o-S-stationary) if and only if there exists $(\bar{y},\bar{u}) \in S^h_o(\bar{x})$ such that $(\bar{x},\bar{y})$ is P-M-stationary (resp. P-C, P-S-stationary).*

Proof. It follows from the direct comparison of the necessary optimality/stationarity conditions obtained in Section 5.2 for the original bilevel formulation (P_o) and the ones of (P) in classical optimistic bilevel programming, cf. Chapter 3. □

A natural question to ask now is how are the CQs to derive the stationarity conditions of (P) and (P_o) related. We address this solely for the P-M and P_o-M-stationarity conditions. Similar comments can be made for the other ones. In fact, note that CQ (3.97) and our qualification condition (A_1^m) in (5.7) are very similar. However, the upper-level constraint function G (1.4) does not appear in the latter. Thus, let us consider a bilevel program with no upper-level constraints. Then (A_1^m) and CQ (3.97) coincide. But for the original optimistic problem (P_o), we additionally need the qualification condition (A_2^m) to ensure the Lipschitz continuity of φ_o (5.1). Provided the latter is not needed (cf. discussion in Remark 5.2.5), the extra assumption which really makes the difference in the aforementioned framework while deriving the P-M-stationarity and the P_o-M-stationarity is that the inner semicompactness or semicontinuity of the mapping S_o^h (5.6) is required in the latter case.

To conclude this section, we illustrate how the local sensitivity analysis of two-level value functions obtained in this chapter can readily be applied for the sensitivity analysis of the auxiliary problem (P). If we consider a perturbation of the *argminimum/solution* set-valued mapping S (1.2) as:

$$S(x,z) := \arg\min_y \{f(x,y,z) |\ g(x,y,z) \leq 0\},$$

then, we can define the following perturbation for the classical model of the bilevel program (P):

$$\min_{x,y} \{F(x,y,z) |\ G(x,z) \leq 0,\ y \in S(x,z)\}. \tag{5.33}$$

One way to derive the sensitivity analysis of the optimal value function associated to the above family of problems can follow from the results in Subsection 5.1.1, by means of the KKT reformulation (see (3.8) for details):

$$\begin{aligned}\min_{x,y,u} \{F(x,y,z) |\quad & u \geq 0,\ g(x,y,z) \leq 0,\ u^\top g(x,y,z) = 0,\\ & G(x,z) \leq 0,\ \nabla_y f(x,y,z) + \nabla_y g(x,y,z)^\top u = 0\}.\end{aligned}$$

This has already been noted in [76]. However, another possibility to do the aforementioned sensitivity analysis by considering the generalized equation reformulation of problem (5.33)

$$\min_{x,y} \{F(x,y,z) | G(x,z) \leq 0,\ \nabla_y f(x,y,z) + N_{K(x,z)}(y) \ni 0\}$$

is given in Section 4.2 (also see Subsection 5.1.1), while for the lower-level value function reformulation

$$\min_{x,y} \{F(x,y,z) |\ f(x,y,z) - \varphi(x,z) \leq 0,\ G(x,z) \leq 0,\ g(x,y,z) \leq 0\},$$

the interested reader in referred to Subsection 5.1.2. Note that the upper-level constraint $G(x,z) \leq 0$ can easily be accommodated in the corresponding settings.

Finally, it is worth mentioning that the results on the sensitivity analysis of two-level value functions obtained in this chapter can readily be applied to derive lower subdifferential necessary optimality conditions for the pessimistic bilevel program (P_p). This is part of what we do in the next chapter.

6 Pessimistic bilevel programming problem

The pessimistic bilevel program (P_p) is a special class of parametric/nonstatic minmax programming problems. Many publications (see e.g. [39, 112, 113]) have been devoted to the derivation of necessary optimality conditions for static minmax problems. Ishizuka [68] was the first to investigate necessary optimality conditions for parametric minmax programs, however while using a generalization of Farkas' alternative theorem. In the paper [131], lower and upper subdifferential necessary optimality conditions for the latter problem are derived using the generalized differentiation calculus from variational analysis. It however appears that the results in this paper cannot be applied to (P_p) because the constraint qualifications needed there do not hold. Secondly, the graph of the lower-level solution map S being nonconvex, part of the results in the aforementioned paper are also not applicable to the pessimistic bilevel programming problem.

In this chapter, we provide a study on stationarity conditions tailored to (P_p) via generalized differentiation calculus. Relationships between the lower subdifferential necessary optimality conditions of the optimistic and pessimistic problems are provided in Subsection 6.2.4. Chapter 5 is a prerequisite to Sections 6.1–6.2. Section 6.3 provides upper subdifferential optimality conditions for (P_p), which appear to have a structure similar to the necessary optimality conditions develop by Ishizuka [68] in the context of the parametric minmax programming problem.

6.1 LLVF reformulation of the pessimistic bilevel program

We consider the following LLVF reformulation for the maximization two-level value function φ_p (1.8):

$$\varphi_p(x) = \max_y \{F(x,y) | \, g(x,y) \le 0, \, f(x,y) - \varphi(x) \le 0\}. \tag{6.1}$$

In order to apply the results obtained in the previous chapter for the optimistic problem (P_o), we introduce the minimization two-level value function

$$\varphi_{op}(x) = \min_y \{-F(x,y) | \, g(x,y) \le 0, \, f(x,y) - \varphi(x) \le 0\}. \tag{6.2}$$

Obviously, the classical interplay between the minimization and maximization of a given objective function induces the equality $\varphi_p(x) = -\varphi_{op}(x)$. The solution set-valued mapping associated to φ_{op} (6.2) follows as

$$S_{op}(x) := \{x \in S(x) | F(x,y) + \varphi_{op}(x) \ge 0\}. \tag{6.3}$$

As usual, we suppose that $S_{op}(x) \neq \emptyset$ for all $x \in X$, in such a way that the two-level value functions in (6.1) and (6.2) will be well-defined.

In the spirit of the previous chapter (also see Chapter 3), we use the prefixes "KM" and "KN" in this section to designate the optimality conditions of the pessimistic bilevel programming problem (P_p) obtained via the LLVF reformulation (6.1) while assuming the inner semico**N**tinuity and inner semico**M**pactness of S_{op} (6.3), respectively. KM and/or KN will be preceded by "KKT-type" if no number is attached to the gradients of the upper-level objective function F. A similar terminology (i.e. analogously to Chapter 5) will be used in the next section in order to facilitate comparisons between the optimality conditions of (P_o) and (P_p). It is worth to remind at this point that the qualification conditions (A_1^v) and (A_2^v) used in this section are defined in (5.25) and (5.26), respectively.

Theorem 6.1.1 (KM-stationarity conditions for the pessimistic bilevel program)**.** *Let $\bar{x}$ be an upper-level regular* (3.72) *local optimal solution for problem (P_p), where the lower-level regularity* (3.65) *is satisfied at $(\bar{x},y)$, $y \in S(\bar{x})$. Assume that S_{op}* (6.3) *is inner semicompact at $\bar{x}$ while conditions (A_1^v) and (A_2^v) hold at $(\bar{x},y)$, for all $y \in S_{op}(\bar{x})$. Then, there exist $\alpha \in \mathbb{R}^k$, $(\beta_t, \mu_s, \lambda_t, v_s, w_t) \in \mathbb{R}^{2p+3}$, $y_t \in S_{op}(\bar{x})$ and $y_s \in S(\bar{x})$ with $s = 1,\dots,n+1$ and $t = 1,\dots,n+1$ such that relationships* (3.14)–(3.16) *and* (3.18)*, together with the following conditions are satisfied:*

$$\sum_{t=1}^{n+1} w_t \big(\nabla_x F(\bar{x},y_t) + \lambda_t \nabla_x f(\bar{x},y_t) + \sum_{i=1}^{p} \beta_{it} \nabla_x g_i(\bar{x},y_t)\big)$$
$$-\Big(\sum_{t=1}^{n+1} w_t \lambda_t\Big) \sum_{s=1}^{n+1} v_s \big(\nabla_x f(\bar{x},y_s) + \sum_{i=1}^{p} \mu_{is} \nabla_x g_i(\bar{x},y_s)\big) + \sum_{j=1}^{k} \alpha_j \nabla G_j(\bar{x}) = 0, \tag{6.4}$$

$$\forall t = 1,\dots,n+1,\ \nabla_y F(\bar{x},y_t) + \lambda_t \nabla_y f(\bar{x},y_t) + \sum_{i=1}^{p} \beta_{it} \nabla_y g_i(\bar{x},y_t) = 0, \tag{6.5}$$

$$\forall t = 1,\dots,n+1,\ i = 1,\dots,p,\ \lambda_t \le 0,\ \beta_{it} \le 0,\ \beta_{it} g_i(\bar{x},y_t) = 0, \tag{6.6}$$

$$\sum_{t=1}^{n+1} w_t = 1,\ \forall t = 1,\dots,n+1,\ w_t \ge 0. \tag{6.7}$$

Proof. Apart from the upper-level regularity of $\bar{x}$, all the other assumptions of the theorem imply that the value function φ_{op} (6.2) is Lipschitz continuous around $\bar{x}$, cf. Theorem 5.1.10 (ii). Hence, φ_p is also Lipschitz continuous around $\bar{x}$. It then follows from Theorem 2.2.1 (ii) that

$$0 \in \partial \varphi_p(\bar{x}) + N_X(\bar{x}). \tag{6.8}$$

Considering the upper-level regularity of $\bar{x}$, the expression of the normal cone to X is given in (5.32). Thus one can find $\alpha \in \mathbb{R}^k$ satisfying the complementarity conditions (3.16) such that

$$\sum_{j=1}^{k} \alpha_j \nabla G_j(\bar{x}) \in \operatorname{co} \partial \varphi_{op}(\bar{x}), \tag{6.9}$$

while taking into account the convex hull property (2.6). Applying the well-known Carathéodory Theorem to (6.9), there exist $w_t, t = 1,\dots,n+1$ satisfying (6.7) and $x_t^* \in \partial \varphi_{op}(\bar{x}), t = 1,\dots,n+1$ such that we have

$$\sum_{j=1}^{k} \alpha_j \nabla G_j(\bar{x}) = \sum_{t=1}^{n+1} w_t x_t^*. \tag{6.10}$$

On the other hand, we also have from Theorem 5.1.10 (ii) that

$$\partial \varphi_{op}(\bar{x}) \subseteq \bigcup_{y \in S_{op}(\bar{x})} \bigcup_{(\lambda,\beta) \in \Lambda_y^o(\bar{x},y)} \big\{ -\nabla_x F(\bar{x},y) + \lambda(\nabla_x f(\bar{x},y) + \partial(-\varphi)(\bar{x})) + \sum_{i=1}^{p} \beta_i \nabla_x g_i(\bar{x},y) \big\},$$

where $\Lambda_y^o(\bar{x},y) := \Lambda_y^o(\bar{x},y,-\nabla_y F(\bar{x},y))$, with $\Lambda_y^o(\bar{x},y,-\nabla_y F(\bar{x},y))$ given by (5.27). Clearly, $x_t^* \in \partial \varphi_{op}(\bar{x})$ implies that there exist $y_t \in S_{op}(\bar{x})$ and $(\lambda_t, \beta_t) \in \mathbb{R}^{1+p}$, with $t = 1,\dots,n+1$ such that the following relationships hold true:

$$\forall t = 1,\dots,n+1,\ -\nabla_y F(\bar{x},y_t) + \lambda_t \nabla_y f(\bar{x},y_t) + \sum_{i=1}^{p} \beta_{it} \nabla_y g_i(\bar{x},y_t) = 0, \tag{6.11}$$

$$\forall t = 1,\dots,n+1,\ i = 1,\dots,p,\ \lambda_t \ge 0,\ \beta_{it} \ge 0,\ \beta_{it} g_i(\bar{x},y_t) = 0, \tag{6.12}$$

$$\forall t = 1,\dots,n+1,\ x_t^* + \nabla_x F(\bar{x},y_t) - \lambda_t \nabla_x f(\bar{x},y_t) - \sum_{i=1}^{p} \beta_{it} \nabla_x g_i(\bar{x},y_t) \in \lambda_t \partial(-\varphi)(\bar{x}). \tag{6.13}$$

At this point, it should be reminded that the inner semicompactness of S (3.1) (ensured by that of S_{op} (6.3)) and the fulfilment of the lower-level regularity at $(\bar{x},y)$, $y\in S(\bar{x})$ imply the local Lipschitz continuity of $-\varphi$, cf. Theorem 3.2.9 (i). Furthermore, we have the upper estimate of $\partial(-\varphi)(\bar{x})$ in Theorem 3.2.9 (i) ensuring that inclusion (6.13) implies the existence of $y_s\in S(\bar{x})$, $\mu_s\in\mathbb{R}^p$ and $v_s\in\mathbb{R}$, with $s=1,\dots,n+1$ such that relationships (3.14)–(3.15) and (3.18) together with the following hold:

$$x_t^*=-\nabla_xF(\bar{x},y_t)-\lambda_t\nabla_xf(\bar{x},y_t)-\sum_{i=1}^{p}\beta_{it}\nabla_xg_i(\bar{x},y_t)+\lambda_t\sum_{s=1}^{n+1}v_s(\nabla_xf(\bar{x},y_s)+\sum_{i=1}^{p}\mu_{is}\nabla_xg_i(\bar{x},y_s)).$$

Substituting the values of x_t^* in (6.10) while multiplying the resulting equation together with (6.11) and (6.12) by -1, we get (6.4), (6.5) and (6.6), respectively, which concludes the proof. □

If in the above theorem we include the full convexity of the functions involved in the lower-level problem (1.3) and we get the following result.

Corollary 6.1.2 (KM-stationarity conditions for the pessimistic bilevel program in presence of full convexity in the lower-level problem)**.** *Let $\bar{x}$ be an upper-level regular* (3.72) *local optimal solution for (P_p), where the functions f and g_i, $i=1,\dots,p$ are all convex in (x,y). Furthermore, let all the assumptions of Theorem 6.1.1 be satisfied. Then, there exist $\alpha\in\mathbb{R}^k$, $(\beta_t,\mu_t,\lambda_t,w_t)\in\mathbb{R}^{2p+2}$ and $y_t\in S_{op}(\bar{x})$ with $t=1,\dots,n+1$ such that relationships* (3.16) *and* (6.5)–(6.7) *together with the following conditions hold:*

$$\sum_{t=1}^{n+1}w_t\Big(\nabla_xF(\bar{x},y_t)+\sum_{i=1}^{p}(\beta_{it}-\lambda_t\mu_{it})\nabla_xg_i(\bar{x},y_s)\Big)+\sum_{j=1}^{k}\alpha_j\nabla G_j(\bar{x})=0, \tag{6.14}$$

$$\forall t=1,\dots,n+1,\ \nabla_yf(\bar{x},y_t)+\sum_{i=1}^{p}\mu_{it}\nabla_yg_i(\bar{x},y_t)=0, \tag{6.15}$$

$$\forall t=1,\dots,n+1,\ i=1,\dots,p,\ \mu_{it}\geq 0,\ \mu_{it}g_i(\bar{x},y_t)=0. \tag{6.16}$$

Proof. It follows as in that of Theorem 6.1.1. However, note that for all $y\in S(\bar{x})$ we have the upper estimate of the subdifferential of $-\varphi$ in (3.67), while taking into account the lower-level regularity (3.65) and the full convexity of f and g_i, $i=1,\dots,p$. Hence, for all $t=1,\dots n+1$, we choose the value $y:=y_t\in S_{op}(\bar{x})\subseteq S(\bar{x})$. □

It is important, in order to eliminate some difficulties in numerical algorithms to solve the problem, that the gradient of the upper-level objective function F not be involved in the convex combinations summations in (6.4) and (6.14). To get results of the latter type, we replace the inner semicompactness of the mapping S_{op} in the above Theorem 6.1.1 by the inner semicontinuity, which leads us to the following result.

Theorem 6.1.3 (KKT-type KN-stationarity conditions for the pessimistic bilevel program)**.** *Let $\bar{x}$ be an upper-level regular* (3.72) *local optimal solution for (P_p), where the lower-level regularity* (3.65) *is satisfied at $(\bar{x},\bar{y})$. Assume that the solution map S_{op}* (6.3) *is inner semicontinuous at $(\bar{x},\bar{y})$ while the qualification conditions (A_1^v) and (A_2^v) hold at $(\bar{x},\bar{y})$. Furthermore we assume $co\,N_{gph\,S}(\bar{x},\bar{y})$ to be a closed set. Then, there exist $\alpha\in\mathbb{R}^k$ and $(\beta_t,\mu_t,\lambda_t,w_t)\in\mathbb{R}^{2p+2}$, $t=1,\dots,n+m+1$ such that relationship* (3.16) *together with the following conditions are satisfied:*

$$\nabla_x F(\bar{x},\bar{y}) + \sum_{j=1}^{k} \alpha_j \nabla G_j(\bar{x}) + \sum_{s=1}^{n+m+1} w_t \sum_{i=1}^{p} (\beta_{it} - \lambda_t \mu_{it}) \nabla_x g_i(\bar{x},\bar{y}) = 0, \tag{6.17}$$

$$\nabla_y F(\bar{x},\bar{y}) + \sum_{t=1}^{n+m+1} w_t \big(\lambda_t \nabla_y f(\bar{x},\bar{y}) + \sum_{i=1}^{p} \beta_{it} \nabla_y g_i(\bar{x},\bar{y})\big) = 0, \tag{6.18}$$

$$\forall t = 1,\ldots,n+m+1,\ \nabla_y f(\bar{x},\bar{y}) + \sum_{i=1}^{p} \mu_{it} \nabla_y g_i(\bar{x},\bar{y}) = 0, \tag{6.19}$$

$$\forall s = 1,\ldots,n+m+1,\ i = 1,\ldots,p,\ \lambda_t \leq 0,\ \beta_{it} \leq 0,\ \beta_{it} g_i(\bar{x},\bar{y}) = 0, \tag{6.20}$$

$$\forall s = 1,\ldots,n+m+1,\ i = 1,\ldots,p,\ \mu_{it} \geq 0,\ \mu_{it} g_i(\bar{x},\bar{y}) = 0, \tag{6.21}$$

$$\sum_{t=1}^{n+m+1} w_t = 1,\ \forall t = 1,\ldots,n+m+1,\ w_t \geq 0. \tag{6.22}$$

Proof. Analogously to the previous theorem, we have from Theorem 5.1.10 (i) that φ_{op} is Lipschitz continuous around $\bar{x}$. Also, the same interplay between φ_{op} and φ_p implies that there exists $\alpha \in \mathbb{R}^k$ such that (3.13) and the following inclusion holds with the convexified subdifferential of φ_{op}:

$$\sum_{j=1}^{k} \alpha_j \nabla G_j(\bar{x}) \in \operatorname{co} \partial \varphi_{op}(\bar{x}) = \bar{\partial} \varphi_{op}(\bar{x}). \tag{6.23}$$

Since the upper-level objective function F is continuously differentiable and the solution set-valued mapping S_{op} is inner semicontinuous at $(\bar{x},\bar{y})$, one has from [91, Theorem 5.1] that

$$\bar{\partial} \varphi_{op}(\bar{x}) \subseteq -\nabla_x F(\bar{x},\bar{y}) + \overline{D}^* S(\bar{x},\bar{y})(-\nabla_y F(\bar{x},\bar{y})). \tag{6.24}$$

Combining (6.23)-(6.24) and the definition of the coderivative, one has

$$\begin{bmatrix} \sum_{j=1}^{k} \alpha_j \nabla G_j(\bar{x}) + \nabla_x F(\bar{x},\bar{y}) \\ \nabla_y F(\bar{x},\bar{y}) \end{bmatrix} \in \overline{N}_{\mathrm{gph}S}(\bar{x},\bar{y}) = \operatorname{co} N_{\mathrm{gph}S}(\bar{x},\bar{y}), \tag{6.25}$$

where the last equality results from the closedness of the set $\operatorname{co} N_{\mathrm{gph}}(\bar{x},\bar{y})$.

On the other hand, one has from Theorem 5.1.8 that under assumption (A_1^y), the lower-level regularity and the inner semicontinuity of S_{op}, the normal cone to the graph of S can be estimated as

$$N_{\mathrm{gph}S}(\bar{x},\bar{y}) \subseteq \left\{ \begin{bmatrix} \nabla_x g(\bar{x},\bar{y})^\top (\beta - \lambda\mu) \\ \lambda \nabla_y f(\bar{x},\bar{y}) + \nabla_y g(\bar{x},\bar{y})^\top \beta \end{bmatrix} \middle| \begin{array}{r} \nabla_y f(\bar{x},\bar{y}) + \nabla_y g(\bar{x},\bar{y})^\top \mu = 0 \\ \lambda \geq 0,\ \beta \geq 0,\ \beta^\top g(\bar{x},\bar{y}) = 0 \\ \mu \geq 0,\ \mu^\top g(\bar{x},\bar{y}) = 0 \end{array} \right\}.$$

By substituting this inclusion in (6.25) and applying the Carathéodory Theorem, we have the result. □

The above KKT-type KN-stationarity conditions coincide with the optimality conditions in Theorem 6.1.1, provided $S_{op}(\bar{x}) = \{\bar{y}\} = S(\bar{x})$ and $\Lambda(\bar{x},\bar{y}) = \{\mu\}$ in the latter result, while $\Lambda_y^o(\bar{x},\bar{y}) = \{(\lambda,\beta)\}$ in both theorems.

6.2 KKT reformulation of the pessimistic bilevel program

Our basic aim in this section is to derive necessary optimality conditions for the pessimistic bilevel optimization (P_p) by using either the primal or the classical KKT reformulation of the two-level value function $\varphi_{op}(x) := \min_y \{-F(x,y) | y \in S(x)\}$. For the reader's convenience, recall that if the lower-level problem (1.3) is convex at $x \in X$, then the primal KKT/OPEC reformulation is obtained as

$$\varphi_{op}(x) = \min_y \{-F(x,y) | 0 \in \nabla_y f(x,y) + N_{K(x)}(y)\}, \tag{6.26}$$

whereas the following KKT reformulation is derived from Lemma 5.1.1

$$\varphi_{op}(x) = \min_{y,u} \{-F(x,y) | \mathscr{L}(x,y,u) = 0,\ u \geq 0,\ g(x,y) \leq 0,\ u^\top g(x,y) = 0\}, \tag{6.27}$$

provided that the Slater CQ (3.6) is additionally satisfied at x.

6.2.1 M-type optimality conditions

As before, the following solution set-valued mapping related to the two-level value function (6.27) will play an important role:

$$S^h_{op}(x) := \{(y,u) \in S^h(x) | F(x,y) + \varphi_{op}(x) \geq 0\}. \tag{6.28}$$

The set-valued mapping S^h here is defined in (5.5). Next we provide the first M-type stationarity conditions for problem (P_p). Note that the qualification conditions (A^m_1) and (A^m_2) widely used in this subsection are defined in (5.7).

Theorem 6.2.1 (M-type stationarity conditions for the pessimistic bilevel program)**.** *Let $\bar{x}$ be an upper-level regular* (3.72) *local optimal solution for (P_p), where the lower-level problem* (1.3) *is convex and the lower-level regularity* (3.65) *is satisfied at $(\bar{x},y)$, $y \in S(\bar{x})$. Assume that S^h_{op}* (6.28) *is inner semicompact at $\bar{x}$ while conditions (A^m_1) and (A^m_2) hold at $(\bar{x},y,u)$, for all $(y,u) \in S^h_{op}(\bar{x})$. Then, there exist $\alpha \in \mathbb{R}^k$, $(\beta^s,\gamma^s,v_s) \in \mathbb{R}^{p+m+1}$ and $(y_s,u_s) \in S^h_{op}(\bar{x})$ with $s = 1,\ldots,n+1$ such that relationship* (3.16) *and the following conditions hold all together:*

$$\sum_{i=1}^{n+1} v_s\Big(\nabla_x F(\bar{x},y_s) + \sum_{i=1}^{p} \beta_{is}\nabla_x g_i(\bar{x},y_s) + \sum_{l=1}^{m} \gamma_{ls}\nabla_x \mathscr{L}_l(\bar{x},y_s,u_s)\Big) + \sum_{j=1}^{k} \alpha_j \nabla G_j(\bar{x}) = 0, \tag{6.29}$$

$$\forall s = 1,\ldots,n+1,\ \nabla_y F(\bar{x},y_s) + \sum_{i=1}^{p} \beta_{is}\nabla_y g_i(\bar{x},y_s) + \sum_{l=1}^{m} \gamma_{ls}\nabla_y \mathscr{L}_l(\bar{x},y_s,u_s) = 0, \tag{6.30}$$

$$\forall s = 1,\ldots,n+1,\ \nabla_y g_{\nu^s}(\bar{x},y_s)\gamma^s = 0,\ \beta^s_{\eta^s} = 0, \tag{6.31}$$

$$\forall s = 1,\ldots,n+1, i \in \theta^s,\ (\beta^s_i < 0 \wedge \nabla_y g_i(\bar{x},y_s)\gamma^s < 0) \vee \beta^s_i(\nabla_y g_i(\bar{x},y_s)\gamma^s) = 0, \tag{6.32}$$

$$\sum_{i=1}^{n+1} v_s = 1,\ \forall s = 1,\ldots,n+1,\ v_s \geq 0, \tag{6.33}$$

$$\forall s = 1,\ldots,n+1,\ \eta^s := \eta(\bar{x},y_s,u_s), \theta^s := \theta(\bar{x},y_s,u_s), \nu^s := \nu(\bar{x},y_s,u_s). \tag{6.34}$$

Proof. By Theorem 5.1.2 (ii), it follows under the assumptions of the theorem that φ_{op} (6.27) is Lipschitz continuous around $\bar{x}$ and we have

$$\partial\varphi_{op}(\bar{x}) \subseteq \bigcup_{(\bar{y},\bar{u})\in S^h_{op}(\bar{x})} \ \bigcup_{(\beta,\gamma)\in\Lambda^{em}_y(\bar{x},\bar{y},\bar{u})} \Big\{-\nabla_x F(\bar{x},\bar{y}) + \nabla_x g(\bar{x},\bar{y})^\top\beta + \nabla_x \mathscr{L}(\bar{x},\bar{y},\bar{u})^\top\gamma\Big\}.$$

Hence, the maximization two-level value function φ_p (1.8) is also Lipschitz continuous around $\bar{x}$ and by Carathédory's Theorem, we have

$$\begin{aligned}\partial\varphi_p(\bar{x}) \subseteq \ \Big\{&\textstyle\sum_{s=1}^{n+1} v_s\big(-\nabla_x F(\bar{x},y_s) + \sum_{i=1}^{p} \beta_{is}\nabla_x g_i(\bar{x},y_s) + \sum_{l=1}^{m}\gamma_{ls}\nabla_x\mathscr{L}_l(\bar{x},y_s,u_s)\big)\big| \\ &(y_s,u_s) \in S^h_{op}(\bar{x}),\ (\beta^s,\gamma^s) \in \Lambda^{em}_y(\bar{x},y_s,u_s),\ s = 1,\ldots,n+1, \\ &\textstyle\sum_{i=1}^{n+1} v_s = 1,\ v_s \geq 0,\ s = 1,\ldots,n+1\Big\}.\end{aligned} \tag{6.35}$$

Here the multipliers set $\Lambda^{em}_y(\bar{x},\bar{y},\bar{u})$ is defined as $\Lambda^{em}_y(\bar{x},\bar{y},\bar{u},-\nabla_y F(\bar{x},\bar{y}))$, see (5.7). The result is then obtained from a combination of (5.32), (6.8) and (6.35). □

As explained in Section 6.1, it may be important for numerical reasons to obtain optimality conditions where no "multiplier" is attached to the gradient of the upper-level objective function. To attain such an objective in the M-type conditions, our first approach is still based on the KKT reformulation (6.26) of the two-level value function. However assumption (A_1^m) in the previous result is replaced by the following one

$$(A_1^m)' \quad \left.\begin{array}{r} \nabla g(\bar{x},\bar{y})^\top \sum_{s=1}^{2p+1} \mu_s v^s + \nabla_{x,y}\mathscr{L}(\bar{x},\bar{y},\bar{u})^\top \gamma = 0 \\ \sum_{s=1}^{2p+1} \mu_s u^s + \nabla_y g(\bar{x},\bar{y})\gamma = 0 \\ \forall s = 1,\dots,2p+1,\, u_\nu^s = 0,\, v_\eta^s = 0 \\ \forall s = 1,\dots,2p+1, i \in \theta,\, (u_i^s > 0 \wedge v_i^s > 0) \vee u_i^s v_i^s = 0 \\ \sum_{s=1}^{2p+1} \mu_s = 1,\, \forall s = 1,\dots,2p+1,\, \mu_s \geq 0 \end{array}\right\} \Longrightarrow \left\{ \begin{array}{l} \gamma = 0 \\ \sum_{s=1}^{2p+1} \mu_s v^s = 0 \end{array}\right.$$

while imposing instead the inner semicontinuity on the solution set-valued mapping S_{op}^h (6.28).

Theorem 6.2.2 (KKT-type M-stationarity conditions for the pessimistic bilevel program via the KKT reformulation)**.** *Let $\bar{x}$ be an upper-level regular* (3.72) *local optimal solution for (P_p), where the lower-level problem* (1.3) *is convex and the lower-level regularity* (3.65) *holds at $(\bar{x},y)$, $y \in S(\bar{x})$. Assume that S_{op}^h* (6.28) *is inner semicontinuous at $(\bar{x},\bar{y},\bar{u})$, where conditions $(A_1^m)'$ and (A_2^m) are satisfied. Furthermore, suppose that* $\mathrm{co}N_\Pi(\bar{u},-g(\bar{x},\bar{y}))$ (Π *defined in* (3.95)) *is a closed set. Then, there exist $(\alpha,\gamma) \in \mathbb{R}^{k+m}$ and $(u^s, v^s, \mu_s) \in \mathbb{R}^{2p+1}$ with $s = 1,\dots,2p+1$ such that condition* (3.16) *together with the following ones are satisfied:*

$$\nabla_x F(\bar{x},\bar{y}) + \sum_{j=1}^{k} \alpha_j \nabla G_j(\bar{x}) + \sum_{i=1}^{p}\sum_{s=1}^{2p+1} \mu_s v_i^s \nabla_x g_i(\bar{x},\bar{y}) + \sum_{l=1}^{m} \gamma_l \nabla_x \mathscr{L}_l(\bar{x},\bar{y},\bar{u}) = 0, \tag{6.36}$$

$$\nabla_y F(\bar{x},\bar{y}) + \sum_{i=1}^{p}\sum_{s=1}^{2p+1} \mu_s v_i^s \nabla_y g_i(\bar{x},\bar{y}) + \sum_{l=1}^{m} \gamma_l \nabla_y \mathscr{L}_l(\bar{x},\bar{y},\bar{u}) = 0, \tag{6.37}$$

$$\sum_{s=1}^{2p+1} \mu_s u^s + \nabla_y g(\bar{x},\bar{y})\gamma = 0, \tag{6.38}$$

$$\forall s = 1,\dots,2p+1,\, u_\nu^s = 0,\, v_\eta^s = 0, \tag{6.39}$$

$$\forall s = 1,\dots,2p+1, i \in \theta,\, (u_i^s > 0 \wedge v_i^s > 0) \vee u_i^s v_i^s = 0, \tag{6.40}$$

$$\sum_{s=1}^{2p+1} \mu_s = 1,\, \forall s = 1,\dots,2p+1,\, \mu_s \geq 0. \tag{6.41}$$

Proof. The proof follows on the path of that of Theorem 6.1.3. Obviously the difference lies in the upper estimate of the normal cone to the graph of the set-valued mapping S^h. To proceed, first note that since the new variable u does not appear in the upper-level objective function F, after the transformation of φ_{op} obtained in (6.27), inclusion (6.25) should be replaced by

$$\begin{bmatrix} \nabla G(\bar{x})^\top \alpha + \nabla_x F(\bar{x},\bar{y}) \\ \nabla_y F(\bar{x},\bar{y}) \\ 0 \end{bmatrix} \in \overline{N}_{\mathrm{gph}S^h}(\bar{x},\bar{y},\bar{u}) \subseteq \nabla\psi(\bar{x},\bar{y},\bar{u})^\top \overline{N}_{\{0_m\}\times\Pi}(\psi(\bar{x},\bar{y},\bar{u})), \tag{6.42}$$

where the last inclusion follows from [91, Theorem 5.2 (ii)], provided the following equality holds true

$$\ker\nabla\psi(\bar{x},\bar{y},\bar{u})^\top \cap \overline{N}_{\{0_m\}\times\Pi}(\psi(\bar{x},\bar{y},\bar{u})) = \{0\}. \tag{6.43}$$

Here the set Π is given in (3.95) while the function ψ is defined by $\psi(x,y,u) := [\mathscr{L}(x,y,u),\, (u,-g(x,y))]$. From the expression of the basic normal cone to Π in (3.96), we have

$$N_\Pi(\bar{u},-g(\bar{x},\bar{y})) = \left\{ (u,v) \,\middle|\, \begin{array}{l} u_\eta = 0, v_\nu = 0, \\ \forall i \in \theta, (u_i < 0 \wedge v_i < 0) \wedge u_i v_i = 0 \end{array} \right\} \tag{6.44}$$

and by Carathéodory's Theorem, we have the following convexified normal cone

$$\mathrm{co}\,N_\Pi(\bar{u},-g(\bar{x},\bar{y})) = \left\{ \left[\begin{array}{c} \sum_{s=1}^{2p+1} \mu_s u^s \\ \sum_{s=1}^{2p+1} \mu_s v^s \end{array} \right] \left| \begin{array}{l} \sum_{s=1}^{2p+1} \mu_s = 1, \forall s=1,\ldots,2p+1,\, \mu_s \geq 0, \\ u^s_\eta = 0, v^s_\nu = 0,\, s=1,\ldots,2p+1 \\ (u^s_i < 0 \wedge v^s_i < 0) \wedge u^s_i v^s_i = 0, \forall i \in \theta,\, s=1,\ldots,2p+1 \end{array} \right. \right\}. \tag{6.45}$$

Since $\mathrm{co}\,N_\Pi(\bar{u},-g(\bar{x},\bar{y}))$ is a closed set, we have the following equality

$$\overline{N}_\Pi(\bar{u},-g(\bar{x},\bar{y})) = \mathrm{co}\,N_\Pi(\bar{u},-g(\bar{x},\bar{y})). \tag{6.46}$$

Hence, inserting the right-hand-side of equality (6.45) in (6.43), we obtain condition $(A^m_1)'$. Repeating the same process with inclusion (6.42), the optimality conditions in the theorem are derived.

It should however be mentioned that to obtain the Lipschitz continuity of φ_{op}, one needs to observe that condition (6.43) implies that

$$\ker \nabla \psi(\bar{x},\bar{y},\bar{u})^\top \cap N_{\{0_m\}\times\Pi}(\psi(\bar{x},\bar{y},\bar{u})) = \{0\}. \tag{6.47}$$

It is a simple exercise to check that this condition is equivalent to (A^m_1). It then follows from Theorem 5.1.2 (i) that φ_{op} is Lipschitz continuous around $\bar{x}$ given that condition (A^m_2) and the inner semicontinuity of S^h_{op} are all satisfied at $(\bar{x},\bar{y},\bar{u})$. □

In the next result we provide KKT-type M-stationarity conditions from the view point of the primal KKT reformulation (6.26). Instead of the inner semicontinuity of S^h_{op} (6.28), we impose the inner semicontinuity of the the following set-valued mapping

$$S^e_{op}(x) := \{y \in S^e(x) | F(x,y) + \varphi_{op}(x) \geq 0\} \tag{6.48}$$

where the set-valued mapping S^e corresponds to the generalized equation solution map in (3.4). For the qualification conditions (A^e_1), (A^e_2) and (A^e_3) imposed in the following theorem, we refer the reader to the comments preceding Corollary 5.1.5.

Theorem 6.2.3 (KKT-type M-stationarity conditions for the pessimistic bilevel program via the OPEC reformulation). *Let $\bar{x}$ be an upper-level regular* (3.72) *local optimal solution for (P_p), where the lower-level problem* (1.3) *is convex. Assume that the lower-level regularity* (3.65) *is satisfied at $(\bar{x},\bar{y})$, where S^e_{op}* (6.48) *is inner semicontinuous. Furthermore, let conditions (A^e_1), (A^e_2), (A^e_3) hold at $(\bar{x},\bar{y})$ while $co N_{gph S^e}(\bar{x},\bar{y})$ is closed. Then, there exist $\alpha \in \mathbb{R}^k$, $(\beta^s, u_s, \gamma^s, v_s) \in \mathbb{R}^{2p+m+1}$, with $s = 1,\ldots,m+n+1$ such that relationship* (3.13) *together with the following conditions hold:*

$$\nabla_x F(\bar{x},\bar{y}) + \sum_{j=1}^{k} \alpha_j \nabla G_j(\bar{x}) + \sum_{s=1}^{m+n+1} v_s \Big(\sum_{i=1}^{p} \beta_{is} \nabla_x g_i(\bar{x},\bar{y}) + \sum_{l=1}^{m} \gamma_l \nabla_x \mathscr{L}_l(\bar{x},\bar{y},u_s) \Big) = 0,$$

$$\nabla_y F(\bar{x},\bar{y}) + \sum_{s=1}^{m+n+1} v_s \Big(\sum_{i=1}^{p} \beta_{is} \nabla_y g_i(\bar{x},\bar{y}) + \sum_{l=1}^{m} \gamma_l \nabla_y \mathscr{L}_l(\bar{x},\bar{y},u_s) \Big) = 0,$$

$$\forall s = 1,\ldots,m+n+1,\, \nabla_y g_{\nu^s}(\bar{x},\bar{y})\gamma^s = 0,\, \beta^s_{\eta^s} = 0,$$

$$\forall s = 1,\ldots,m+n+1, i \in \theta^s,\, (\beta^s_i < 0 \wedge \nabla_y g_i(\bar{x},\bar{y})\gamma^s < 0) \vee \beta^s_i(\nabla_y g_i(\bar{x},\bar{y})\gamma^s) = 0,$$

$$\sum_{i=1}^{m+n+1} v_s = 1, \forall s = 1,\ldots,m+n+1,\, v_s \geq 0,$$

$$\forall s = 1,\ldots,m+n+1, \eta^s := \eta(\bar{x},\bar{y},u_s), \theta^s := \theta(\bar{x},\bar{y},u_s), \nu^s := \nu(\bar{x},\bar{y},u_s).$$

Proof. Since we are working here with the primal KKT reformulation (6.26) of the two-level value function φ_{op}, one simply has to note that an upper bound of the normal cone to the graph of S^e (3.4) is obtained as

$$\begin{aligned} N_{\mathrm{gph}\, S^e}(\bar{x},\bar{y}) \subseteq \Big\{ & \textstyle\sum_{i=1}^p \beta_i \nabla g_i(\bar{x},\bar{y}) + \sum_{l=1}^m \gamma_l \nabla_{x,y}\mathscr{L}_l(\bar{x},\bar{y},\bar{u}) \Big| \\ & \bar{u} \geq 0,\, \bar{u}^\top g(\bar{x},\bar{y}) = 0,\, \mathscr{L}(\bar{x},\bar{y},\bar{u}) = 0, \\ & \nabla_y g_\nu(\bar{x},\bar{y})\gamma = 0, \beta_\eta = 0, \\ & (\nabla_y g_i(\bar{x},\bar{y})\gamma > 0 \wedge \beta_i > 0) \vee \beta_i(\nabla_y g_i(\bar{x},\bar{y})\gamma) = 0, \forall i \in \theta \Big\} \end{aligned}$$

while considering the convexity of the lower-level problem (1.3) and the assumptions (A_1^e) and (A_2^e) [93]. The rest of the proof follows as in the proof of Theorem 6.1.3, taking into account the sensitivity analysis of φ_{op} (6.26) from Corollary 5.1.5 (i). □

To close this subsection, note that by means of Corollary 5.1.5 (ii), the necessary optimality conditions (3.16), (6.29)–(6.34) obtained in Theorem 6.2.1 could also be derived using the primal KKT/OPEC reformulation (6.26) of the two-level value function. In this case assumptions (A_1^e), (A_2^e) and (A_3^e) would replace (A_1^m) and (A_2^m) while instead the inner semicompactness S_{op}^e (6.48) would be needed.

6.2.2 C-type optimality conditions

We proceed in this subsection with the derivation of C and KKT-type C-stationarity conditions for the pessimistic bilevel programming problem (P_p). We concentrate here and in the next subsection on the KKT reformulation (6.27) of the minimization two-level value function φ_{op}. As mentioned in the previous subsection, similar results can be derived via the OPEC reformulation (6.26). As usual the C-qualification conditions imposed here are those in Chapter 5 (cf. (5.14)).

Theorem 6.2.4 (C-stationarity conditions for the pessimistic bilevel program)**.** *Let $\bar{x}$ be an upper-level regular* (3.72) *local optimal solution for (P_p), where the lower-level problem* (1.3) *is convex and the lower-level regularity* (3.65) *is satisfied at $(\bar{x},y)$, $y \in S(\bar{x})$. Assume that S_{op}^h* (6.28) *is inner semicompact at $\bar{x}$ while conditions (A_1^c) and (A_2^c) hold at $(\bar{x},y,u)$, for all $(y,u) \in S_{op}^h(\bar{x})$. Then, there exist $\alpha \in \mathbb{R}^k$, $(\beta^s,\gamma^s,v_s) \in \mathbb{R}^{p+m+1}$ and $(y_s,u_s) \in S_{op}^h(\bar{x})$ with $s = 1,\ldots,n+1$ such that relationships* (6.29)-(6.31) *and* (6.33)-(6.34) *together with the following condition are satisfied:*

$$\forall s = 1,\ldots,n+1, \forall i \in \theta^s,\ \beta_i^s(\nabla_y g_i(\bar{x},y_s)\gamma^s) \geq 0.$$

Proof. Analogous to the proof of Theorem 6.2.1 (ii). For the Lipschitz continuity and the upper estimate of the basic subdifferential of φ_{op}, see Theorem 5.1.3 (ii). □

Theorem 6.2.5 (KKT-type C-stationarity conditions for the pessimistic problem)**.** *Let $\bar{x}$ be an upper-level regular* (3.72) *local optimal solution for (P_p), where the lower-level problem* (1.3) *is convex and the lower-level regularity holds at $(\bar{x},y)$, $y \in S(\bar{x})$. Assume that S_{op}^h* (6.28) *is inner semicontinuous at $(\bar{x},\bar{y},\bar{u})$, where conditions (A_1^c) and (A_2^c) are satisfied. Then, there exist $(\bar{y},\bar{u}) \in S_{op}^h(\bar{x})$ and $(\alpha,\beta,\gamma) \in \mathbb{R}^{k+p+m}$ such that the C-stationarity conditions* (3.16)*,* (3.23)–(3.25) *and* (3.29) *are satisfied.*

Proof. It follows in the pattern of that of Theorem 6.2.2. However, one should note that the graph of S^h (5.5) can take the following form

$$\mathrm{gph}\, S^h = \{(x,y,u) | \mathscr{L}(x,y,u) = 0, \min\{u_i, -g_i(x,y)\} = 0, i = 1,\ldots,p\}.$$

Applying [91, Proposition 5.8], it follows that

$$\overline{N}_{\mathrm{gph}\, S^h}(\bar{x},\bar{y},\bar{u}) \subseteq \left\{ \begin{bmatrix} \mathscr{L}(\bar{x},\bar{y},\bar{u})^\top \gamma \\ \nabla_y g_\nu(\bar{x},\bar{y})\gamma \end{bmatrix} + \sum_{i=1}^p \lambda_i \bar{\partial} V_i(\bar{x},\bar{y},\bar{u}) \right\} \tag{6.49}$$

provided that the following quafication condition holds true:

$$0 \in \begin{bmatrix} \mathscr{L}(\bar{x},\bar{y},\bar{u})^\top \gamma \\ \nabla_y g_\nu(\bar{x},\bar{y})\gamma \end{bmatrix} + \sum_{i=1}^{p} \lambda_i \bar{\partial} V_i(\bar{x},\bar{y},\bar{u}) \Longrightarrow \gamma = 0,\, \lambda_i = 0,\, i = 1,\dots,p, \tag{6.50}$$

where $V_i(x,y,u) := \min\{u_i, -g_i(x,y)\}$ for $i = 1,\dots,p$. We have from inclusion (3.119) (also see the proof of Theorem 4.1.6) that

$$\sum_{i=1}^{p} \lambda_i \bar{\partial} V_i(\bar{x},\bar{y},\bar{u}) \subseteq \left\{ \begin{bmatrix} -\nabla_y g(\bar{x},\bar{y})^\top \beta \\ \xi \end{bmatrix} \middle| \begin{array}{l} \beta_\eta = 0,\, \xi_\nu = 0, \\ \beta_i \xi_i \geq 0,\, \forall i \in \theta \end{array} \right\}. \tag{6.51}$$

Combining (6.50) and (6.51), it is easy to observe that (A_1^c) is a sufficient condition for implication (6.50) to be satisfied. Moreover, the optimality conditions in the theorem are obtained by successively inserting (6.49) and (6.51) in the inclusion

$$\begin{bmatrix} \nabla G(\bar{x})^\top \alpha + \nabla_x F(\bar{x},\bar{y}) \\ \nabla_y F(\bar{x},\bar{y}) \\ 0 \end{bmatrix} \in \overline{N}_{\mathrm{gph}S^h}(\bar{x},\bar{y},\bar{u}),$$

obtained as in the proof of Theorem 6.2.2. □

Clearly, the C-stationarity conditions obtained in the latter result coincide with those of the optimistic problem. However, for a more clear picture on the relationship between P_o-C-stationarity and a similar stationarity concepts for the pessimistic problem, we refer the reader to Subsection 6.2.4.

6.2.3 S-type optimality conditions

We start this subsection by deriving the S-counterpart of Theorem 6.2.1 and Theorem 6.2.4. Recall that the S-qualification conditions (A_1^s) and (A_2^s) are given in (5.17).

Theorem 6.2.6 (S-stationarity conditions for the pessimistic bilevel program)**.** *Let $\bar{x}$ be an upper-level regular* (3.72) *local optimal solution for (P_p), where the lower-level problem* (1.3) *is convex and the lower-level regularity* (3.65) *is satisfied at $(\bar{x},y)$, $y \in S(\bar{x})$. Assume that S^h_{op}* (6.28) *is inner semicompact at $\bar{x}$ while* EITHER $[(A_1^m) \wedge (A_2^m)]$ OR $[(A_1^c) \wedge (A_2^c)]$ *hold at $(\bar{x},y,u)$, for all $(y,u) \in S^h_{op}(\bar{x})$. Furthermore, we suppose that (A_1^s) also holds at $(\bar{x},y,u)$, for all $(y,u) \in S^h_{op}(\bar{x})$. Then, there exist $\alpha \in \mathbb{R}^k$, $(\beta^s,\gamma^s,v_s) \in \mathbb{R}^{p+m+1}$ and $(y_s,u_s) \in S^h_{op}(\bar{x})$ with $s = 1,\dots,n+1$ such that relationships* (6.29)-(6.31) *and* (6.33)-(6.34) *together with the following condition are satisfied:*

$$\forall s = 1,\dots,n+1, \forall i \in \theta^s,\, \beta_i^s \leq 0 \wedge \nabla_y g_i(\bar{x},y_s)\gamma^s \leq 0.$$

Proof. Follows in the line of that of Theorem 6.2.1 while applying Theorem 5.1.4 (ii). □

In Theorem 6.2.6, the combination of assumptions (A_1^m) and (A_2^m) on one hand and of (A_1^c) and (A_2^c) on the other can be replaced by the single assumption (A_3^m) and (A_3^c), respectively. Next we provide a more general setting of S-stationarity conditions for the pessimistic problem that coincide with those of Theorem 6.2.6 provided the vector $(\beta,\gamma,\bar{y},\bar{u})$ is unique in the latter result. The following occur however without requiring the aforementioned uniqueness.

Theorem 6.2.7 (KKT-type S-stationarity conditions for the pessimistic bilevel program)**.** *Let $\bar{x}$ be an upper-level regular* (3.72) *local optimal solution for (P_p), where the lower-level problem* (1.3) *is convex and the lower-level regularity* (3.65) *holds at $(\bar{x},y)$, $y \in S(\bar{x})$. Assume that S^h_{op}* (6.28) *is inner semicontinuous at $(\bar{x},\bar{y},\bar{u})$, where* EITHER $[(A_1^c) \wedge (A_2^c) \wedge (A_1^s)]$, $[(A_3^c) \wedge (A_1^s)]$ OR (A_2^s) *hold. Then, there exists $(\alpha,\beta,\gamma) \in \mathbb{R}^{k+p+m}$ such that conditions* (3.16) *and* (3.23)–(3.25) *together with the following one hold:*

$$\nabla_y g_i(\bar{x},\bar{y})\gamma \leq 0 \wedge \beta_i \leq 0, \forall i \in \theta. \tag{6.52}$$

Proof. First note that in addition to the convexity of problem (1.3) and the lower-level regularity (3.65) at $(\bar{x}, y)$, $y \in S(\bar{x})$, if EITHER $[(A_1^c) \wedge (A_2^c)]$ OR $[(A_3^c)]$ is satisfied, we have the C-stationarity conditions in Theorem 6.2.5. Moreover if (A_s^1) holds, then we have our KKT-type S-stationarity conditions for (P_p). Secondly, if instead (A_2^s) holds, we have from Theorem 5.1.4 (i) (cf. (5.19)) that the two-level value function φ_{op} (6.27) is strictly differentiable at $\bar{x}$ and

$$\nabla \varphi_{op}(\bar{x}) = -\nabla_x F(\bar{x}, \bar{y}) + \sum_{j=1}^{k} \alpha_j \nabla G_j(\bar{x}) + \sum_{i=1}^{p} \beta_i \nabla_x g_i(\bar{x}, \bar{y}) + \sum_{l=1}^{m} \gamma_l \nabla_x \mathscr{L}_l(\bar{x}, \bar{y}, \bar{u}),$$

where (β, γ) is the unique multiplier vector satisfying (3.25) and (3.30) together with condition:

$$-\nabla_x F(\bar{x}, \bar{y}) + \sum_{i=1}^{p} \beta_i \nabla_x g_i(\bar{x}, \bar{y}) + \sum_{l=1}^{m} \gamma_l \nabla_x \mathscr{L}_l(\bar{x}, \bar{y}, \bar{u}) = 0.$$

Thus the result then follows while changing the signs of β and γ as done in the previous cases. □

Based on this simple form of the optimality conditions of the pessimistic problem and the observations already made above for the other types of stationarity concepts, we will try, in the next subsection, to highlight some links between the optimality conditions of the latter formulation of the bilevel program and those of the optimistic one.

6.2.4 Pessimistic versus optimistic bilevel programming

As mentioned in the Introduction, the optimistic and pessimistic bilevel optimization problems (P_o) and (P_p) are usually distinct from each other when we can not guaranty uniqueness in the lower-level problem (1.3). In this subsection we provide a particular setting where the necessary optimality conditions of both problems coincide. For simplicity in our comparison we consider only the C- and S-type stationarity conditions. Furthermore, we focus here only on the KKT-type C- and S-stationarity conditions of the pessimistic problem obtained in Theorem 6.2.5 and Theorem 6.2.7, that we denote here by P_p-C and P_p-S-stationarity conditions, respectively, in order to avoid confusion with those of (P_o) denoted similarly while replacing P_p by P_o, cf. Definition 5.2.1. Precisely, $\bar{x}$ will be said to be P_p-C-STATIONARY (resp. P_p-S-STATIONARY) if there exist $(\bar{y}, \bar{u}) \in S^h_{op}(\bar{x})$ and $(\alpha, \beta, \gamma) \in \mathbb{R}^{k+p+m}$ such that relationships (3.16) and (3.23)-(3.25) together with the condition (3.29) (resp. (6.52)) are satisfied. For the latter definition, recall that the solution set-valued mapping S^h_{op} is defined in (6.28), while the mapping S^h_o needed in the following corollary is given by (5.6).

The aim of the next consequences of previous results is to show that necessary optimality conditions of a pessimistic bilevel program can be obtained from those of its optimistic counterpart and vice-versa.

Corollary 6.2.8 (derivation of P_p-C and P_p-S-stationarity from P_o-C and P_o-S-stationarity, respectively)**.** *Let $\bar{x}$ be an upper-level regular* (3.72) *local optimal solution of the optimistic bilevel program (P_o), where the lower-level problem* (1.3) *is convex. Assume that the Slater CQ* (3.6) *holds at $\bar{x}$ while relationships (A_1^c) and (A_2^c) are satisfied at $(\bar{x}, \bar{y}, \bar{u})$. Furthermore, we suppose that we have $S^h_o(\bar{x}) = \{(\bar{y}, \bar{u})\} = S^h_{op}(\bar{x})$. Then the following assertions are satisfied:*

(i) *$\bar{x}$ is P_p-C-stationary.*

(ii) *If in addition, (A_2^s) is satisfied at $(\bar{x}, \bar{y}, \bar{u})$ and* EITHER $\theta = \{i | \beta_i = 0 \wedge \nabla_y g_i(\bar{x}, \bar{y})\gamma = 0\}$ OR $\theta = \emptyset$ *(strict complementarity) holds. Then $\bar{x}$ is P_p-C-stationary.*

Proof. For (i), note that from the C-counterpart of Theorem 5.2.4 (ii) (also see the related comments following the result) where condition $S^h_{op}(\bar{x}) = \{(\bar{y}, \bar{u})\}$ here ensures the inner semicontinuity of S^h_{op} (6.28), it follows that $\bar{x}$ is P_o-C-stationary. Furthermore, also observe that $S^h_o(\bar{x}) = \{(\bar{y}, \bar{u})\} = S^h_{op}(\bar{x})$ guaranties the fulfillment of inclusion $(\bar{y}, \bar{u}) \in S^h_o(\bar{x})$ contained in the P_p-C-stationarity conditions. This completes the proof of (i). Similarly, (ii) is obtained from the S-counterpart of Theorem 5.2.4 (ii). □

Using Theorem 6.2.5 and Theorem 6.2.7, we can state a result similar to Corollary 6.2.8 (i) and (ii), respectively, while considering but a local optimal solution of (P_p), which provides P_o-C and P_o-S-stationarity from the P_p-C and P_p-S-stationarity, respectively. Note that all the assumptions of the above corollary are satisfied for the problem in Example 3.

To close this section, we consider the bilevel optimization problem (1.1) where the upper-level feasible set is defined by nonegativity inequality constraints $G_j(x) \geq 0$ for $j = 1, \ldots, k$. Next consider the corresponding pessimistic reformulation

$$\min_x \{\varphi_p(x) | \, G(x) \geq 0\} \text{ with } \varphi_p(x) := \max_y \{F(x,y) | \, y \in S(x)\}. \tag{6.53}$$

Furthermore we introduce an optimistic problem related to the aforementioned bilevel program, where the upper-level objective function is the negative of that of the initial problem:

$$\min_x \{\varphi_{op}(x) | \, G(x) \geq 0\} \text{ with } \varphi_{op}(x) := \min_y \{-F(x,y) | \, y \in S(x)\}. \tag{6.54}$$

In the next result we show that the optimality conditions of the pessimistic problem (6.53) can be obtained from those of (6.54) and vice-versa.

Proposition 6.2.9 (optimality conditions for pessimistic problem via an optimistic problem)**.** *Let $\bar{x}$ be an upper-level regular* (3.72) *local optimal solution problem* (6.54)*, where the lower-level problem* (1.3) *is convex and the Slater CQ holds at $\bar{x}$. Furthermore, suppose that S^h_{op}* (6.28) *is inner semicompact at $\bar{x}$. Then, the following assertions hold:*

(i) *If (A^c_1) and (A^c_2) hold at $(\bar{x}, y, u)$, for all $(y,u) \in S^h_{op}(\bar{x})$. Then, $\bar{x}$ is P_p-C-stationary.*

(ii) *If (A^c_1), (A^c_2) and (A^s_1) hold at $(\bar{x}, y, u)$, for all $(y,u) \in S^h_{op}(\bar{x})$. Then, $\bar{x}$ is P_p-S-stationary.*

Proof. For (i), note that under the assumptions of the theorem, it follows from Theorem 5.1.3 (ii) that there exist $(\bar{y}, \bar{u}) \in S^h_{op}(\bar{x})$ and $(\alpha, \beta, \gamma) \in \mathbb{R}^{k+p+m}$ such that relationships (3.16), (3.25) and (3.29) together with the following conditions are satisfied:

$$-\nabla_x F(\bar{x},\bar{y}) - \sum_{j=1}^{k} \alpha_j \nabla G_j(\bar{x}) + \sum_{i=1}^{p} \beta_i \nabla_x g_i(\bar{x},\bar{y}) + \sum_{l=1}^{m} \gamma_l \nabla_x \mathcal{L}_l(\bar{x},\bar{y},\bar{u}) = 0,$$
$$-\nabla_y F(\bar{x},\bar{y}) + \sum_{i=1}^{p} \beta_i \nabla_x g_i(\bar{x},\bar{y}) + \sum_{l=1}^{m} \gamma_l \nabla_x \mathcal{L}_l(\bar{x},\bar{y},\bar{u}) = 0.$$

Multiplying the two lines of the latter system by -1 and hence changing the signs of the multiplies β and γ, we have the P_p-C-stationarity conditions. Note that (ii) follows analogously. □

The power of this result is that it allows us to get the C- and S-stationarity conditions of the pessimistic problem of Theorems 6.2.5 and 6.2.7, under much weaker assumptions. To understand this first observe that only the inner semicompactness of S^h_{op} is needed here. Secondly assumption (A^c_2) can be dropped, cf. Remark 5.2.5 for related discussion. More clarifications on the latter point are given in the next section.

6.3 Upper subdifferential conditions for pessimistic programs

The Lipschitz continuity of the two-level value function φ_p (1.8) has been crucial in deriving the necessary optimality conditions in the previous sections of the current chapter. Most often it was obtained via the combination of $(A^{\ltimes}_1)$ and $(A^{\ltimes}_2)$ (with $\ltimes = v, m, c$). If we drop $(A^{\ltimes}_2)$, we may lost the Lipschitz continuity, cf. Remark 5.2.5. We might then end up just with the upper semicontinuity of φ_p, which is achieved only under the inner semicompactness of S_{op} (6.3) or S^h_{op} (5.5), cf. Theorem 4.0.1 for the related general framework. Hence, we would be out of the scope of Theorem 2.2.1 (ii) (used in Sections

6.1 and 6.2), where we absolutely need φ_p to be lower semicontinuous. To deal with such a case, we will consider the notion of *upper subdifferential necessary optimality* introduced by Mordukhovich [84] (see Theorem 2.2.1 (i) for the definition), where we only need $|\varphi_p(\bar{x})| < \infty$. Before stating the main result of this section, it is worth mentioning that we do not face the above difficulty while investigating necessary optimality conditions for the optimistic bilevel optimization problem (P_o), as φ_o is automatically lsc under the inner semicompactness of the solution set-valued mapping S_o (5.2) or S_o^h (5.6). Hence assumption ($A_1^{\ltimes}$) may be enough to derive the counterparts of the results of Sections 6.1–6.2 for the optimistic bilevel programming problem (P_o), cf. Section 5.2 and comments therein. In the next result, we provide a detailed version of the upper subdifferential necessary optimality conditions for the pessimistic bilevel programming problem (P_p) from (1.8). For the convenience of the reader, recall that the notion of *local upper Lipschtzian selection* imposed here is defined in Subsection 2.1.4, see (2.35).

Theorem 6.3.1 (S-type upper subdifferential optimality conditions for the pessimistic bilevel program)**.** *Let $\bar{x}$ be an upper-level regular* (3.72) *local optimal solution for (P_p), where the lower-level problem* (1.3) *is convex and the Slater CQ* (3.6) *holds at $\bar{x}$. Assume that $S_{op}^h : \operatorname{dom} S^h \rightrightarrows \mathbb{R}^{m+p}$ admits a local upper Lipschtzian selection at $(\bar{x},\bar{y},\bar{u})$. Then we have:*

$$\begin{array}{l}\forall(\beta,\gamma)\in\mathbb{R}^{m+p} \text{ satisfying } (3.24)-(3.25) \text{ and } (6.52),\\ \exists\alpha\in\mathbb{R}^k \text{ verifying } (3.16) \text{ such that } (3.23) \text{ holds true.}\end{array} \tag{6.55}$$

Proof. Since $|\varphi_p(\bar{x})| < \infty$ (given that $S_{op}^h(\bar{x}) \neq \emptyset$ by assumption), it follows from Theorem 2.2.1 (i) that the following upper subdifferential optimality condition is satisfied:

$$-\widehat{\partial}^+\varphi_p(\bar{x}) \subseteq \widehat{N}_X(\bar{x}). \tag{6.56}$$

By the definition of the upper subdifferential, we have

$$\widehat{\partial}^+\varphi_p(\bar{x}) = -\widehat{\partial}(-\varphi_p)(\bar{x}) = -\widehat{\partial}\varphi_{op}(\bar{x}). \tag{6.57}$$

Inserting (6.57) in (6.56) we have the following inclusion

$$\widehat{\partial}\varphi_{op}(\bar{x}) \subseteq N_X(\bar{x}). \tag{6.58}$$

Since $-F$ is Fréchet differentiable and S_{op}^h admits a local upper Lipschitz selection at $(\bar{x},\bar{y},\bar{u})$, it follows from [92, Theorem 2] that

$$\widehat{\partial}\varphi_{op}(\bar{x}) = -\nabla_x F(\bar{x},\bar{y}) + \widehat{D}^* S^h(\bar{x},\bar{y},\bar{u})(-\nabla_y F(\bar{x},\bar{y}),0). \tag{6.59}$$

Now recall that S^h can be reformulated as $S^h(x) := \{(y,u)|\, \psi(x,y,u) \in \{0\}\times\Pi\}$ with Π in (3.95), while $\psi(x,y,u) := [\mathscr{L}(x,y,u),\ (u,-g(x,y))]$. Hence, by [89, Corollary 1.15], we have the following lower estimate of the Fréchet normal cone to the graph of S^h

$$\nabla\psi(\bar{x},\bar{y},\bar{u})^\top \widehat{N}_{\{0\}\times\Pi}(\psi(\bar{x},\bar{y},\bar{u})) \subseteq \widehat{N}_{\operatorname{gph} S^h}(\bar{x},\bar{y},\bar{u}), \tag{6.60}$$

while taking into account that ψ is Fréchet differentiable at $(\bar{x},\bar{y},\bar{u})$. Clearly, we have from the definition of $\widehat{D}^*S^h$ (see (2.25)) that

$$\widehat{D}^*S^h(\bar{x},\bar{y},\bar{u})(-\nabla_y F(\bar{x},\bar{y}),0) \supseteq \begin{array}{l}\Big\{\sum_{l=1}^m \gamma_l\nabla_x\mathscr{L}_l(\bar{x},\bar{y},\bar{u}) - \sum_{i=1}^p \beta_i\nabla_x g_i(\bar{x},\bar{y})\Big|\\ \gamma\in\mathbb{R}^m,\ (-\nabla_y g(\bar{x},\bar{y})\gamma,\beta)\in\widehat{N}_\Pi(\bar{u},-g(\bar{x},\bar{y})),\\ \sum_{l=1}^m \gamma_l\nabla_y\mathscr{L}_l(\bar{x},\bar{y},\bar{u}) - \sum_{i=1}^p \beta_i\nabla_y g_i(\bar{x},\bar{y}) - \nabla_y F(\bar{x},\bar{y}) = 0\Big\}.\end{array} \tag{6.61}$$

Further note that from the expression of the Fréchet normal cone to Π, see e.g. [123], we have

$$\widehat{N}_\Pi(\bar{u},-g(\bar{x},\bar{y})) = \left\{(u,v)\,\middle|\, \begin{array}{l} u_\eta = 0, v_\nu = 0,\\ \forall i\in\theta, u_i\leq 0\wedge v_i\leq 0\end{array}\right\}. \tag{6.62}$$

Obviously, combining (6.58)–(6.59) and (6.61)–(6.62), we have the result. □

One can easily observe that the S-type upper subdifferential optimality conditions in (6.55) imply the S-type lower-ones in Theorem 6.2.7. To conclude this section we consider a special case of the so-called *simple bilevel programming problem* [23] and derive upper subdifferential necessary optimality conditions for its pessimistic version

$$\min_{x\in X}\ \varphi_p(x) := \max_{y\in S} F(x,y) \text{ with } S := \arg\min\{f(y)|\,g(y)\le 0\}. \tag{6.63}$$

To simplify the presentation of the result we set $S^{n+1} := \underbrace{S\times\ldots\times S}_{n+1-\text{times}}$.

Theorem 6.3.2 (upper subdifferential optimality conditions in pessimistic simple bilevel programming)**.** *Let $\bar{x}$ be an upper-level regular* (3.72) *local optimal solution of problem* (6.63)*, where the function $-F$ is convex in (x,y) and the solution set S in* (6.63) *is convex and compact. Then we have:*

$$\begin{array}{l}\forall y := (y_s)_{s=1}^{n+1}\in S^{n+1},\, \forall\mu := (\mu_s)_{s=1}^{n+1}\in\mathbb{R}_+^{n+1} \text{ with } \sum_{s=1}^{n+1}\mu_s = 1,\\ \exists\alpha\in\mathbb{R}^k \text{ satisfying (3.16) such that the following holds:}\end{array}$$

$$\textstyle\sum_{s=1}^{n+1}\mu_s\nabla_x F(\bar{x},y_s) + \sum_{j=1}^{k}\alpha_j\nabla G_j(\bar{x}) = 0.$$

Proof. S being compact and $-F$ continuously differentiable, it follows from Danskin's Theorem (see e.g. [129, Proposition 2.1]) that the value function

$$\varphi_{op}(x) := \min\{-F(x,y)|\,y\in S\}$$

is locally Lipschitz continuous and its Clarke/convexified subdifferential is obtained as

$$\begin{array}{rl}\bar{\partial}\varphi_{op}(\bar{x}) = & \text{co}\,\{-\nabla_x F(\bar{x},y)|\,y\in S\}\\ := & \left\{-\sum_{s=1}^{n+1}\mu_s\nabla_x F(\bar{x},y_s)|\,\sum_{s=1}^{n+1}\mu_s = 1, \mu_s\ge 0,\, y_s\in S, s=1,\ldots,n+1\right\}.\end{array} \tag{6.64}$$

On the other hand, the convexity of the set S and the function $-F$ in (x,y) imply that φ_{op} is a lower regular function, and hence we have

$$\widehat{\partial}\varphi_{op}(\bar{x}) = \partial\varphi_{op}(\bar{x}) = \bar{\partial}\varphi_{op}(\bar{x}). \tag{6.65}$$

Combining (6.58), (6.64) and (6.65), completes the proof of the theorem. □

7 Applications of bilevel programming in transportation

Bilevel programming has been used to model various issues in transportation. Among the most popular ones are the bilevel road pricing problem (also known as the toll optimization problem [72]), the O-D-matrix adjustment problem and the network design problem. Important contributions to toll optimization have been made by Labbé, Marcotte and Savard, see e.g. [8, 38, 53, 72] for some of their publications with co-authors. The models they usually consider are tailored to situations where there is no congestion in the network. Thus the objective function of the traffic assignment problem, which represents the follower's problem here, is linear in terms of link flows. In the above mentioned papers, the KKT reformulation of the bilevel road pricing problem is considered, and the resulting problem is converted into a mixed-integer programming problem for which various algorithms have been proposed.

However, there are also great contributions to toll optimization by Yang et al. [57, 117, 118, 119] and many others, where congestion is taken into account, i.e. with the traffic assignment problem's objective function having a certain nonlinear structure. The technique has usually been based on sensitivity analysis of the solution mapping of the lower-level problem while applying the implicit function approach (see (1.6)) for a one-level reformulation. The aforementioned approaches for the toll problem have also been applied to the *network design* [14, 69, 79] and *O-D matrix adjustment* [11, 17, 96] problems. It should be mentioned that the major difference between these three problems resides in the structure of the the traffic assignment problem's objective function and the role played by the road authority's strategy vector corresponding to the *toll* in the road pricing problem. In the O-D adjustment problem, the latter vector (upper-level variable) is the demand vector while it denotes the capacity enhancement for the network design problem. Furthermore, in the case of the O-D adjustment problem, the demand represents a perturbation parameter for the set of feasible route flows, whereas mathematically, the network design and road pricing problems have the same structure in the sense that the lower-level problem is parameterized only in the route users total cost function.

A major draw back of the KKT reformulation usually considered by Labbé et al. [53, 8, 72, 38] is that the link between the resulting KKT reformulation (see (3.8) for a general problem) and the initial problem is unbalanced, in the sense that a local solution of the latter problem may not correspond to a local solution of the former one, cf. Section 3.1.1. Considering that we are dealing with nonconvex problems, this is an important issue in a solution process. As far as the approach by Yang et al. [57, 117, 119, 118] is concerned, note that the assumptions often made to ensure the differentiability of the traffic assignment solution function are too strong and can not be satisfied, for example, when dealing with a network where there is no congestion. Moreover, the sensitivity analysis cannot be applied to practical problems like the road accident reduction problem (to be discussed in the next section), which can be modeled as a special case of the road pricing problem. The latter is justified in the study by Netter [94], where it is showed that it is not possible to have a unique Wardrop equilibrium in a network with more than one category of road users.

To circumvent these difficulties, we suggest in this chapter to apply the LLVF reformulation (3.3) as it is completely equivalent (locally and globally) to the initial problem without any condition. In fact note that in the literature, the classical/auxiliary problem (P) has usually been the working model for the above mentioned transportation problems. Thus, we perpetuate this tradition here. However, in view of Chapters 5 and 6, it would be a simple exercise for the interested reader to deduce the results corresponding to the original optimistic problem (P_o) and the pessimistic problem (P_p), respectively.

Let us now remind that one of the things that called our attention is that very little has been done in the direction of optimality conditions for the transportation problems mentioned above. The only work that we found on this question is in the PhD thesis of Chen [11], where Fritz-John's type necessary optimality conditions were suggested for the demand adjustment problem. Part of the work in [11] was later published in [12]. In this chapter we suggest an improvement of the results by Chen and Florian by providing KKT-type conditions under even weaker assumptions on the traffic assignment problem. Before, we first suggest KKT conditions for the toll optimization problem in the next section. The results that we summarize in this chapter were first obtained in the papers [35, 36], were more details are provided.

To conclude the current section, we present the traffic assignment problem. More details on the latter problem can be found in the seminal works by Wardrop [116] and Beckman et al. [6]. See also the book by Patrisson [102] and references therein for further developments on the topic.

We consider a transportation network $\mathscr{G} = (\mathscr{N}, \mathscr{A})$, where $\mathscr{N}$ and $\mathscr{A}$ denote the set of nodes and directed links (arcs), respectively. Let $\mathscr{W} \subseteq \mathscr{N}^2$ denote the set of origin-destination (O-D) pairs. Each O-D pair $w \in \mathscr{W}$ is connected by a set of routes (paths) $\mathscr{P}_w$, each member of which is a set of sequentially connected links. We denote by $\mathscr{P} = \bigcup_{w\in\mathscr{W}} \mathscr{P}_w$ the set of all routes of the network and by $n = |\mathscr{A}|$, $l = |\mathscr{W}|$ and $m = |\mathscr{P}|$, the cardinalities of $\mathscr{A}$, $\mathscr{W}$ and $\mathscr{P}$, respectively. Let the matrix $(\Lambda = [\Lambda_{wp}]) \in \mathbb{R}^{l\times m}$ denote the O-D-route incidence matrix in which $\Lambda_{wp} = 1$ if route $p \in \mathscr{P}_w$ and $\Lambda_{wp} = 0$ otherwise, and the matrix $(\Delta = [\Delta_{ap}]) \in \mathbb{R}^{n\times m}$ denotes the arc-route incidence matrix with $\Delta_{ap} = 1$ if arc a is in route p and $\Delta_{ap} = 0$ otherwise. The network is assumed to be strongly connected, that is, at least one route joins each O-D pair.

We also consider the column vectors $(d = [d_w]) \in \mathbb{R}^l$, $(q = [q_p]) \in \mathbb{R}^m_+$ and $(v = [v_a]) \in \mathbb{R}^n$ to denote the travel demand, the route flow and arc flow, respectively. The column vectors $(c = [c_p]) \in \mathbb{R}^m_+$ and $(\tau = [\tau_a]) \in \mathbb{R}^n$ denote the route capacity and arc toll, respectively. A route flow q is feasible if it does not exceed the capacity and satisfies the O-D demand constraint $\Lambda q = d$. Let us denote by Q the set of such flows, then

$$Q = \{q \in \mathbb{R}^m_+ |\, q \leq c,\ \Lambda q = d\}. \tag{7.1}$$

A link flow v is feasible if there exists a feasible route flow q such that the flow conservation constraint $\Delta q = v$, is satisfied. Hence,

$$V = \{v \in \mathbb{R}^n |\, \exists q \in Q,\ \Delta q = v\} \tag{7.2}$$

denotes the set of feasible link flows. We let the function t from $\mathbb{R}^n \times \mathbb{R}^n$ to $\mathbb{R}^n$ denote the route cost, that is for each $a \in \mathscr{A}$, the component $t_a(\tau, v)$ of the vector $t(\tau, v)$ gives the traffic cost on the arc a, under the flow-toll couple (τ, v). We assume that the route cost is additive, thus the components of $\bar{c}(\tau, v) = \Delta^\top t(\tau, v)$ give the cost on each route $p \in \mathscr{P}$. Finally, we introduce the vector $\vartheta(\tau, v) = [\vartheta_w(\tau, v)] \in \mathbb{R}^l$ of minimum cost between each O-D pair $w \in \mathscr{W}$, that is $\vartheta_w(\tau, v) = \min_{p\in\mathscr{P}_w} \bar{c}_p(\tau, v)$.

Wardrop's user equilibrium principle [116] states that for every O-D pair $w \in \mathscr{W}$, the travel cost of the routes utilized are equal and minimal for each individual user, that is for each $w \in \mathscr{W}$ and $p \in \mathscr{P}_w$, we have

$$\begin{cases} \bar{c}_p(\tau, v) = \vartheta_w(\tau, v) & \text{if } q_p > 0 \\ \bar{c}_p(\tau, v) \geq \vartheta_w(\tau, v) & \text{if } q_p = 0 \end{cases} \tag{7.3}$$

for any fixed toll pattern $\tau \in \Gamma$. It follows from Beckmann et al. [6] that for every fixed toll pattern $\tau \in \Gamma$, the Wardrop's user equilibrium problem (7.3) is equivalent to the parametric optimization problem

$$\min_{v\in V} f(\tau, v) := \sum_{a\in\mathscr{A}} \int_0^{v_a} t_a(\tau, s)ds \tag{7.4}$$

provided that for each link $a \in \mathscr{A}$, the link cost takes the form $t_a(\tau, v_a)$; that is, it does not depend on the flow on the other links. In other words, the link costs are separable with respect to the link flows. In addition, they should also be continuous and positive. These assumptions will be maintained for the rest of the chapter such that for each toll pattern τ, Wardrop's user equilibrium arc flow will be defined as the solution of the optimization problem (7.4) also called the traffic assignment problem.

To close this section, it is important to mention that in this chapter, the expression of the total road cost f in (7.4) is simply illustrative. In fact when a network is congested, many expression for the total cost exist in the literature; we refer the interested reader to the book by Patriksson [102] for further details. In what follows, we rather assume very general expressions of f defining a continuously differentiable functions. However, the results are applicable to any Lipschitz continuous function.

7.1 Bilevel road pricing problem

We consider a road authority who intends to improve the circulation on the network $\mathscr{G}$ described above. He/she chooses road pricing as a method to modify the behavior of the road users, by setting tolls on some links of the network to discourage the use of the tolled links in favor of some perhaps abandoned of less used ones. The bilevel formulation of this problem enables the road authority to decide while considering the reactions of the road users. If we assume that for each toll pattern τ, the road users choose their origin-destination pairs in a way that favors the road authority, then the problem to be solved by the authority is the so-called optimistic bilevel problem to

$$\min_{\tau,v} \{F(\tau,v)|\ \tau \in \Gamma,\ v \in S(\tau)\}, \tag{7.5}$$

where $F(\tau,v)$ is the disutility function of the road authority who is also called the *leader*, Γ is a closed set representing the set of feasible tolls, and for any given toll $\tau \in \Gamma$, $S(\tau)$ denotes the set of optimal link flows for the collection of all the road users also called the *follower*. In other words, $S(\tau)$ is the solution set of the traffic assignment problem (7.4) under the toll pattern τ. For simplification in the presentation of the model, we first assume that all the links are tolled (we can have $\tau_a = 0$ for some link $a \in \mathscr{A}$). Below, we will discuss a possible way to introduce some fairness in the model. Obviously the model in (7.5) corresponds to the classical optimistic problem (P) discussed in the previous chapters, cf. in particular to Chapter 3. But as mentioned in the introduction of this chapter we will use here the standard terminology and left the translation of the above theory to the appropriate original optimistic and pessimistic models to the interested reader.

As discussed in [37], the model in (7.5) can be altered to tackle other hierarchical problems like the reduction of road accidents in some developing countries. The modification consist of separating the road users in two categories: the heavy goods vehicles and the rest of the users. Hence, only the heavy goods vehicles may be charged a toll equivalent to the level of risk to which they expose the other road users, by using the corresponding link. In the same way, the problem of reducing the level of pollution caused by heavy goods vehicles on some links of a network can be addressed. Many other economical or traffic improvements goals can be achieved by road pricing, see the aforementioned literature for further discussions.

A major deciding factor is the leader or road authority's objective function. Various expressions of $F(\tau,v)$ have been considered in the literature, including that of minimizing the total travel time experienced by all vehicles

$$F(\tau,v) = \sum_{a\in\mathscr{A}} v_a t_a(\tau_a, v_a). \tag{7.6}$$

The total revenue arising from toll charges can also be maximized, hence the authority's cost function takes the form

$$F(\tau,v) = \sum_{a\in\mathscr{A}} \tau_a v_a. \tag{7.7}$$

A combination of objectives (7.6) and (7.7) is also possible through a weighted sum or the maximization of the ratio of the total revenue to the total cost, that is the function:

$$F(\tau,v) = \sum_{a\in\mathscr{A}} \tau_a v_a \Big/ \sum_{a\in\mathscr{A}} v_a t_a(\tau_a, v_a). \tag{7.8}$$

In view of the structures of the upper-level objective functions in (7.6)–(7.8), we will rather focus our attention here on the structure of the lower-level problem, given that it is not a problem computing the derivatives of the above functions. However it is simply assumed that this function is continuously differentiable. As for the set of feasible tolls, it usually has the following form

$$\Gamma := \{\tau \in \mathbb{R}^n | \varsigma \leq \tau \leq \kappa\}, \tag{7.9}$$

where $\varsigma, \kappa \in \mathbb{R}^n$ represent the minimum and the maximum tolls, respectively. To proceed with our analysis, let us denote by

$$\mathscr{A}^{\varsigma} := \{a \in \mathscr{A} | \bar{\tau}_a = \varsigma_a\}$$

the set of all the links of the network having the minimum toll. We remind that since the aim of the toll setting may be to encourage road users to utilize some abandoned or less utilized routes, then we may have $\varsigma_a = 0$, for some $a \in \mathscr{A}$. Hence to correct the unfairness in our model, we may assume that

$$\emptyset \neq \mathscr{A}^o := \{a \in \mathscr{A} | \; \bar{\tau}_a = \varsigma_a = 0\} \subseteq \mathscr{A}^{\varsigma};$$

thus, allowing some links to be toll-free. We further define the set

$$\mathscr{A}^{\kappa} := \{a \in \mathscr{A} | \; \bar{\tau}_a = \kappa_a\}$$

of links with maximum tolls. It is worth mentioning that the restriction that the tolls should not exceed some certain amount is of great importance for social considerations since the road users and the community in general should not have the feeling that the road authority just intends to make as much money as possible. Finally, let $\mathscr{A}^{\gamma} := \{a \in \mathscr{A} | \varsigma_a < \bar{\tau}_a < \kappa_a\}$; then, $\mathscr{A}^{\varsigma}, \mathscr{A}^{\kappa}$ and $\mathscr{A}^{\gamma}$ form a partition of $\mathscr{A}$. Thus we have $\mathscr{A} = \mathscr{A}^{\varsigma} \cup \mathscr{A}^{\kappa} \cup \mathscr{A}^{\gamma}$.

To make the further explanations more clear, we make the following technical assumption: We assume that $\mathscr{A}$ is an ordered set; hence, each link $a \in \mathscr{A}$ is associated with an index $|a| \in \mathbb{N}$ and we define the n-dimensional vector

$$e^a := (0, \ldots, 0, 1, 0, \ldots, 0)^\top, \tag{7.10}$$

where 1 is at position $|a|$, in order to symbolize the utilization of the corresponding link by a road user.

Similarly to $\mathscr{A}$, we also assume that $\mathscr{P}$ is an ordered set such that for a route $r \in \mathscr{P}$, we associate an index $|r| \in \mathbb{N}$ and we define the m-dimensional vector e^r as in (7.10). Similarly, we also consider the set of unused routes of the network (resp. routes used at their full capacity)

$$\mathscr{P}^o := \{r \in \mathscr{P} | \bar{q}_r = 0\} \;\; (\text{resp. } \; \mathscr{P}^c := \{r \in \mathscr{P} | \bar{q}_r = c_r\}). \tag{7.11}$$

Then $\mathscr{P}$ can be partitioned into $\mathscr{P}^0, \mathscr{P}^c$ and $\mathscr{P}^u$, where $\mathscr{P}^u$ is the set of routes used but which are not at full capacity. Thus, $\mathscr{P} = \mathscr{P}^o \cup \mathscr{P}^c \cup \mathscr{P}^u$. Also of importance in the sequel is the collection $[\Lambda_w]_{w\in\mathscr{W}}$ of rows of the O-D-route incidence matrix Λ.

Now recall that we only focus here on the following LLVF reformulation of problem (7.5)

$$\min_{\tau,v} \{F(\tau,v) | \; \tau \in \Gamma, \; v \in V, \; f(\tau,v) \leq \varphi(\tau)\}, \tag{7.12}$$

where the optimal value function φ of the traffic assignment problem (7.4) is defined by

$$\varphi(\tau) := \min\{f(\tau,v) | \; v \in V\}. \tag{7.13}$$

As usual, it is assumed that the solution set-valued mapping of the traffic assignment problem is nonempty, i.e. $S(\tau) := \arg\min \{f(\tau,v) | \; v \in V\} \neq \emptyset$ for all $\tau \in \Gamma$, such that (7.13) is well-defined.

Theorem 7.1.1 (normal cone to the set of feasible link flows)**.** *Let $\bar{v} \in V$, then for all $\bar{q} \in \mathbb{R}^m$ with $\bar{q} \in Q$ and $\Delta\bar{q} = \bar{v}$, we have*

$$N_V(\bar{v}) = \Big\{ u \in \mathbb{R}^n \Big| \begin{array}{l} (\lambda_r^o)_{r\in\mathscr{P}^o} \geq 0,\ (\lambda_r^c)_{r\in\mathscr{P}^c} \geq 0, \\ \Delta^\top u = \sum\limits_{r\in\mathscr{P}^c} \lambda_r^c e^r - \sum\limits_{r\in\mathscr{P}^o} \lambda_r^o e^r + \sum\limits_{w\in\mathscr{W}} \lambda_w \Lambda_w^\top \end{array} \Big\}. \tag{7.14}$$

Proof. Recall that $V = \Delta Q$. Thus by Theorem 2.2.7, we have

$$N_Q(\bar{q}) = \Big\{ \sum_{r\in\mathscr{P}^c} \lambda_r^c e^r - \sum_{r\in\mathscr{P}^o} \lambda_r^o e^r + \sum_{w\in\mathscr{W}} \lambda_w \Lambda_w^\top \Big| \ (\lambda_r^o)_{r\in\mathscr{P}^o} \geq 0,\ (\lambda_r^c)_{r\in\mathscr{P}^c} \geq 0 \Big\}. \tag{7.15}$$

Applying [109, Theorem 6.43], while considering the fact that the mapping $q \to \Delta q$ is linear and the set Q is convex, it follows that

$$N_V(\bar{v}) = \{u \in \mathbb{R}^n |\ \Delta^\top u \in N_Q(\bar{q})\},\ \forall \bar{q} \in \mathbb{R}^m :\ \bar{q} \in Q,\ \Delta\bar{q} = \bar{v}. \tag{7.16}$$

Inserting (7.15) in (7.16), the result follows. □

Theorem 7.1.2 (sensitivity analysis of the traffic assignment value function with no perturbation on the feasible link flows)**.** *The optimal value function φ* (7.13) *is Lipschitz continuous around any $\bar{\tau} \in \Gamma$, and its basic subdifferential is estimated as*

$$\partial\varphi(\bar{\tau}) \subseteq \{\nabla_\tau f(\bar{\tau}, \bar{v}) |\ \bar{v} \in S(\bar{\tau})\}. \tag{7.17}$$

Proof. We have $Q \subseteq |c|\mathbb{B}$, where $\mathbb{B}$ is the unit ball of $\mathbb{R}^m$ and $|c| := \max\{c_i | i = 1,\dots,m\}$ (c is the route capacity vector). Hence $V := \Delta Q \subseteq \Delta(|c|\mathbb{B})$ is a closed and bounded set given that $q \to \Delta q$ is a continuous function and Q is also a closed set. In addition to the continuous differentiability of f, inclusion (7.17) follows from Theorem 4.0.1 (ii), while noting that $D^*V(\bar{d}, \bar{v})(v) = 0$ for all $\bar{v} \in V(\bar{d}) := V$ and $\bar{v} \in \mathbb{R}^n$. Thanks to the latter expression of the coderivative of the constant mapping V, we have from (2.30) that φ is Lipschitz continuous around $\bar{\tau}$. □

Theorem 7.1.3 (necessary optimality conditions for the bilevel road pricing problem in a congested network)**.** *Let $(\bar{\tau}, \bar{v})$ be a local optimal solution to problem* (7.12)*, which is assumed to be partially calm at $(\bar{\tau}, \bar{v})$. Then there exist $\lambda > 0$, $(\lambda^\varsigma, \lambda^\kappa, \lambda^o, \lambda^c, \lambda^l)$, $\bar{q} \in \mathbb{R}^m$, $v_s \in S(\bar{\tau})$ and u_s with $s = 1,\dots,n+1$ such that we have:*

$$\nabla_\tau F(\bar{\tau}, \bar{v}) + \lambda \nabla_\tau f(\bar{\tau}, \bar{v}) - \lambda \sum_{s=1}^{n+1} u_s \nabla_\tau f(\bar{\tau}, v_s) = \sum_{a\in\mathscr{A}^\varsigma} \lambda_a^\varsigma e^a - \sum_{a\in\mathscr{A}^\kappa} \lambda_a^\kappa e^a, \tag{7.18}$$

$$\Delta^\top \big(\nabla_v F(\bar{\tau}, \bar{v}) + \lambda \nabla_v f(\bar{\tau}, \bar{v})\big) = \sum_{r\in\mathscr{P}^o} \lambda_r^o e^r - \sum_{r\in\mathscr{P}^c} \lambda_r^c e^r - \sum_{w\in\mathscr{W}} \lambda_w \Lambda_w^\top, \tag{7.19}$$

$$0 \leq \bar{q} \leq c,\ \Lambda\bar{q} = d,\ \Delta\bar{q} = \bar{v}, \tag{7.20}$$

$$\forall s = 1,\dots,n+1,\ u_s \geq 0,\ \sum_{s=1}^{n+1} u_s = 1, \tag{7.21}$$

$$\lambda^\varsigma = (\lambda_a^\varsigma) \geq 0,\ \lambda^\kappa = (\lambda_a^\kappa) \geq 0,\ \lambda^o = (\lambda_r^o) \geq 0,\ \lambda^c = (\lambda_r^c) \geq 0,\ \lambda^l = (\lambda_w). \tag{7.22}$$

If in addition $S(\bar{\tau}) = \{\bar{v}\}$, then there exist $\lambda > 0$, $(\lambda^\varsigma, \lambda^\kappa, \lambda^o, \lambda^c, \lambda^l)$, and $\bar{q} \in \mathbb{R}^m$ such that relationships (7.19)-(7.22)*, together with the following condition are satisfied:*

$$\nabla_\tau F(\bar{\tau}, \bar{v}) = \sum_{a\in\mathscr{A}^\varsigma} \lambda_a^\varsigma e^a - \sum_{a\in\mathscr{A}^\kappa} \lambda_a^\kappa e^a. \tag{7.23}$$

Proof. Under the partial calmness condition, it follows from Theorem 3.2.2 and Proposition 2.2.5, that there exists $\lambda > 0$ such that

$$0 \in \nabla F(\bar{\tau},\bar{v}) + \lambda(\nabla f(\bar{\tau},\bar{v}) + \partial(-\varphi)(\bar{\tau}) \times 0) + N_{\Gamma\times V}(\bar{\tau},\bar{v}). \tag{7.24}$$

The result follows from (7.17) (while using Carathéodory's Theorem to deduce $\partial(-\varphi)(\bar{\tau})$) and equality (7.14), while noting that $N_{V\times\Gamma}(\bar{\tau},\bar{v}) = N_V(\bar{v}) \times N_\Gamma(\bar{\tau})$ with

$$N_\Gamma(\bar{\tau}) = \Big\{ -\sum_{a\in\mathscr{A}^\varsigma} \lambda_a^\varsigma e^a + \sum_{a\in\mathscr{A}^\kappa} \lambda_a^\kappa e^a \big| (\lambda_a^\varsigma)_{a\in\mathscr{A}^\varsigma} \geq 0, (\lambda_a^\kappa)_{a\in\mathscr{A}^\kappa} \geq 0 \Big\}, \tag{7.25}$$

which is obtained by applying Theorem 2.2.7. Obviously, if $S(\bar{\tau}) = \{\bar{v}\}$, we have from Theorem 7.1.2 that $\mathrm{co}\,\partial\varphi(\bar{\tau}) = \{\nabla_\tau f(\bar{\tau},\bar{v})\}$, which generates condition (7.23). □

It should be clear that the condition $S(\bar{\tau}) = \{\bar{v}\}$ imposed in this Corollary is far away from the usual strong assumptions made in the sensitivity analysis approaches mentioned in the Introduction. In fact, in the latter cases, it is usually required that the traffic assignment problem admits a unique optimal solution in a certain neighborhood. In particular, the strict monotonicity of the link costs $t_a (a \in \mathscr{A})$, often needed is not satisfied in the framework of Theorem 7.1.4 below.

For the next result, we assume that the cost function of the traffic assignment problem is bilinear, i.e. $f(\tau,v) = v^\top\tau$, which corresponds to the ideal case, where there is no congestion in the network. This framework has been considered by many authors; in particular by Labbé et al. [53, 8, 72, 38].

Theorem 7.1.4 (necessary optimality conditions for the bilevel road pricing problem in the case of no congestion)**.** *Let* $(\bar{\tau},\bar{v})$ *be an optimal solution to problem* (7.12)*, where* $f(\tau,v) = v^\top\tau$*. Then there exist* $\widetilde{v} \in S(\bar{\tau})$*,* $\lambda > 0$*,* $(\lambda^\varsigma,\lambda^\kappa,\lambda^o,\lambda^c,\lambda^l)$ *and* $\bar{q} \in \mathbb{R}^m$ *such that* (7.20) *and* (7.22)*, together with the following conditions are satisfied:*

$$\nabla_\tau F(\bar{\tau},\bar{v}) + \lambda(\bar{v} - \widetilde{v}) = \sum_{a\in\mathscr{A}^\varsigma} \lambda_a^\varsigma e^a - \sum_{a\in\mathscr{A}^\kappa} \lambda_a^\kappa e^a,$$

$$\Delta^\top\big(\nabla_v F(\bar{\tau},\bar{v}) + \lambda\bar{\tau}\big) = \sum_{r\in\mathscr{P}^o} \lambda_r^o e^r - \sum_{r\in\mathscr{P}^c} \lambda_r^c e^r - \sum_{w\in\mathscr{W}} \lambda_w \Lambda_w^\top.$$

Proof. We recall from Theorem 3.2.6 that with $f(\tau,v) = v^\top\tau$, problem (7.12) is partially calm at $(\bar{\tau},\bar{v})$. Hence, proceeding as in the proof of Theorem 7.1.3, it remains to observe that inlcusion (7.17) implies $\partial\varphi(\bar{\tau}) \subseteq S(\bar{\tau}) = \mathrm{co}\,S(\bar{\tau})$, given that $S(\bar{\tau})$ is convex in this case. □

7.2 Estimation of the O-D matrix

The (fixed) demand vector d needed in the road pricing problem and particularly in the traffic assignment problem (7.4) is a crucial datum given that a good decision process highly depends on how accurate it is estimated. The origin-destination (O-D) matrix estimation or O-D demand adjustment problem (DAP) is important not only for the road pricing problem, but also for many other decision-making frameworks of transportation planing. The modeling of this problem has evolved over the years, see [1, 11, 12] for extensive reviews. The bilevel formulation was pioneered by Fisk [47]. Since then, many researchers have adopted this model which usually takes the classical optimistic form:

$$\min_{d,v} \{F(d,v) |\ d \in D,\ v \in S(d)\}, \tag{7.26}$$

where $D \subseteq \mathbb{R}^l$ is a closed set and $S(d)$ is the solution set of the traffic assignment problem

$$\min_v \{f(d,v) |\ v \in V(d)\} \tag{7.27}$$

parameterized by d, also called O-D demand and representing the O-D matrix organized as a vector. In the light of (7.1)–(7.2), the set-valued mappings

$$V(d) := \{v \in \mathbb{R}^n |\ \exists q \in Q(d),\ \Delta q = v\} \text{ and } Q(d) := \{q \in \mathbb{R}^m_+ |\ q \leq c,\ \Lambda q = d\} \tag{7.28}$$

denote the set of feasible link flows and feasible route flows, respectively, for a given demand vector d.

In the DAP, the lower-level objective function is often chosen as

$$f(d,v) := \sum_{a\in\mathscr{A}} \int_0^{v_a} t_a(s)ds. \tag{7.29}$$

Clearly, this function is convex in (d,v). Thus having in mind the discussion in the introduction of this chapter, we still insist on working with general expressions of the total route cost functions. However, with this observation on f in (7.29), it is clear that requiring the full convexity of this function in a general model of DAP in an affordable assumption. Hence, we assume throughout this section that f in (7.27) is a general function of f which is convex. As before, we refer the reader e.g. to [102] for further discussions on this issue. As for the upper-level objective function F in (7.26), it is usually has the form

$$F(d,v) := \gamma_1 F_1(d,\hat{d}) + \gamma_2 F_2(v,\hat{v}),$$

where $\hat{d}$ represents the target O-D matrix that may be obtained from sample surveys, and $\hat{v}$ denotes the vector of flows observed on some links of the network. The function $F_1(d,\hat{d})$ represents the error measurement between the target O-D matrix $\hat{d}$ and the estimated matrix d, while $F_2(v,\hat{v})$ denotes the error measurement between the observed link flow $\hat{v}$ and the estimated flow v. The parameters γ_1 and γ_2 represent the uncertainty in the information contained in $\hat{d}$ and $\hat{v}$, respectively. In the line of Migdalas [80], the set D can be considered analogously to Γ (7.9). To remain closer to the framework of Chen and Florian [12], we consider $D := \{d \in \mathbb{R}^m |\ d \geq 0\}$. However the results obtained here can easily be extended to a more general expression in view of the previous chapters.

As mentioned in the introduction of the chapter, the sensitivity analysis of the solution of the traffic assignment problem has also been used to tackle the O-D matrix estimation problem. Recall the references [1, 11, 12, 17, 96] for various methodological approaches in solving the problem. Problem (7.26) can be reformulated using the following LLVF reformulation

$$\min_{d,v} \{F(d,v)|\ d \in D,\ v \in V(d),\ f(d,v) - \varphi(d) \leq 0\} \tag{7.30}$$

while assuming that $S(d) := \arg\min_v \{f(d,v)|\ v \in V(d)\} \neq \emptyset$ for all $d \in D$; thus the optimal value function of the traffic assignment problem (7.27) φ is well-defined by

$$\varphi(d) := \min_v \{f(d,v)|\ v \in V(d)\}. \tag{7.31}$$

Chen and Florian [12] also considered this reformulation, however in the following simplified form

$$\min_{d,v} \{F(d,v)|\ d,q \geq 0,\ \Lambda q = d,\ f(d,v) - \varphi(d) \leq 0\} \tag{7.32}$$

while considering the constraint $v = \Delta q$ as exogenous. Fritz-John's type optimality conditions were then derived for (7.32). Our aim here is to suggest KKT-type optimality conditions for the more general problem with the flow conservation constraint $v = \Delta q$ being fully part of the feasible set of the traffic assignment problem (7.27). To proceed, the following set of route flows

$$\mathscr{H}(\bar{d},\bar{v}) := \{q \in \mathbb{R}^m |\ \Delta q = \bar{v},\ -\bar{d} + \Lambda q = 0,\ 0 \leq q \leq c\} \tag{7.33}$$

obtained for a given couple $(\bar{d},\bar{v})$ of demand-link flow, will play an important role. Next, we provide a sensitivity analysis of the mapping V (7.28). Here, the vectors e^r and Λ_w are defined as in the previous section, whereas $[e^w]_{w\in\mathscr{W}}$ denotes the collection of rows of the identity matrix of $\mathbb{R}^{l\times l}$.

Theorem 7.2.1 (sensitivity analysis of the feasible link flows mapping)**.** *Consider* $(\bar{d},\bar{v}) \in \text{gph}V$ *and* $v \in \mathbb{R}^n$*, then the coderivative of the mapping* V (7.28) *can be estimated as:*

$$D^*V(\bar{d},\bar{v})(v) \subseteq \bigcup_{\bar{q}\in\mathscr{H}(\bar{d},\bar{v})} \Big\{ -\sum_{w\in\mathscr{W}} \lambda_w e^w \Big| \; (\lambda_w)_{w\in\mathscr{W}} \in \mathbb{R}^l,\, (\lambda_r^o)_{r\in\mathscr{P}^0(\bar{q})},\, (\lambda_r^c)_{r\in\mathscr{P}^c(\bar{q})} \geq 0, \quad \Delta^\top v - \sum_{r\in\mathscr{P}^c(\bar{q})} \lambda_r^c e^r + \sum_{r\in\mathscr{P}^o(\bar{q})} \lambda_r^o e^r - \sum_{w\in\mathscr{W}} \lambda_w \Lambda_w^\top = 0 \Big\}. \tag{7.34}$$

Moreover, V *is Lipschitz-like around* $(\bar{d},\bar{v})$*, provided the following CQ holds at* $(\bar{d},\bar{v})$*:*

$$\left.\begin{array}{r} \sum_{r\in\mathscr{P}^c(\bar{q})} \lambda_r^c e^r - \sum_{r\in\mathscr{P}^o(\bar{q})} \lambda_r^o e^r + \sum_{w\in\mathscr{W}} \lambda_w \Lambda_w^\top = 0 \\ (\lambda_r^o)_{r\in\mathscr{P}^0(\bar{q})}, (\lambda_r^c)_{r\in\mathscr{P}^c(\bar{q})} \geq 0 \\ \bar{q} \in \mathscr{H}(\bar{d},\bar{v}) \end{array}\right\} \Longrightarrow [\lambda_w = 0,\, w \in \mathscr{W}]. \tag{7.35}$$

Proof. Since $Q(d) \subseteq |c|\mathbb{B}$, for all $d \in \mathbb{R}^l$, where $\mathbb{B}$ is the unit ball of $\mathbb{R}^m$ and $|c| := \max\{c_i | i = 1,\dots,m\}$ (c is the route capacity vector), then the set-valued mapping $(d,v) \rightrightarrows Q(d) \cap J^{-1}(v) = \mathscr{H}(d,v)$ is uniformly bounded around $(\bar{d},\bar{v})$. Hence, including the continuous differentiability of the function J, it follows from chain rule in [109, Exercise 10.39] that:

$$\begin{aligned} D^*V(\bar{d},\bar{v})(v) &\subseteq \bigcup_{\bar{q}\in\mathscr{H}(\bar{d},\bar{v})} D^*Q(\bar{d},\bar{q})(-\Delta^\top v) \\ &\subseteq \bigcup_{\bar{q}\in\mathscr{H}(\bar{d},\bar{v})} \{d \in \mathbb{R}^l |\, (d, -\Delta^\top v) \in N_{\text{gph}\,Q}(\bar{d},\bar{q})\}. \end{aligned} \tag{7.36}$$

On the other hand, note that $\text{gph}Q = \{(d,q) \in \mathbb{R}^l \times \mathbb{R}^m | -d + \Lambda q = 0, 0 \leq q \leq c\}$. Thus by Theorem 2.2.7, we have

$$N_{\text{gph}\,Q}(\bar{d},\bar{q}) = \left\{ \left[\begin{array}{l} -\sum_{w\in\mathscr{W}} \lambda_w e^w \\ \sum_{r\in\mathscr{P}^c(\bar{q})} \lambda_r^c e^r - \sum_{r\in\mathscr{P}^o(\bar{q})} \lambda_r^o e^r + \sum_{w\in\mathscr{W}} \lambda_w \Lambda_w^\top \end{array} \right] \middle| \begin{array}{l} (\lambda_r^o)_{r\in\mathscr{P}^o(\bar{q})} \geq 0, \\ (\lambda_r^c)_{r\in\mathscr{P}^c(\bar{q})} \geq 0 \end{array} \right\} \tag{7.37}$$

Obviously, inclusion (7.34) is obtained by inserting the expression of the normal cone to gph$Q$ from (7.37) in (7.36).

For the Lipschitz-like property of V, observe that from inclusion (7.34), we have

$$D^*V(\bar{d},\bar{v})(0) \subseteq \bigcup_{\bar{q}\in\mathscr{H}(\bar{d},\bar{v})} \Big\{ -\sum_{w\in\mathscr{W}} \lambda_w e^w \Big| \sum_{r\in\mathscr{P}^c(\bar{q})} \lambda_r^c e^r - \sum_{r\in\mathscr{P}^o(\bar{q})} \lambda_r^o e^r + \sum_{w\in\mathscr{W}} \lambda_w \Lambda_w^\top = 0 \quad (\lambda_w)_{w\in\mathscr{W}} \in \mathbb{R}^l,\, (\lambda_r^o)_{r\in\mathscr{P}^0(\bar{q})}, (\lambda_r^c)_{r\in\mathscr{P}^c(\bar{q})} \geq 0 \Big\}.$$

Hence, since the coderivative is a homogeneous mapping, it follows from the coderivative criterion (2.30) that condition (7.35) is sufficient for V to be Lipschitz-like around $(\bar{d},\bar{v})$. □

Example 7 (validity of CQ (7.35))**.** *We consider the problem of Example 1 in the paper of Lu [75], with a network of four links* ($n = 4$)*, two O-D pairs* ($l = 2$) *and four routes* ($m = 4$)*. Furthermore, let the O-D-route incidence matrix* (Λ) *and the link-route incidence matrix* (Δ) *be defined as in Lu's Example, then it follows from [75] that for* $\bar{d} = [20,20]^\top$ *and* $\bar{v} = [10,10,10,10]^\top$*, one has* $\bar{v} \in V(\bar{d})$ *and* $\mathscr{H}(\bar{d},\bar{v}) = \{\bar{q} = [10,10,10,10]^\top\}$*. Clearly, all the routes are used. If we set the route capacity vector to be* $[10,30,30,10]^\top$*, meaning that routes 1 and 4 are used at full capacity, while routes 2 and 3 are underused, the left hand side of implication* (7.35) *reduces to*

$$[\lambda_1^c + \lambda_1,\, \lambda_1,\, \lambda_2,\, \lambda_4^c + \lambda_2] = [0,0,0,0], \text{ with } \lambda_1^c, \lambda_4^c \geq 0.$$

This obviously implies that $\lambda_1 = \lambda_2 = 0$*. Thus, CQ* (7.35) *is satisfied at* $(\bar{d},\bar{v})$*.*

Remark 7.2.2 (simplified form of CQ (7.35) when the routes are used but not at full capacity)**.** *For a given demand-link flow couple $(\bar{d},\bar{v})$, if we assume that all routes are used but not at full capacity, that is*

$$\mathscr{P}^c(\bar{q}) = \emptyset \text{ and } \mathscr{P}^o(\bar{q}) = \emptyset, \forall \bar{q} \in \mathscr{H}(\bar{d},\bar{v}),$$

then CQ (7.35) *reduces to the fulfillment of the implication:*

$$\sum_{w\in\mathscr{W}} \lambda_w \Lambda_w^\top = 0 \Longrightarrow [\lambda_w = 0,\, w \in \mathscr{W}].$$

This amounts to saying that the O-D-route incidence matrix has full rank. This assumption is satisfied for Λ in the network of the above example.

In the next result on the sensitivity analysis of the value function in (7.31), the following set of Lagrange multipliers tailored to the corresponding traffic assignment problem is needed:

$$\Lambda(\bar{d},\bar{q}) := \Big\{(\lambda^l,\lambda^c,\lambda^o)\Big|\ \begin{array}{l}\lambda^l = (\lambda_w),\, \lambda^c = (\lambda_r^c) \geq 0,\, \lambda^o = (\lambda_r^o) \geq 0,\\ -\sum_{r\in\mathscr{P}^c(\bar{q})} \lambda_r^c e^r + \sum_{r\in\mathscr{P}^o(\bar{q})} \lambda_r^o e^r - \sum_{w\in\mathscr{W}} \lambda_w \Lambda_w^\top = \Delta^\top \nabla_v f(\bar{d},\bar{v})\end{array}\Big\}. \tag{7.38}$$

Theorem 7.2.3 (sensitivity analysis of the traffic assignment value function parameterized by the demand)**.** *Let $\bar{d} \in D$, then the following assertions are satisfied:*

(i) *An upper bound of the basic subdifferential of φ* (7.31) *at $\bar{d}$ is obtained as*

$$\partial\varphi(\bar{d}) \subseteq \bigcup_{\bar{v}\in S(\bar{d})} \bigcup_{\bar{q}\in\mathscr{H}(\bar{d},\bar{v})} \bigcup_{(\lambda^l,\lambda^c,\lambda^o)\in\Lambda(\bar{d},\bar{q})} \Big\{-\sum_{w\in\mathscr{W}} \lambda_w e^w + \nabla_d f(\bar{d},\bar{v})\Big\}. \tag{7.39}$$

If for all $\bar{v} \in S(\bar{d})$, CQ (7.35) *holds at $(\bar{d},\bar{v})$. Then, φ is Lipschitz continuous around $\bar{d}$.*

(ii) *Assume that* EITHER *the total road cost function f is convex in (d,v)* OR *S is inner semicontinuous at $(\bar{d},\bar{v}) \in$* gph*S, while CQ* (7.35) *also holds at this point. Then φ is Lipschitz continuous around $\bar{d}$ and*

$$\partial\varphi(\bar{d}) \subseteq \bigcup_{\bar{q}\in\mathscr{H}(\bar{d},\bar{v})} \bigcup_{(\lambda^l,\lambda^c,\lambda^o)\in\Lambda(\bar{d},\bar{q})} \Big\{-\sum_{w\in\mathscr{W}} \lambda_w e_w^\top + \nabla_d f(\bar{d},\bar{v})\Big\}. \tag{7.40}$$

Proof. We first start by noting that since J is a continuous function, we have $\Psi(d) \subseteq V(d) := J(Q(d)) \subseteq J(|c|\mathbb{B}), \forall d \in \mathbb{R}^l$, with $J(|c|\mathbb{B})$ bounded. This means that the set-valued mapping Ψ is uniformly bounded, hence inner semicompact, cf. Section 2. Thus (i) follows from the combination of Theorem 4.0.1 (ii) and Theorem 7.2.1. Similarly, (ii) is obtained from Theorem 4.0.1 (i) and Theorem 7.2.1. Under the full convexity of f, note also that gph$V$ is a convex set. Thus, in this case, the basic subdifferential of φ in (7.40) can be derived by proceeding as in the proof of Theorem 3.2.9 (ii). □

As already mentioned above, the *full* convexity assumption on the total road cost function f is automatically satisfied for the expression in (7.29). Thus, we now provide the KKT-type necessary optimality for the O-D matrix adjustment problem only in the case where the latter condition is fulfilled. In other frameworks one can easily proceed similarly while using Theorems 7.2.1 and 7.2.3. Also see the papers [35, 36] for further details.

Theorem 7.2.4 (necessary optimality conditions for the O-D matrix adjustment problem)**.** *Let $(\bar{d},\bar{v})$ be a local optimal solution to problem* (7.30)*, where f is convex in (d,v) and $D := \{d \in \mathbb{R}^l | d \geq 0\}$. Assume that the problem is partially calm at $(\bar{d},\bar{v})$. Then there exist $\lambda > 0$, $\bar{q},\tilde{q} \in \mathbb{R}^m$, $(\lambda^l,\lambda^c,\lambda^o)$ and $(\widetilde{\lambda}^l,\widetilde{\lambda}^c,\widetilde{\lambda}^o)$*

such that the following conditions are satisfied:

$$\nabla_d F(\bar{d},\bar{v}) - \sum_{w\in\mathscr{W}} (\lambda_w - \lambda\widetilde{\lambda}_w)e_w^\top \geq 0,$$

$$\bar{d}^\top \left(\nabla_d F(\bar{d},\bar{v}) - \sum_{w\in\mathscr{W}} (\lambda_w - \lambda\widetilde{\lambda}_w)e_w^\top\right) = 0,$$

$$\Delta^\top \Big(\nabla_v F(\bar{d},\bar{v}) + \lambda\nabla_v f(\bar{d},\bar{v})\Big) = \sum_{r\in\mathscr{P}^o(\bar{q})} \lambda_r^o e^r - \sum_{r\in\mathscr{P}^c(\bar{q})} \lambda_r^c e^r - \sum_{w\in\mathscr{W}} \lambda_w \Lambda_w^\top,$$

$$\Delta^\top \nabla_v f(\bar{d},\bar{v}) = -\sum_{r\in\mathscr{P}^c(\widetilde{q})} \widetilde{\lambda}_r^c e^r + \sum_{r\in\mathscr{P}^o(\widetilde{q})} \widetilde{\lambda}_r^o e^r - \sum_{w\in\mathscr{W}} \widetilde{\lambda}_w \Lambda_w^\top,$$

$$0 \leq \bar{q} \leq c,\, \Lambda\bar{q} = \bar{d},\, \Delta\bar{q} = \bar{v},$$

$$0 \leq \widetilde{q} \leq c,\, \Lambda\widetilde{q} = \bar{d},\, \Delta\widetilde{q} = \widetilde{v},$$

$$\lambda^l = (\lambda_w),\, \lambda^c = (\lambda_r^c) \geq 0,\, \lambda^o = (\lambda_r^o) \geq 0,$$

$$\widetilde{\lambda}^l = (\widetilde{\lambda}_w),\, \widetilde{\lambda}^c = (\widetilde{\lambda}_r^c) \geq 0,\, \widetilde{\lambda}^o = (\widetilde{\lambda}_r^o) \geq 0.$$

Proof. Proceeding as in the proof of Theorem 7.1.3, there exists a number $\lambda > 0$ such that we have the condition

$$0 \in \nabla F(\bar{d},\bar{v}) + \lambda\nabla f(\bar{d},\bar{v}) + \lambda\partial(-\varphi)(\bar{d}) \times \{0\} + N_{\Psi(0)}(\bar{d},\bar{v}), \tag{7.41}$$

where the set-valued mapping Ψ describing a perturbation of the joint upper and lower-level feasible sets is defined by

$$\Psi(d',v') := \{(d,v) \in D\times\mathbb{R}^n |\, (d+d',v+v') \in \operatorname{gph} V\}.$$

Considering the expression of the lower-level feasible set-valued map $V(d) := \Delta Q(d)$, the graph of Ψ is obtained as:

$$\begin{aligned}\operatorname{gph}\Psi &= \{(d,v,d',v')|\exists q : 0 \leq q \leq c,\, d \geq 0,\, \Delta q = v+v',\, \Lambda q = d+d'\}\\ &= \Pi_{1,2,3,4}\{(d,v,d',v',q)|0 \leq q \leq c,\, d \geq 0,\, \Delta q = v+v',\, \Lambda q = d+d'\},\end{aligned}$$

where $\Pi_{1,2,3,4}$ denotes the canonical projection from $\mathbb{R}^l \times \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^n \times \mathbb{R}^m$ to $\mathbb{R}^l \times \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^n$. Clearly, Ψ is a polyhedral set-valued mapping in the sense of Robinson [106]. Thus, Ψ is calm at $(0,\bar{d},\bar{v})$. Thus, by Theorem 2.2.7, we have

$$N_{\Psi(0)}(\bar{d},\bar{v}) \subseteq N_{\operatorname{gph} V}(\bar{d},\bar{v}) + N_{D\times\mathbb{R}^n}(\bar{d},\bar{v}).$$

Inserting this inclusion in (7.41), then the proof is completed by considering the estimate of the subdifferential of φ in Theorem 7.2.3 (ii) and the normal cone to $\operatorname{gph} V$ resulting from the coderivative of V in (7.34). Also note that $N_D(\bar{d}) = \{u|\, u \leq 0,\, \bar{d}^\top u = 0\}$. □

In [11, 12], Fritz-John's type necessary optimality conditions were derived while imposing among other things that the value function φ (7.31) is differentiable. Note that this condition which is obviously very strong, would not change the outcome of our result above, where we obtain KKT-type conditions while imposing only the *partial calmness*, which is a CQ; thus useful for such a purpose. Also recall that in [11, 12], the value of f is the one in (7.29), thus the convexity assumption in Theorem 7.2.4 is automatically satisfied. Finally, observe that similarly to Theorem 7.1.4, the partial calmness assumption is automatical for the O-D matrix adjustment problem in a network with no congestion.

8 Final comments

In this thesis, we have developed two approaches to derive necessary optimality conditions for the classical optimistic bilevel program (P). The first one studied in Chapter 3 is based on techniques from variational analysis which essentially consists of providing constructive representations of the feasible set in terms of operator constraints. Doing so, we have derived various types of stationarity conditions depending on the possible reformulations of the lower-level solution set-valued mapping S (1.2) via the LLVF, KKT and OPEC reformulations. Explicit bounds have also been given for some of the multipliers.

The second approach to derive stationarity conditions for (P) which is mentioned in Section 3.5, see (3.173), has not been extensively discussed. Note that by inserting the corresponding estimate of the coderivative of S (1.2) from Theorem 4.1.1 (see Theorem 5.1.2), Theorem 4.1.6 (see Theorem 5.1.3), Theorem 4.1.8 (see Theorem 5.1.4) and Theorem 5.1.8, one gets the M, C, S, KM and KN-stationarity conditions, respectively. A complete development of this approach is left as exercise to the interested reader.

The coderivatives estimates and robust Lipschitz stability results for the mapping S given in Chapters 4 and 5 have been the backbones of the sensitivity analysis of the two-level value function (5.1), which has led to the necessary optimality conditions of the original optimistic bilevel program in Chapter 5 and to those of the pessimistic bilevel program in Chapter 6. Therefore, one of the most important conclusions of this thesis is that the theory on necessary optimality conditions for the bilevel optimization problem can essentially reduce to the sensitivity analysis of the solution set-valued mapping of the lower-level problem. For the optimistic models (P) and (P_o) an upper estimation of the coderivative of S may be sufficient, cf. Remark 5.2.5. However, for the pessimistic case we may need the Lipschitz-like property for S in the case where we wish to obtain lower sudifferential optimality conditions. In any case, the gap between the KN and KM-stationarity conditions and the M, C and S-stationarity is solely explained by the structure chosen for the mapping S.

In these concluding comments, it is important to mention the result in Theorem 5.3.2 stating the equivalence between the stationarity conditions of original (P_o) and classical (P) optimistic bilevel programs under a particular setting. Also note the interesting discussions in Subsection 6.2.4 on how to derive necessary optimality conditions of the pessimistic problem from those of an appropriate problem of the optimistic-type.

In order to have a clear picture on the links between the stationarity conditions of the bilevel program (1.1) via its optimistic reformulations ((P) and (P_o)) and the pessimistic reformulation (P_p), we have considered only the setting where all the functions involved are continuously differentiable. In closely related frameworks, stationarity conditions of these problems are investigated in [26, 28, 132] in the case of locally Lipschitz continuous functions while also considering the very general operator constraints. In the latter papers, the LLVF reformulations are discussed. The nonsmooth KKT reformulation will be considered in a future research.

To complete the work started in this thesis, the following points will also be considered in future research: (1) Provide sufficient conditions ensuring the inner semicontinuity of the two-level solution set-valued mappings S_o (5.2) which are different from the single-valuedness. As for the lower-level mapping S (1.2), recall that the dual conditions from Chapters 4 and 5 ensuring robust Lipschitz stability are also sufficient conditions for the inner semicontinuity of this mapping which do not necessary induce that it is single-valued. (2) Provide characterizations of the partial calmness in bilevel transportation problems of Chapter 7 in terms of problem data.

Bibliography

[1] T. Abrahamsson. *Estimation of origin-destination matrices using traffic counts: Aliterature survey*. Interim Report IR-98-021/May. International Institue for Applied Systems Analysis, Laxenburg, Austria (1998).

[2] H. Babahadda and N. Gadhi. *Necessary optimality conditions for bilevel optimization problems using convexificators*. J. Glob. Optim. **34** (2006), no. 4, 535–549.

[3] L. Ban, B.S. Mordukhovich and W. Song. *Lipschitzian stability of parametric variational inequalities over generalized polyhedra in Banach spaces*. Nonlinear Anal. **74** (2011), no. 2, 441–461.

[4] T.Q. Bao, P. Gupta and B.S. Mordukhovich. *Necessary conditions in multiobjective optimization with equilibrium constraints*. J. Optim. Theory Appl. **135** (2007), no.2, 179–203.

[5] J.F. Bard. *Practical bilevel optimization: Algorithms and applications*. Dordrecht: Kluwer Academic Publishers, 1998.

[6] M. Beckmann, C.B. Mcguire and C.B. Winsten. *Studies in the economics of transportation*. New Haven: Yale University Press, 1956.

[7] H. Bonnel. *Optimality conditions for the semivectorial bilevel optimization problem*. Pac. J. Optim. **2** (2006), no. 3, 447–467.

[8] L. Brotcorne, M. Labbée, P. Marcotte and G. Savard. *A bilevel model and algorithm for a freight tariff-setting problem.* Transp. Sci. **34** (2000), no. 3, 289-302.

[9] J.V. Burke and M.C. Ferris. *Weak sharp minima in mathematical programming*. SIAM J. Control Optim. **31** (1993), no. 5, 1340–1359.

[10] M. Cervinka, M. Matonoha and J.V. Outrata. *On the computation of relaxed pessimistic solutions to MPECs*. Optim. Math. Soft. (to appear). doi: 10.1080/10556788.2011.627585.

[11] Y. Chen. *Bilevel programming problems: Analysis, algorithms and applications*. Ph.D. thesis. Centre de Recherche sur les Transports, Université de Montréal, CRT-984, 1994.

[12] Y. Chen and M. Florian. *Congested O-D trip demand adjustement problem: Bilevel programming formulation and optimality conditions*. A. Migdalas et al. (eds), Multilevel Optimization: Algorithms and Applications. Dordrecht: Kluwer Academic Publishers, 1998, pp. 1–22.

[13] Y. Chen and M. Florian. *The nonlinear bilevel programming problem: Formulations, regularity and optimality conditions*. Optimization **32** (1995), no. 3, 193–209.

[14] S.-W. Chiou. *Bilevel programming for the continuous transport network design problem*. Tansp. Res. B **39** (2005), 361–383.

[15] P.A. Clark and A.W. Westerberg. *A note on the optimality conditions for the bilevel programming problem*. Nav. Res. Logist. **35** (1988), no. 5, 413–418.

[16] F.H. Clarke. *Optimization and nonsmooth analysis*. New York: Wiley, 1994.

[17] E. Codina and L. Montero. *Approximation of the steepest descent direction for the O-D matrix adjustment problem.* Ann. Oper. Res. **144** (2006), 329–362.

[18] B. Colson, P. Marcotte and G. Savard. *An overview of bilevel optimization*. Ann. Oper. Res. **153** (2007), 235–256.

[19] S. Dassanayaka. *Methods of variational analysis in pessimistic bilevel progamming*. Ph.D. thesis. Department of Mathematics, Wayne State University, 2010.

[20] S. Dempe. *A necessary and a sufficient optimality condition for bilevel programming problems*. Optimization **25** (1992), no. 4, 341–354.

[21] S. Dempe. *Foundations of bilevel programming*. Dordrecht: Kluwer Academic Publishers, 2002.

[22] S. Dempe. *Annotated bibliography on bilevel programming and mathematical programs with equilibrium constraints*. Optimization **52** (2003), no. 3, 333–359.

[23] S. Dempe, N. Dinh and J. Dutta. *Optimality conditions for a simple convex bilevel programming problem*. R.S. Burachik and J.-C. Yao (eds), Variational analysis and generalized differentiation in optimization and control. In honor of B.S. Mordukhovich. Berlin: Springer, 2010, pp. 149–161.

[24] S. Dempe and J. Dutta. *Is bilevel programming a special case of mathematical programming with equilibrium constraints?* Math. Program. **131** (2012), no. 1-2, 37–48.

[25] S. Dempe, J. Dutta and S. Lohse. *Optimality conditions for bilevel programming problems*. Optimization **55** (2006), no. 5-6, 505–524.

[26] S. Dempe, J. Dutta and B.S. Mordukhovich. *New necessary optimality conditions in optimistic bilevel programming*. Optimization **56** (2007), no. 5-6, 577–604.

[27] S. Dempe and N. Gadhi. *Necessary optimality conditions for bilevel set optimization problems*. J. Glob. Optim. **39** (2007), no. 4, 529–542.

[28] S. Dempe, N. Gadhi and A.B. Zemkoho. *New optimality conditions in semivectorial bilevel optimization.* Preprint 2011-06. Fakultät für Mathematik und Informatik, TU Bergakademie Freiberg, 2011.

[29] S. Dempe, V.V. Kalashnikov and N. Kalashnykova. *Optimality conditions for bilevel programming problems*. S. Dempe and V.V. Kalashnikov (eds.), Optimization with multivalued mappings. New York: Springer, 2006, pp. 3-28.

[30] S. Dempe, B.S. Mordukhovich and A.B. Zemkoho. *Necessary optimality conditions in pessimistic bilevel programming.* Optimization, to appear.

[31] S. Dempe, B.S. Mordukhovich and A.B. Zemkoho. *Sensitivity analysis for two-level value functions with applications to bilevel programming.* Preprint 2011-07. Fakultät für Mathematik und Informatik, TU Bergakademie Freiberg, 2011.

[32] S. Dempe and S. Vogel. *The generalized Jacobian of the optimal solution in parametric optimization*. Optimization **50** (2001), no.5-6, 387–405.

[33] S. Dempe and A.B. Zemkoho. *The generalized Mangasarian-Fromowitz constraint qualification and optimality conditions for bilevel programs.* J. Optim. Theory Appl. **148** (2011), no. 1, 46–68.

[34] S. Dempe and A.B. Zemkoho. *On the Karush-Kuhn-Tucker reformulation of the bilevel optimization problem*. Nonlinear Anal. **75** (2012), no. 3, 1202–1218.

[35] S. Dempe and A.B. Zemkoho. *Bilevel programming: reformulations, constraint qualifications and optimality conditions*. Math. Program. (2012). doi: 10.1007/s10107-011-0508-5.

[36] S. Dempe and A.B. Zemkoho. *Bilevel road pricing: theoretical analysis and optimality conditions*. Ann. Oper. Res. (2011). doi: 10.1007/s10479-011-1023-z.

[37] S. Dempe and A.B. Zemkoho. *A bilevel approach for traffic management in capacitated networks.* Preprint 2008-05. Fakultät für Mathematik und Informatik, TU Bergakademie Freiberg, 2008.

[38] S. Dewez, M. Labbé, P. Marcotte and G. Savard. *New formulations and valid inequalities for a bilevel pricing problem.* Oper. Res. Lett. **36** (2008), 141–149.

[39] A. Dhara and A. Mehra. *Approximate optimality conditions for minimax programming problems.* Optimization **59** (2010), no. 7-8, 1013–1032.

[40] N. Dinh, B.S. Mordukhovich and T.T.A. Nghia. *Subdifferentials of value functions and optimality conditions for DC and bilevel infinite and semi-infinite programs.* Math. Program. **123** (2010), no. 1, 101–138.

[41] A.L. Dontchev and R.T. Rockafellar. *Ample parametrization of variational inclusions.* SIAM J. Optim. **12** (2001), 170–187.

[42] M. Durea, J. Dutta and Chr. Tammer. *Bounded sets of Lagrange multipliers for vector optimization problems in infinite dimension.* J. Math. Anal. Appl. **348** (2008), no. 2, 589–606.

[43] J. Dutta. *Generalized derivatives and nonsmooth optimization, a finite dimensional tour* (with comments and rejoinder). Top **13** (2005), no. 2, 185–314.

[44] J. Dutta and S. Dempe. *Bilevel programming with convex lower level problems.* S. Dempe and V. Kalashnikov (eds.), Optimization with multivalued mappings. New York: Springer, 2006, pp. 51-71.

[45] M.C. Ferris. *Weak sharp minima and penalty functions in mathematical programming.* Ph.D. thesis. Churchill College, University of Cambridge, 1989.

[46] A.V. Fiacco. *Introduction to sensitivity and stability analysis in nonlinear programming.* New York-London: Academic Press, 1983.

[47] C.S. Fisk. *On combining maximum maximum entropy trip matrix with user optimal assigment.* Transp. Res. B **22** (1988), no. 1, 69–73.

[48] M. Flegel. *Constraint qualification and stationarity concepts for mathematical programs with equilibrium constraints.* Ph.D. thesis. Institute of Applied Mathematics and Statistics, University of Würzburg, 2005.

[49] M.L. Flegel and C. Kanzow. *On the Guignard constraint qualification for mathematical programs with equilibrium constraints.* Optimization **54** (2005), no. 6, 517–534.

[50] M.L. Flegel and C. Kanzow. *A direct proof for M-stationarity under MPEC-GCQ for mathematical programs with equilibrium constraints.* S. Dempe and V.V. Kalashnikov (eds), Optimization with multivalued mappings. New York: Springer, 2006, pp. 111–122.

[51] M.L. Flegel, C. Kanzow and J.V. Outrata. *Optimality conditions for disjunctive programs with application to mathematical programs with equilibrium constraints.* J. Set-Valued Anal. **15** (2007), no. 2, 139–162.

[52] C.A. Floudas and P.M. Pardalos (eds). *Encyclopedia of Optimization.* Dordrecht-London-Boston: Kluwer Academic Publishers, 2001.

[53] G. Heilporn, M. Labbé, P. Marcotte and G. Savard. *A polyhedral study of the network pricing problem with connected toll arcs.* Networks **55** (2010), 234–246.

[54] J. Gauvin. *A necessary and sufficient regularity condition to have bounded multipliers in nonconvex programming.* Math. Program. **12** (1977), 136–138.

[55] J. Gauvin and F. Dubeau. *Differential properties of the marginal function in mathematical programming.* Math. Program. Study **18** (1982), 101–119.

[56] S.-P. Han and O.L. Mangasarian. *Exact penalty functions in nonlinear programming.* Math. Program. **17** (1979), 251–269.

[57] D. Han and H. Yang. *The multi-class, multicriterion traffic equilibrium and efficiency of congestion pricing.* Transp. Res. E **44** (2008), no. 5, 753–773.

[58] R. Henrion. *On constraint qualifications.* J. Optim. Theory Appl. **72** (1992), no. 1, 187–197.

[59] R. Henrion, A. Jourani and J.V. Outrata. *On the calmness of a class of multifunctions.* SIAM J. Optim. **13** (2002), no. 2, 603–618.

[60] R. Henrion, B.S. Mordukhovich and N.M. Nam. *Second-order analysis of polyhedral systems in finite and infinite dimensions with applications to robust stability of variational inequalities.* SIAM J. Optim. **20** (2010), no. 5, 2199–2227.

[61] R. Henrion and J.V. Outrata. *A subdifferential condition for calmness of multifunctions.* J. Math. Anal. Appl. **258** (2001), no. 1, 110–130.

[62] R. Henrion and J.V. Outrata. *Calmness of constraint systems with applications.* Math. Program. **104** (2005), no. 2-3, 437-464.

[63] R. Henrion, J. Outrata, and T. Surowiec. *On the co-derivative of normal cone mappings to inequality systems.* Nonlinear Anal. **71** (2009), no. 3-4, 1213–1226.

[64] R. Henrion, J.V. Outrata and T.M. Surowiec. *A note on the relation between strong and M-stationarity for a class of mathematical programs with equilibrium constraints.* Kybernetika **46** (2010), no. 3, 423–434.

[65] R. Henrion and T.M. Surowiec. *On calmness conditions in convex bilevel programming.* Appl. Anal. **90** (2011), no. 6, 951 – 970.

[66] T. Hoheisel, C. Kanzow and J.V. Outrata. *Exact penalty results for mathematical programs with vanishing constraints.* Nonlinear Anal. **72** (2010), 2514–2526.

[67] X. Hu and D. Ralph. *A note on sensitivity of value functions of mathematical programs with complementarity constraints.* Math. Program. **93** (2002), 265–279.

[68] Y. Ishizuka. *Farkas' theorem of nonconvex type and its application to a min-max problem.* J. Optim. Theory Appl. **57** (1988), no. 2, 341–354.

[69] M. Josefsson and M. Patriksson. *Sensivity analysis of separable traffic equilibrium equilibria with application to bilevel optimization in network design.* Transp. Res. B **41** (2007), no. 1 4-31.

[70] A. Jourani. *Constraint qualifications and Lagrange multipliers in nondifferentiable programming problems.* J. Optim. Theory Appl. **81** (1994), no. 3, 533–548.

[71] M. Kojima. *Strongly stable stationary solutions in nonlinear programs.* Analysis and computation of fixed points, Proc. Symp., Univ. Wis. 1979, 93-138.

[72] M. Labbé, P. Marcotte and G. Savard. *A bilevel model of taxation and its application to optimal highway pricing.* Management Sci. **44** (1998), no. 12, 1608–1622.

[73] S. Lohse. *Eine spezielle klasse von zwei-ebenen-optimierungsaufgaben.* Ph.D. thesis. Fakultät für Mathematik und Informatik, TU Bergakademie Freiberg, 2010.

[74] P. Loridan and J. Morgan. *Weak via strong Stackelberg problem: New results.* J. Glob. Optim. **8** (1996), no. 3, 263–287.

[75] S. Lu. *Sensitivity of static traffic user quilibria with pertubations in arc cost function and travel demand.* Transp. Sci. **42** (2008), no. 1, 105–123.

[76] Y. Lucet and J.J. Ye. *Sensitivity analysis of the value function for optimization problems with variational inequality constraints.* SIAM J. Control Optim. **40** (2001), no. 3, 699–723. (Erratum in SIAM J. Control Optim. **41** (2002), no. 4, 1315-1319).

[77] Z.Q. Luo, J.-S. Pang and D. Ralph. *Mathematical programs with equilibrium constraints.* Cambridge University Press, 1996.

[78] O.L. Mangasarian and T.-H. Shiau. *Lipschitz continuity of solutions of linear inequalities, programs and complementarity problems.* SIAM J. Control Optim. **25** (1987), 583–595.

[79] Q. Meng, H. Yang and M.G.H Bell. *An equivalent continuously differentiable model and a locally convergent algorithm for the continuous network design problem.* Transp. Res. B **35** (2001), 83–105.

[80] A. Migdalas. *Bilevel programming in traffic planning: Models, methods and challenge*. J. Glob. Optim. **7** (1995), 381–405.

[81] J.A. Mirrlees. *The theory of moral hazard and unobservable behaviour. I.* Rev. Econ. Stud. **66** (1999), no. 1, 3–21.

[82] B.S. Mordukhovich. *Lipschitzian stability of constraint systems and generalized equations.* Nonlinear Anal. **22** (1994), no. 2, 173–206.

[83] B.S. Mordukhovich. *Stability theory for parametric generalized equations and variational inequalities via nonsmooth analysis.* Trans. Am. Math. Soc. **343** (1994), no. 2, 609–657.

[84] B.S. Mordukhovich. *Necessary conditions in nonsmooth minimization via lower and upper subgradients.* Set-Valued Anal. **12** (2004), 163-193.

[85] B.S. Mordukhovich. *Maximum principle in the problem of time optimal response with nonsmooth constraints.* J. Appl. Math. Mech. **40** (1976), 690–669.

[86] B.S. Mordukhovich. *Nonsmooth analysis with nonconvex generalized differentials and conjugate mappings.* Dokl. Akad. Nauk BSSR **28** (1984), 976–979.

[87] B.S. Mordukhovich. *Generalized differential calculus for nonsmooth and set-valued mappings.* J. Math. Anal. Appl. **183** (1994), no. 1, 250–288.

[88] B.S. Mordukhovich. *Metric approximations and necessary optimality conditions for general classes of nonsmooth extremal problems.* Sov. Math. Dokl. **22** (1980), 526–530.

[89] B.S. Mordukhovich. *Variational analysis and generalized differentiation. I: Basic theory. II: Applications.* Berlin: Springer, 2006.

[90] B.S. Mordukhovich and N.M. Nam. *Variational stability and marginal functions via generalized differentiation.* Math. Oper. Res. **30** (2005), no. 4, 800–816.

[91] B.S. Morukhovich, M.N. Nam and H.M. Phan. *Variational analysis of marginal function with applications to bilevel programming problems.* J. Optim. Theory Appl. **152** (2012), no. 3, 557–586.

[92] B.S. Morukhovich, M.N. Nam and N.D. Yen. *Subgradients of marginal functions in parametric mathematical programming.* Math. Program. **116** (2009), no. 1-2, 369–396.

[93] B.S. Mordukhovich and J.V. Outrata. *Coderivative analysis of quasi-variational inequalities with applications to stability and optimization.* SIAM J. Optim. **18** (2007), no. 2, 389–412.

[94] M. Netter. *Affectations de traffic et tarification au coût marginal social: critiques de quelques idées admises.* Transp. Res. **6** (1972), 411–429.

[95] V.H. Nguyen, J.-J. Strodiot and R. Mifflin. *On conditions to have bounded multipliers in locally Lipschitz programming.* Math. Program. **18** (1980), 100–106.

[96] Y. Noriega and M. Florian. *Some enhancements of the gradient method for the O-D matrix adjustement.* CIRRELT-2009-04 (2009).

[97] J.V. Outrata. *Optimality conditions for a class of mathematical programs with equilibrium constraints.* Math. Oper. Res. **24** (1999), no. 3, 627–644.

[98] J.V. Outrata. *A note on the usage of nondifferentiable exact penalties in some special optimization problems.* Kybernetika **24** (1988), no. 4, 251–258.

[99] J.V. Outrata. *Necessary optimality conditions for Stackelberg problems.* J. Optim. Theory Appl. **76** (1993), no. 2, 305–320.

[100] J.V. Outrata. *A generalized mathematical program with equilibrium constraints.* SIAM J. Control Optim. **38** (2000), no. 5, 1623–1638.

[101] J.V. Outrata, M. Kocvara and J. Zowe,.*Nonsmooth approach to optimization problems with equilibrium constraints.* Dordrecht: Kluwer Academic Publishers, 1998.

[102] M. Patriksson. *The traffic assignment problem - Models and methods, Topics in Transportation.* Utrecht: VSP BV, 1994.

[103] M. Pilecka. *Eine spezielle Transformation der Zwei-Ebenen-Optimierungsaufgabe.* M.Sc thesis. Fakultät für Mathematik und Informatik, TU Bergakademie Freiberg, 2011.

[104] J.-Ch. Pomerol. *The boundedness of the Lagrange multipliers set and duality in mathematical programming.* Z. Oper. Res., Ser. A **25** (1981), 191–204.

[105] D. Ralph and S. Dempe. *Directional derivatives of the solution of a parametric nonlinear program.* Math. Program. **70** (1995), no. 2, 159–172.

[106] S.M. Robinson. *Some continuity properties of polyhedral multifunctions.* Math. Program. Study **14** (1981), 206–214.

[107] S.M. Robinson. *Generalized equations and their solutions. I: Basic theory.* Math. Program. Study **10** (1979), 128–141.

[108] R.T. Rockafellar. *Lagrange multipliers and optimality.* SIAM Rev. **35** (1993), no. 2, 183–238.

[109] R.T. Rockafellar and R.J.-B. Wets. *Variational analysis.* Berlin: Springer, 1998.

[110] H. Scheel and S. Scholtes. *Mathematical programs with complementarity constraints: stationarity, optimality, and sensitivity.* Math. Oper. Res. **25** (2000), no. 1, 1–22.

[111] W. Schirotzek. *Nonsmooth analysis.* Berlin: Springer, 2007.

[112] W.E. Schmitendorf. *A simple derivation of necessary conditions for static minmax problems.* J. Math. Anal. Appl. **70** (1979), 486–489.

[113] K. Shimizu, Y. Ishizuka and J.F. Bard. *Nondifferentiable and two-level mathematical programming.* Dordrecht: Kluwer Academic Publishers, 1997.

[114] T.M. Surowiec. *Explicit stationary conditions and solution characterization for equilibrium problems with equilibrium constraints.* Ph.D. thesis. Mathematisch-Naturwissentschaften Fakultät II, Humbolt Universität zu Berlin, 2010.

[115] L.N. Vicente and P.H. Calamai. *Bilevel and multilevel programming: A bibliography review.* J. Glob. Optim. **5** (1994), no. 3, 291–306.

[116] J.G. Wardrop. *Some theoretical aspects of road traffic research.* Proceedings of the Institute of Civil Engeneers Part II (1952), 325–378.

[117] H. Yang and W.H.K. Lam. *Optimal road tolls under conditions of queueing and congestion.* Tansp. Res. A **30** (1996), no. 5, 319–332.

[118] H. Yang, T. Sasaki, Y. Iida and Y. Asakura. *Estimation of origin-destination matrices from traffic counts on congested networks.* Tansp. Res. B **26** (1992), no. 6, 417–434.

[119] H. Yang and S. Yagar. *Traffic assignment and traffic control in general freeway-arterial corridor systems.* Tansp. Res. B **28** (1994), no. 6, 463–486.

[120] J.J. Ye. *Constraint qualifications and KKT conditions for bilevel programming problems.* Math. Oper. Res. **31** (2006), no. 4, 811–824.

[121] J.J. Ye. *Necessary optimality conditions for multiobjective bilevel programs.* Math. Oper. Res. **36** (2011), no. 1, 165–184.

[122] J.J. Ye. *New uniform parametric error bounds.* J. Optim. Theory Appl. **98** (1998), no. 1, 197–219.

[123] J.J. Ye. *Optimality conditions for optimization problems with complementarity constraints.* SIAM J. Optim. **9** (1999), no. 2, 374–387.

[124] J.J. Ye. *Constraint qualifications and necessary optimality conditions for optimization problems with variational inequality constraints.* SIAM J. Optim. **10** (2000), no. 4, 943–962.

[125] J.J. Ye. *Nondifferentiable multiplier rules for optimization and bilevel optimization problems.* SIAM J. Optim. **15** (2004), no. 1, 252–274.

[126] J.J. Ye. *Necessary and sufficient optimality conditions for mathematical programs with equilibrium constraints.* J. Math. Anal. Appl. **307** (2005), no. 1, 350–369.

[127] J.J. Ye and X.Y. Ye. *Necessary optimality conditions for optimization problems with variational inequality constraints.* Math. Oper. Res. **22** (1997), no. 4, 977–997.

[128] J.J. Ye and D.L. Zhu. *Optimality conditions for bilevel programming problems.* Optimization **33** (1995), no. 1, 9–27 (Erratum in *Optimization* **39** (1997), no. 4, 361-366).

[129] 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** (2010), no. 4, 1885–1905.

[130] 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** (1997), no. 2, 481–507.

[131] A.B. Zemkoho. *A simple approach to optimality conditions in minmax programming.* Optimization (2012). doi: 10.1080/02331934.2011.653788.

[132] A.B. Zemkoho. *Optimization problems with value functions objectives*. www.optimization-online.org.

[133] R. Zhang. *Problems of hierarchical optimization in finite dimensions.* SIAM J. Optim. **4** (1994), 521–536.

[134] X.Y. Zheng and K.F. Ng. *Calmness for L-subsmooth multifunctions in Banach spaces.* SIAM J. Optim. **19** (2009), no. 4, 1648–1673.