Optimization

A Journal of Mathematical Programming and Operations Research

ISSN: 0233-1934 (Print) 1029-4945 (Online) Journal homepage: https://www.tandfonline.com/loi/gopt20

New necessary optimality conditions in optimistic bilevel programming

S. Dempe , J. Dutta & B. S. Mordukhovich

Optimization,
Vol. 56, Nos. 5–6, October–December 2007, 577–604

Taylor & Francis
Taylor & Francis Group

New necessary optimality conditions in optimistic bilevel programming¶

S. DEMPE†, J. DUTTA‡ and B. S. MORDUKHOVICH*§

†Department of Mathematics and Computer Science, Technical University Bergakademie Freiberg, Freiberg, Germany
‡Department of Mathematics, Indian Institute of Technology, Kanpur, India
§Department of Mathematics, Wayne State University, Detroit, USA

(*Received 18 December 2006; in final form 21 May 2007*)

The article is devoted to the study of the so-called optimistic version of bilevel programming in finite-dimensional spaces. Problems of this type are intrinsically nonsmooth (even for smooth initial data) and can be treated by using appropriate tools of modern variational analysis and generalized differentiation. Considering a basic optimistic model in bilevel programming, we reduce it to a one-level framework of nondifferentiable programs formulated via (nonsmooth) optimal value function of the parametric lower-level problem in the original model. Using advanced formulas for computing basic subgradients of value/marginal functions in variational analysis, we derive new necessary optimality conditions for bilevel programs reflecting significant phenomena that have never been observed earlier. In particular, our optimality conditions for bilevel programs do not depend on the partial derivatives with respect to parameters of the smooth objective function in the parametric lower-level problem. We present efficient implementations of our approach and results obtained for bilevel programs with differentiable, convex, linear, and Lipschitzian functions describing the initial data of the lower-level and upper-level problems.

Keywords: Bilevel programming; Value functions; Variational analysis; Generalized differentiation; Necessary optimality conditions

Mathematics Subject Classifications 2000: Primary: 90C29; 49J52; Secondary: 49J53

1. Introduction and overview

Bilevel programming deals with a broad class of problems in *hierarchical optimization* that consist of minimizing some (*upper-level*) objective function $F: \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ subject to upper-level constraints $x \in X \subset \mathbb{R}^n$ and $(x, y) \in \text{gph } \Psi \subset \mathbb{R}^m \times \mathbb{R}^n$. Here, gph Ψ stands

*Corresponding author. Email: boris@math.wayne.edu.
¶This work is dedicated to the memory of Prof. Dr Alexander Moiseevich Rubinov.

Optimization
ISSN 0233-1934 print/ISSN 1029-4945 online © 2007 Taylor & Francis
http://www.tandf.co.uk/journals
DOI: 10.1080/02331930701617551

for the graph of the *solution/argminimum* set-valued mapping to another (*lower-level*) parametric optimization problem given by

$$\Psi(x) := \operatorname*{Argmin}_{y}\{f(x, y) \mid g(x, y) \leq 0\} \tag{1.1}$$

with the cost function $f: \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ and the constraint vector function $g: \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^p$.

More precisely, minimization at the upper level is done with respect to the upper-level variable x only, and $y = y(x) \in \Psi(x)$ is a reaction of the lower-level decision maker to the upper-level selection. If this reaction is not uniquely determined (and, hence, the upper-level decision maker is not able to predict it in advance of the lower-level selection), the upper-level objective function value is not well determined with the choice of $x \in X$. This implies that the bilevel programming problem is generally *not well-defined*. To overcome this obstacle, two different solution concepts are introduced. These are the *optimistic* solution and the *pessimistic* solution concepts.

In the optimistic framework we consider the following problem:

$$\begin{aligned} &\text{minimize } \varphi_0(x) \text{ subject to } x \in X \\ &\text{with } \varphi_0(x) := \inf\{F(x, y) \mid y \in \Psi(x)\}. \end{aligned} \tag{1.2}$$

A point $\overline{x} \in X$ is called a *local optimistic solution* of the bilevel programming problem if $\varphi_0(x) \leq \varphi_0(\overline{x})$ for $x \in X$ sufficiently close to $\overline{x}$. Local optimistic solutions are of our main interest in this article, while *global* optimistic solutions can be defined in a similar way.

In the pessimistic formulation we construct the function

$$\varphi_p(x) := \sup\{F(x, y) \mid y \in \Psi(x)\}$$

and then solve the problem of minimizing $\varphi_p(x)$ on x. Pessimistic optimal solutions can now be defined similarly to the optimistic case with the replacement of φ_0 by φ_p in (1.2).

In this article we concentrate on the *optimistic formulation* of the bilevel programming problem and intend to establish efficient necessary optimality conditions for local optimistic solutions. Note that in the most interesting situations, from the viewpoints of both the theory and applications, the parametric sets of solutions $\Psi(x)$ to the lower-level problem in (1.1) are compact in finite dimensions, and thus the infimum in the construction of the upper-level value function φ_0 in (1.2) is realized.

To derive necessary optimality conditions for optimistic bilevel programming problems of type (1.2), several approaches are possible and have been already investigated. Let us mention and partly discuss the following *three* approaches to the study of optimistic bilevel programs in form (1.2).

The *first idea* is to replace the lower-level problem (1.1) by its *Karush–Kuhn–Tucker* (*KKT*) *conditions* (where the symbol $\top$ stands as usual for transposition):

$$\nabla_y f(x, y) + \lambda^\top \nabla_y g(x, y) = 0, \; g(x, y) \leq 0, \quad \lambda \geq 0, \; \lambda^\top g(x, y) = 0$$

arising from necessary optimality conditions for the parametric lower-level problems. The resulting upper-level problem belongs to the class of the so-called *mathematical programs with equilibrium constraints* (*MPECs*):

$$\begin{cases} \text{minimize } \; F(x, y) \; \text{ subject to } \; (x, y, \lambda) \in X \times \mathbb{R}^m \times \mathbb{R}^p, \\ \nabla_y f(x, y) + \lambda^\top \nabla_y g(x, y) = 0, \; g(x, y) \leq 0, \; \lambda \geq 0, \; \lambda^\top g(x, y) = 0; \end{cases} \tag{1.3}$$

see, e.g., [14,26] and the references therein.

This approach faces serious difficulties in the case when problem (1.1) is not convex, since then a point can be detected as a candidate for a local minimum (i.e., a stationary point), which is not even feasible to the original bilevel program. Considering the case when the lower-level problem in bilevel programming is smooth and convex, Dempe [6] developed this approach under the assumption that the parameterized solution set $\Psi(x)$ is a *singleton* and that $y(x)$ given by $\Psi(x)=\{y(x)\}$ is a PC^1-function, i.e., a continuous selection from finitely many smooth functions:

$$y(x) \in \left\{y^1(x), \dots, y^q(x)\right\} \tag{1.4}$$

whenever x belongs to a neighborhood of a local minimizer $\bar{x}$ to the bilevel program under consideration. In (1.4), the natural number $q>1$ corresponds to violating the linear independence constraint qualification and/or the strict complementarity slackness assumption for problem (1.1) at the point $\left(\overline{x}, y(\overline{x})\right)$. If the linear independence constraint qualification is violated, the resulting necessary optimality conditions for the MPEC may be very weak if applied to the corresponding bilevel program. Indeed, it is proved under the assumptions in Dempe [6] that each of the functions $y^i(x)$ in (1.4) corresponds to one of the vertices λ^i of the Lagrange multiplier set $\Lambda\left(\overline{x}, y(\overline{x})\right)$ in the lower-level problem defined by

$$\Lambda(x, y) := \left\{\lambda \geq 0 \mid \nabla_y f(x, y) + \lambda^\top \nabla_y g(x, y) = 0,\ \lambda^\top g(x, y) = 0\right\}.$$

Furthermore, the following condition is necessary for optimality: one has

$$\begin{aligned}&\nabla_x f(\overline{x}, y(\overline{x}))r + \nabla_y f(\overline{x}, y(\overline{x}))y'(x; r)\\ &\quad = \nabla_x f(\overline{x}, y(\overline{x}))r + \nabla_y f(\overline{x}, y(\overline{x}))\nabla y^i(\overline{x})r \geq 0\end{aligned} \tag{1.5}$$

whenever $i \in I_y^e(\overline{x})$ and $r \in T_{\{x \mid y(x)=y^i(x)\}}(\overline{x}) \cap T_X(\overline{x})$, where $T_W(z)$ denotes the Bouligand contingent cone to the set W at $z \in W$, and where $I_y^e(\overline{x})$ is the set of essentially active selection functions $y^i(\cdot)$ at $\overline{x}$; see, e.g. Scholtes [30].

On the other hand, assume that $\left(\overline{x}, \overline{y}, \overline{\lambda}\right)$ is a *stationary point* for the MPEC in (1.3). Then there exists an open neighborhood Z of $(\overline{x}, \overline{y}, \overline{\lambda})$ such that $f(x, y) \geq f(\overline{x}, \overline{y})$ for all $(x, y, \lambda) \in Z$ feasible to problem (1.3). Restricting λ to a neighborhood of $\overline{\lambda}$, we thus restrict the choice of $y(x)$ in (1.4) to a proper subset of $\{y^1(x), \dots, y^q(x)\}$. By using then standard differential calculus, we can formulate a necessary optimality condition for the MPEC as

$$\nabla_x f(\overline{x}, y(\overline{x}))r + \nabla_y f(\overline{x}, y(\overline{x}))\nabla y^i(\overline{x})r \geq 0$$

held for *some* $i \in I_y^e(\overline{x})$ and for all $r \in T_{\{x \mid y(x)=y^i(x)\}}(\overline{x}) \cap T_X(\overline{x})$. In this way, there may exist another $\widehat{i} \in I_y^e(\overline{x})$ – corresponding to another multiplier $\widehat{\lambda} \in \Lambda(\overline{x}, y(\overline{x}))$ – such that

$$\nabla_x f(\overline{x}, y(\overline{x}))r + \nabla_y f(\overline{x}, y(\overline{x}))\nabla y^{\hat{i}}(\overline{x})r < 0$$

for some $r \in T_{\{x \mid y(x)=y^{\hat{i}}(x)\}}(\overline{x}) \cap T_X(\overline{x})$. Thus, the necessary optimality condition (1.5) for MPEC (1.3) is satisfied while there is a *direction of descent* for the bilevel program under consideration, which causes the MPEC model (1.3) and the necessary optimality condition (1.5) to be largely *unreliable* for the study of the original optimistic bilevel program (1.2).

Another well-recognized manifestation of this phenomena is the *violation* of the classical Mangasarian–Fromovitz constraint qualification for the MPEC problem (1.3) considered as a nonlinear program with one of its (equality) constraints defined by the *complementarity* condition $\lambda^\top g(x, y) = 0$. Generally speaking, the main difficulty in the MPEC approach to bilevel programming is an appropriate introduction of new variables λ while reducing a bilevel program to the corresponding MPEC that allows an adequate treatment.

The *second idea* involves the application of the *normal cone* introduced by Mordukhovich [16] to the graph of the argminimum map Ψ defined in (1.1), i.e., computing in fact the *coderivative* of this set-valued mapping. This idea has been implemented in the framework of bilevel programming in the papers by Zhang [39] and by Dutta and Dempe [10]. We also refer the reader to [3,17,19,24,36] and the bibliographies therein for related developments in this direction in more general frameworks of MPEC problems and the like.

The *third idea* initiated by Outrata [23] in the framework of bilevel programming/Stackelberg games (cf. also [31,37]) is to consider the *optimal value function*

$$\varphi(x) := \inf_y \{f(x, y) \mid g(x, y) \le 0\} \tag{1.6}$$

in the lower-level problem (1.1) and to reformulate the optimistic bilevel program (1.2) as

$$\begin{cases} \text{minimize} \;\; F(x, y) \;\; \text{subject to} \;\; (x, y) \in X \times \mathbb{R}^m, \\ g(x, y) \le 0, \quad f(x, y) \le \varphi(x). \end{cases} \tag{1.7}$$

It is easy to see that the nonlinear programming problem (1.7) is *unconditionally* equivalent to the optimistic bilevel program (1.2) when we are concerned with the global minimum; the relationships between local minima in problems (1.2) and (1.7) are more subtle, they are discussed in more detail in section 3 with further references therein. However, the price to pay in replacing problem (1.2) by the one in (1.7) is that the nonlinear program (1.7) is *intrinsically nonsmooth* (even for infinitely smooth initial data of the original bilevel program) due to the intrinsic nondifferentiability of the value function (1.6). It is worth mentioning that the appearance of functions like (1.6), known also as *marginal functions*, in problems of optimization and control, was among the strongest original motivations for the development of nonsmooth analysis.

Formulation (1.7) of the bilevel program (1.2) has been developed by Ye and Zhu [37]; see more discussions further. Based on this description and certain tools of nonsmooth analysis, some necessary optimality conditions in bilevel programming have been recently derived by Babbahadda and Gadhi [1] and Ye [34,35].

In the paper by Babbahadda and Gadhi [1], it is shown that the function φ admits a so-called *convexificator* provided that the functions f and $g_i, i = 1, \ldots, p$, have one. Then a necessary optimality condition of the corresponding nondifferentiable Fritz John type is derived for (1.7) by using convexificators. Furthermore, it is proved in [1] that the obtained Fritz John condition can be written in the *normal* (i.e. KKT) form under an appropriate regularity/constraint qualification requirement.

As has been observed by Ye and Zhu [37], the classical Mangasarian–Fromovitz constraint qualification is violated at every feasible point to problem (1.7). In [34], Ye introduced new constraint qualifications that can be satisfied at feasible points to problem (1.7) and derived, imposing those qualification assumptions, KKT-type necessary optimality conditions for this problem. The conditions obtained employ, in particular, the so-called Michel–Penot subdifferential for the optimal value function φ from (1.6). If the value function φ happens to be *convex*, this subdifferential can be efficiently estimated by using, e.g. the corresponding results from Gauvin and Dubeau [11], Shimizu *et al.* [31], and Tanino and Ogawa [32].

In her recent paper [35], Ye developed another idea to derive KKT-type conditions for bilevel programs in the framework of the value function formulation (1.7). This idea is to approximate the inequality $f(x, y) - \varphi(x) \leq 0$ by a function Ψ for which the Clarke-generalized gradient can be efficiently estimated. Then, after showing that problem (1.7) is locally equivalent to the nondifferentiable program

$$\begin{cases} \text{minimize } \ F(x, y) \ \text{ subject to } \ (x, y) \in X \times \mathbb{R}^m, \\ g(x, y) \leq 0, \quad \psi(x, y) \leq 0, \end{cases} \tag{1.8}$$

she derived KKT-type necessary optimality conditions for (1.7), assuming that the auxiliary problem (1.8) satisfies a nonsmooth counterpart of the classical Abadie constraint qualification; see [35] for more details and discussions.

In this article, we further develop the *value function approach* to derive new necessary optimality conditions for bilevel programs in finite-dimensional spaces. In fact, the results obtained can be extended to infinite-dimensional settings (see Remark 5.3 for more discussions), but we do not pursue this goal here. The main tools of our analysis are the basic generalized differential constructions (normals, subgradients, and coderivatives) by Mordukhovich, which satisfy comprehensive calculus rules. In particular, we employ advanced formulas for evaluating *basic subgradients* of the value function in the lower-level problem, which – in conjunction with appropriate constraint qualifications and a nonsmooth version of the Lagrange multiplier rule applied to the upper-level problem – play a key role in deriving necessary optimality conditions in bilevel programming. In this way, we arrive at new optimality conditions or bilevel programs whose certain significant features never been observed before. In particular, the resulting conditions obtained below for bilevel programs with smooth initial data in the parametric lower-level problem *do not depend*, under appropriate assumptions, on the partial derivatives with respect to parameters of the lower-level cost function. Other new features of the optimality conditions derived in the article are discussed in the subsequent sections.

The rest of the article is organized as follows. In section 2, we present some basic material on *generalized differentiation* and related properties widely used in the article. Section 3 is devoted to bilevel programs with *smooth* (actually *strictly differentiable* at the reference optimal solution) initial data in both lower-level and upper-level problems. In section 4, we consider in more detail bilevel programs with *fully convex* (in both variables) functions, exploring both smooth and nonsmooth settings that happen to be significantly different. We discuss specifications of the results obtained in the case of *linear programming* problems at the lower level. The concluding section 5 concerns bilevel programs with *Lipschitzian* initial data. The results derived for such problems by using the value function approach partly extend the corresponding results derived for smooth problems in section 3 and for fully convex problems in section 4, while certain important issues of the smooth and convex results turn out to be crucial for these underlying structures.

Our notation is basically standard; see the books [7,12,19].

2. Generalized differentiation and related properties

Advanced methods of modern variational analysis unavoidably relate to the study of nonsmooth objects (sets, functions, and set-valued mappings) and thus require appropriate tools of generalized differentiation. In this article, we use the basic/limiting constructions by Mordukhovich, which enjoy *full robust calculus* and turn out to be *minimal* among any constructions of this type satisfying natural properties employed in this article. We refer the reader to the books by Mordukhovich [18,19] and Rockafellar and Wets [29] for more details, additional material, extensive comments, and bibliographies. Unless otherwise stated, all the spaces under consideration are *finite-dimensional*.

We start with generalized normals to nonempty sets. Given $\Omega \subset \mathbb{R}^n$ and $\bar{x} \in \Omega$, the (basic, limiting) *normal cone* to Ω at $\bar{x}$ is defined by

$$N(\bar{x}; \Omega) := \limsup_{x \xrightarrow{\Omega} \bar{x}} \widehat{N}(x; \Omega), \tag{2.1}$$

where $x \xrightarrow{\Omega} \bar{x}$ means that $x \to \bar{x}$ with $x \in \Omega$, where $\widehat{N}(x; \Omega)$ stands for the *prenormal/ Fréechet normal cone* to Ω at $x \in \Omega$ defined by

$$\widehat{N}(x; \Omega) := \left\{ v \in \mathbb{R}^n \middle| \limsup_{u \xrightarrow{\Omega} x} \frac{\langle v, u - x \rangle}{\|u - x\|} \leq 0 \right\}, \tag{2.2}$$

and where "lim sup" signifies the *Kuratowski-Painlevée outer/upper limit* for a set-valued mapping $S: \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ as $u \to x$ given by

$$\limsup_{u \to x} S(u) := \left\{ v \in \mathbb{R}^m \mid \exists\, u_k \to u,\ \exists\, v_k \to v \ \text{ with } \ v_k \in S(u_k) \ \text{ as } \ k \to \infty \right\}.$$

The basic normal cone (2.1) is often *nonconvex* (e.g., for $\Omega = \text{gph}\,|x|$ at $\bar{x} = (0, 0) \in \mathbb{R}^2$), while the prenormal one (2.2) is always convex being in fact *polar/dual* to the classical Bouligand *contingent cone* to Ω at $x \in \Omega$. For convex sets Ω, both cones (2.1) and (2.2) reduce to the normal cone of convex analysis. Let us mention a convenient representation of the basic normal cone held for sets Ω locally closed around $\bar{x} \in \Omega$:

$$N(\bar{x}; \Omega) = \limsup_{x \to \bar{x}} \left[\left(x - \Pi(x; \Omega) \right) \right],$$

where "cone" stands for the *conic hull* of the set in question, and where $\Pi(x; \Omega)$ is the *Euclidean projector* of $x \in \mathbb{R}^n$ on the set Ω.

Given a set-valued mapping $S: \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ and a point $(\bar{x}, \bar{y}) \in \text{gph}\, S$ from its graph

$$\text{gph}\, S := \left\{ (u, v) \in \mathbb{R}^n \times \mathbb{R}^m \mid v \in S(u) \right\},$$

define the *coderivative* of S at $(\bar{x}, \bar{y})$ as a positive homogeneous mapping $D^*S(\bar{x}, \bar{y}): \mathbb{R}^m \rightrightarrows \mathbb{R}^n$ with the values

$$D^*S(\bar{x}, \bar{y})(v) := \left\{ u \in \mathbb{R}^n \mid (u, -v) \in N\left((\bar{x}, \bar{y}); \text{gph}\, S \right) \right\}, \quad v \in \mathbb{R}^m. \tag{2.3}$$

If S is single-valued and *strictly differentiable* at $\bar{x}$ with the gradient $\nabla S(\bar{x})$, in the now classical sense that

$$\lim_{x, u \to \bar{x}} \frac{S(u) - S(x) - \langle \nabla S(\bar{x}), u - x \rangle}{\|u - x\|} = 0$$

(this is obviously the case when S is continuously differentiable around $\bar{x}$), then

$$D^*S(\bar{x})(v) = \{\nabla S(\bar{x})^* v\} \quad \text{for all} \quad v \in \mathbb{R}^n; \tag{2.4}$$

we omit $\bar{y}$ in the coderivative notation (2.3) for single-valued mappings. Representation (2.4) shows that the coderivative is a proper extension of the *adjoint* derivative operator to nonsmooth and set-valued mappings.

Given a function $\varphi: \mathbb{R}^n \to \mathbb{R}$ *Lipschitz-continuous* around $\bar{x}$ (we do not consider more general nonsmooth functions in this article), define its (basic, limiting) *subdifferential* by

$$\partial\varphi(\bar{x}) = \limsup_{x \to \bar{x}} \widehat{\partial}\varphi(x) \tag{2.5}$$

via the Painlevé–Kuratowski outer limit of the so-called *Fréchet/viscosity subdifferentials*

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

of φ at x. The basic subdifferential (2.5) is always *nonempty* and *compact* for every locally Lipschitzian function. It reduces to the classical gradient

$$\partial\varphi(\bar{x}) = \{\nabla\varphi(\bar{x})\} \tag{2.6}$$

for strictly differentiable functions and to the subdifferential of convex analysis for convex ones. Note that the basic subdifferential (2.5) can be equivalently defined geometrically

$$\partial\varphi(\bar{x}) = \{ v \in \mathbb{R}^n \mid (v, -1) \in N((\bar{x}, \varphi(\bar{x})); \varphi) \}$$

via the basic normal cone (2.1) to the epigraph of φ. At the same time, the geometrically defined coderivative (2.3) of a single-valued locally Lipschitzian mapping $S: \mathbb{R}^n \to \mathbb{R}^m$ can be represented analytically as

$$D^*S(\bar{x})(v) = \partial\langle v, S\rangle(\bar{x}) \neq \emptyset \qquad \text{for all} \quad v \in R^m,$$

via the basic subdifferential (2.5) of the *Lagrange scalarization* $\langle v, S\rangle(x) := \langle v, S(x)\rangle$.

In this article, we use the *convex hull property*

$$\text{co}\,\partial(-\varphi)(\bar{x}) = -\text{co}\,\partial\varphi(\bar{x}) \tag{2.7}$$

of the basic subdifferential; this follows from the fact that the convex hull of $\partial\varphi(\bar{x})$ agrees with the Clarke-generalized gradient for locally Lipschitzian functions, which enjoys the classical plus-minus symmetry.

Given a set-valued mapping $S: \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ and a point $\bar{x}$ with $S(\bar{x}) \neq \emptyset$, we say that S is *inner semicompact* at $\bar{x}$ if for every sequence $x_\nu \to \bar{x}$ with $S(x_\nu) \neq \emptyset$ there is a sequence of $y_\nu \in S(x_\nu)$ that contains a convergent subsequence as $\nu \to \infty$. It is clear, in the finite-dimensional setting under consideration, that the inner semicompactness holds whenever S is *uniformly bounded* around $\bar{x}$, i.e. there exists a neighborhood U of $\bar{x}$ and a bounded set $C \subset \mathbb{R}^m$ such that

$$S(x) \subset C \quad \text{whenever} \quad x \in U.$$

We say that $S: \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ is *inner semicontinuous* at $(\bar{x}, \bar{y}) \in \text{gph}\, S$ if for any sequence $x_\nu \to \bar{x}$ there is a sequence $y_\nu \in S(x_\nu)$ converging to $\bar{y}$ as $\nu \to \infty$. For single-valued mappings, this clearly reduces to the standard continuity of S at $\bar{x}$. In the set-valued case, the inner semicontinuity at $\bar{x}$ of an inner semicompact closed-graph mapping S surely holds when $S(\bar{x})$ is a singleton just *at* the reference point $\bar{x}$, which occurs in many applications; see more discussions in section 3 on the fulfillment of the inner semicontinuity assumption in the framework of bilevel programming. Note that the inner semicontinuity of S at $(\bar{x}, \bar{y})$ for *every* $\bar{y} \in S(\bar{x})$ goes back to the standard notion of *lower/inner* semicontinuity of S at $\bar{x}$ widely used in the literature. It is clear that the inner semicontinuity of S at $(\bar{x}, \bar{y})$ is implied by its following Lipschitz-like behavior around this point that extends the classical *local Lipschitz continuity* of set-valued mappings in the Hausdorff sense.

Recall that $S: \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ satisfies the *Aubin/Lipschitz-like property* around the point $(\bar{x}, \bar{y}) \in \operatorname{gph} S$ if there are neighborhoods U of $\bar{x}$, V of $\bar{y}$ and a constant $\ell \geq 0$ such that

$$S(x) \cap V \subset S(u) + \ell \|x - u\| \mathbb{B} \quad \text{for all} \quad x, u \in U,$$

where $\mathbb{B}$ stands for the closed unit ball of the space in question. This property is known to be equivalent to *metric regularity* and *linear openness* of the inverse mapping S^{-1}, which are of crucial importance for many aspects of nonlinear analysis and optimization. It is worth mentioning that the *coderivative criterion* $D^*S(\bar{x}, \bar{y})(0) = \{0\}$ provides a *complete characterization* of the Lipschitz-like property of S around $(\bar{x}, \bar{y})$.

3. Bilevel programs with smooth initial data

In this section, we consider the *optimistic version* of the basic *bilevel programming* problem given in the following form:

$$\text{minimize} \;\; F(x, y) \;\; \text{subject to} \;\; x \in X, \quad y \in \Psi(x), \tag{3.1}$$

where $F: \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ is the upper-level objective function, $X \subset \mathbb{R}^n$ is the upper level constraint set, and $\Psi: \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ is the set-valued mapping given by

$$\Psi(x) := \operatorname*{Argmin}_y \big\{ f(x, y) \,|\, y \in K(x) \big\}, \tag{3.2}$$

which describes the parameterized (by x) set of *optimal solutions* to the lower-level problem:

$$\text{minimize} \;\; f(x, y) \;\; \text{subject to} \;\; y \in K(x) \tag{3.3}$$

with $f: \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ and $K: \mathbb{R}^n \rightrightarrows \mathbb{R}^m$. For simplicity, we focus in what follows on the case when the constraint mapping $K(x)$ in the lower-level problem and the constraint set x in the upper-level problem are described by functional constraints of the *inequality type*:

$$K(x) = \big\{ y \in \mathbb{R}^m \,|\, g_i(x, y) \leq 0, \; i = 1, \ldots, p \big\}, \quad x \in \mathbb{R}^n, \tag{3.4}$$

$$X = \{ x \in \mathbb{R}^n \,|\, G_j(x) \leq 0, \; j = 1, \ldots, k \big\} \tag{3.5}$$

via real-valued functions g_i and G_j. In fact, our methods and results obtained below hold in more general constraint settings; it is discussed in Remark 3.3 and 5.2 further.

In the introductory section, we discussed the relationship between the above optimistic version of bilevel programs and other versions and approaches in bilevel programming; see also Dempe [7] and Dutta and Dempe [10] for more details. To proceed further without triviality, we always assume that the constraint sets $K(x)$ in the lower-level problem are nonempty for all $x \in X \neq \emptyset$. It follows immediately from the classical Weierstrass theorem that the argminimum sets $\Psi(x)$ in (3.2) are nonempty whenever the constraint sets $K(x)$ are compact and the lower-level cost function f is lower semicontinuous in y on $K(x)$. However, from the viewpoint of deriving necessary optimality conditions – which is the main issue in this article – it does not make sense to impose any compactness assumptions. Note that, given a local optimal solution, we automatically presume the *nonemptiness* of all the (moving) constraint sets at the points under consideration.

The main attention in this section is paid to deriving *necessary optimality conditions* for the optimistic bilevel program described in (3.1)–(3.5), where all the functions are supposed to be *smooth*, in the sense that they are *strictly differentiable* at the reference points, which always happens when they are continuously differentiable around these points.

Developing the *value function approach* to bilevel programming, consider the problem

$$\begin{cases} \text{minimize } \; F(x,y) \; \text{ subject to } \; x \in X, \; y \in K(x), \\ \text{and } \; f(x,y) - \varphi(x) \leq 0, \end{cases} \tag{3.6}$$

where $K(x)$ and x are given in (3.4) and (3.5), respectively, and where

$$\varphi(x) := \inf\left\{ f(x,y) \mid y \in K(x) \right\} \tag{3.7}$$

is the *optimal value function* in the parametric lower-level problem (3.3). It is clear that the latter problem is *globally* equivalent to the original optimistic bilevel program (3.1) with $\Psi(x)$ given in (3.2), while *local* optimal solutions to problem (3.1), (3.2) are always locally optimal to the value function problem (3.6), (3.7). Since the original optimistic bilevel program we are dealing with is problem (1.2), it is important to clarify how local optimal solutions for (1.2) are related to those for (3.1). This issue has been addressed in the paper by Dutta and Dempe [10], where it has been shown that if the argminimum mapping Ψ is upper semicontinuous in the sense of set-valued analysis, then a local optimistic solution to (1.2) is also a local optimistic solution to the problem (3.1) under consideration. It has been later realized in our personal communication with Outrata [25] that the upper semicontinuity is more than we need. In fact, to keep the required results of [10], it is sufficient to impose the *uniform boundedness* assumption on the solution/argminimum mapping Ψ around the reference local optimistic solution to (1.2). The latter assumption is fully in accordance with the *inner semicompactness* requirement on Ψ imposed in the present article from the viewpoint of necessary optimality conditions. This observation allows us to derive necessary conditions for local optimality in the optimistic bilevel program (3.1), (3.2) by doing it in the framework of nonlinear programming of the *special nonstandard type* (3.6), (3.7).

In what follows, we thus concentrate on problem (3.6) involving the value/marginal function (3.7), where the constraint mapping $K(x)$ and set x are given in the functional forms (3.4) and (3.5), respectively. Note that in spite of *smoothness* of all the functions involving in the original formulation of the bilevel program under consideration, the nonlinear programming problem (3.4)–(3.7) is nonstandard for at least *two reasons*. *First*, as mentioned in section 1, the value function $\varphi(x)$ from (3.7) involved in the inequality constraints in (3.6) is *intrinsically nonsmooth*. The *second* principal obstacle is that, due to the specific structure of (3.6), the *classical constraint qualification conditions* (like the Mangasarian–Fromovitz and Slater ones) naturally extended to the nonsmooth case *fail to hold* for (3.6); see [37,38] for more details. In this setting, Ye and Zhu [37] suggested a new constraint qualification condition for (3.6), which is formulated as follows.

Consider the *perturbed* version of (3.6), linearly parameterized by $u \in \mathbb{R}$, in the form:

$$\min_{x,y} F(x,y) \text{ subject to } f(x,y) - \varphi(x) + u = 0, \quad y \in K(x), \; x \in X. \tag{3.8}$$

Following [37], we say that the unperturbed problem (3.6) is *partially calm* at its given feasible point $(\bar{x},\bar{y})$ if there are a constant $\lambda > 0$ and a neighborhood U of the triple $(\bar{x},\bar{y},0) \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}$, such that for all $(x,y,u) \in U$ feasible to (3.8) one has

$$F(x,y) - F(\bar{x},\bar{y}) + \lambda|u| \geq 0. \tag{3.9}$$

In [37,38], the reader can find more details and discussions on partial calmness, its relationships with other constraint qualifications, and efficient conditions ensuring its validity for important classes of optimization problems.

To derive necessary optimality conditions of the *KKT type* in the bilevel programming problem (3.1)–(3.5), we also need to impose appropriate *regularity assumptions* on the *inequality constraints* in the lower-level and upper-level problems.

Given a point $(\bar{x},\bar{y})$ satisfying the lower-level inequality constraints (3.4), we denote

$$I(\bar{x},\bar{y}) := \Big\{ i \in \{1,\dots,p\} \mid g_i(\bar{x},\bar{y}) = 0 \Big\} \tag{3.10}$$

and say that $(\bar{x},\bar{y}) \in \mathbb{R}^n \times \mathbb{R}^m$ is *lower-level regular* if for any nonnegative λ_i, one has

$$\Bigg[\sum_{i \in I(\bar{x},\bar{y})} \lambda_i \nabla_y g_i(\bar{x},\bar{y}) = 0 \Bigg] \Longrightarrow [\lambda_i = 0 \text{ whenever } i \in I(\bar{x},\bar{y})]. \tag{3.11}$$

Similarly, given $\bar{x} \in \mathbb{R}^n$ satisfying the upper-level inequality constraints (3.5), we denote

$$J(\bar{x}) := \Big\{ j \in \{1,\dots,k\} \mid G_j(\bar{x}) = 0 \Big\} \tag{3.12}$$

and say that $\bar{x}$ is *upper-level regular* if for any nonnegative λ_j one has

$$\Big[\sum_{j \in J(\bar{x})} \lambda_j \nabla G_j(\bar{x}) = 0 \Big] \Longrightarrow \Big[\lambda_j = 0 \text{ whenever } j \in J(\bar{x}) \Big]. \tag{3.13}$$

One can easily recognize that the regularity conditions (3.11) and (3.13) are dual versions of the classical Mangasarian–Fromovitz constraint qualification for the inequality constraints in the lower-level and upper-level problems, respectively. Now we are ready to establish the following new necessary optimality conditions for smooth bilevel programs.

THEOREM 3.1 (Necessary conditions in smooth bilevel programming). *Let* $(\bar{x},\bar{y})$ *be a local optimal solution to the bilevel program described in* (3.1)–(3.5), *where all the functions are strictly differentiable at* $(\bar{x},\bar{y})$ *and* $\bar{x}$, *respectively. Assume that the bilevel program under consideration is partially calm at* $(\bar{x},\bar{y})$ *and that* $(\bar{x},\bar{y})$ *is lower-level regular* (3.11) *while* $\bar{x}$ *is upper-level regular* (3.13). *Furthermore, we suppose that the* *argminimum mapping* Ψ *in* (3.2) *is inner semicontinuous at* $(\bar{x}, \bar{y})$. *Then there are real numbers* $\lambda > 0$, λ_{is} *with* $i = 1, \ldots, p$ *and* $s = 1, \ldots, n+1$, μ_i *with* $i = 1, \ldots, p$, η_s *with* $s = 1, \ldots, n+1$, *and* α_j *with* $j = 1, \ldots, k$, *such that the following conditions hold*:

$$\nabla_x F(\bar{x}, \bar{y}) + \sum_{i=1}^{p} \mu_i \nabla_x g_i(\bar{x}, \bar{y}) - \lambda \sum_{s=1}^{n+1} \eta_s \Big(\sum_{i=1}^{p} \lambda_{is} \nabla_x g_i(\bar{x}, \bar{y}) \Big) + \sum_{j=1}^{k} \alpha_j \nabla G_j(\bar{x}) = 0, \tag{3.14}$$

$$\nabla_y F(\bar{x}, \bar{y}) + \lambda \nabla_y f(\bar{x}, \bar{y}) + \sum_{i=1}^{p} \mu_i \nabla_y g_i(\bar{x}, \bar{y}) = 0, \tag{3.15}$$

$$\nabla_y f(\bar{x}, \bar{y}) + \sum_{i=1}^{p} \lambda_{is} \nabla_y g_i(\bar{x}, \bar{y}) = 0 \ \text{ for all } \ s = 1, \ldots, n+1. \tag{3.16}$$

$$\lambda_{is} \geq 0, \ \lambda_{is} g_i(\bar{x}, \bar{y}) = 0 \ \text{ for all } \ i = 1, \ldots, p, \ s = 1, \ldots, n+1, \tag{3.17}$$

$$\mu_i \geq 0, \quad \mu_i g_i(\bar{x}, \bar{y}) = 0 \ \text{ for all } \ i = 1, \ldots, p, \tag{3.18}$$

$$\eta_s \geq 0 \ \text{ for all } \ s = 1, \ldots, n+1, \quad \sum_{s=1}^{n+1} \eta_s = 1, \tag{3.19}$$

$$\alpha_j \geq 0 \ \text{ and } \ \alpha_j G_j(\bar{x}) = 0 \ \text{ for all } \ j = 1, \ldots, k. \tag{3.20}$$

Proof Since $(\bar{x}, \bar{y})$ is a local minimizer to the bilevel program (3.1)–(3.5), it is a local optimal solution to the single-level mathematical program (3.6) involving the value function (3.7) and the constraints given by (3.4) and (3.5). The latter problem reads as follows:

$$\begin{cases} \text{minimize } \ F(x, y) \ \text{ subject to} \\ f(x, y) - \varphi(x) \leq 0, \\ g_i(x, y) \leq 0, \quad i = 1, \ldots, p, \\ G_j(x) \leq 0, \quad j = 1, \ldots, k, \end{cases} \tag{3.21}$$

where the value function $\varphi(x)$ from (3.7) is represented by

$$\varphi(x) = \inf \big\{ f(x, y) \,|\, g_i(x, y) \leq 0 \ \text{ as } \ i = 1, \ldots, p \big\}. \tag{3.22}$$

Observe that the value function (3.7) is *locally Lipschitzian* around $\bar{x}$ under the assumptions made. Indeed, it follows from [18, Corollary 4.39] that the constraint mapping $K(x)$ in (3.4) is *Lipschitz-like* around $(\bar{x}, \bar{y})$ under the lower-level regularity of $(\bar{x}, \bar{y})$ – this follows from the coderivative criterion mentioned in the end of section 2. Therefore, the result of [20, Theorem 5.2(i)] ensures the Lipschitz continuity of the value/marginal function (3.7), since the argminimum mapping Ψ from (3.2) in case (3.4), is assumed to be inner semicontinuous at $(\bar{x}, \bar{y})$, while the cost function f is locally Lipschitzian around this point.

Applying next [37, Proposition 3.3] to problem (3.21), we conclude that the imposed partial calmness condition at $(\bar{x}, \bar{y})$ is *equivalent* to the existence of $\lambda > 0$, such that $(\bar{x}, \bar{y})$

solves (locally) the following *penalized* constrained problem, where the value function φ is moved from the constraint to the cost functional:

$$\begin{cases} \text{minimize } \ F(x,y) + \lambda\big(f(x,y) - \varphi(x)\big) \ \text{ subject to} \\ g_i(x,y) \le 0, \quad i = 1,\dots,p, \\ G_j(x) \le 0, \quad j = 1,\dots,k. \end{cases} \tag{3.23}$$

By the discussion above, (3.27) is a problem of *Lipschitzian programming*. Applying the *generalized Lagrange multiplier rule* from [19, Theorem 5.21(iii)] – in terms of the basic subdifferential (2.5) – to the local optimal solution $(\bar{x},\bar{y})$ for (3.23) and taking into account the subdifferential sum rule from [18, Proposition 1.107(ii)] and the subdifferential representation (2.6) for strictly differentiable functions, we find *nonnegative* multipliers $(\lambda_0, \mu_1, \dots, \mu_p, \alpha_1, \dots, \alpha_k) \neq 0$ satisfying the Lagrangian inclusion

$$\begin{aligned} 0 \in \lambda_0 \nabla F(\bar{x},\bar{y}) &+ \lambda_0 \lambda \nabla f(\bar{x},\bar{y}) + \big(\lambda_0 \lambda \partial(-\varphi)(\bar{x}), 0\big) \\ &+ \sum_{i=1}^{p} \mu_i \nabla g_i(\bar{x},\bar{y}) + \sum_{j=1}^{k} \alpha_j \big(\nabla G_j(\bar{x}), 0\big) \end{aligned} \tag{3.24}$$

and the *complementary slackness conditions* in (3.18) and (3.20). It easily follows from the assumed *lower-level regularity* of $(\bar{x},\bar{y})$ and *upper-level regularity* of $\bar{x}$, combined with sign and complementarity slackness conditions on $(\mu_1, \dots, \mu_p, \alpha_1, \dots, \alpha_k)$, that $\lambda_0 > 0$ in (3.24). Hence, the latter inclusion is equivalent to the *KKT-type condition*

$$\begin{aligned} 0 \in \nabla F(\bar{x},\bar{y}) &+ \lambda \nabla f(\bar{x},\bar{y}) + \big(\lambda \partial(-\varphi)(\bar{x}), 0\big) \\ &+ \sum_{i=1}^{p} \mu_i \nabla g_i(\bar{x},\bar{y}) + \sum_{j=1}^{k} \big(\alpha_j \nabla G_j(\bar{x}), 0\big) \end{aligned} \tag{3.25}$$

along with the above sign and complementary slackness conditions for μ_i and α_j. To proceed, we observe the inclusion

$$\partial(-\varphi)(\bar{x}) \subset -\mathrm{co}\, \partial\varphi(\bar{x}) \tag{3.26}$$

valid due to the symmetry property (2.7), where the sets $\partial\varphi(\bar{x})$ and $\mathrm{co}\,\partial\varphi(\bar{x})$ are *nonempty* and *compact* due to the local Lipschitz continuity of φ around $\bar{x}$. Furthermore, applying [22, Corollary 4] that holds under the *lower-level regularity* and the *inner semicontinuity* conditions assumed in the theorem, we get the following upper estimate of the basic subdifferential of the value/marginal function φ at $\bar{x}$:

$$\partial\varphi(\bar{x}) \subset \bigcup_{(\lambda_1,\dots,\lambda_p)\in\Lambda(\bar{x},\bar{y})} \Big[\nabla_x f(\bar{x},\bar{y}) + \sum_{i=1}^{p} \lambda_i \nabla_x g_i(\bar{x},\bar{y})\Big], \tag{3.27}$$

where the union is taken over the set

$$\begin{aligned} \Lambda(\bar{x},\bar{y}) := \Big\{(\lambda_1, \dots, \lambda_p) \in \mathbb{R}^p \Big| \nabla_y f(\bar{x},\bar{y}) + \sum_{i=1}^{p} \lambda_i \nabla_y g_i(\bar{x},\bar{y}) = 0, \\ \lambda_i \ge 0, \ \ \lambda_i g_i(\bar{x},\bar{y}) = 0, \ \ i = 1, \dots, p\Big\}. \end{aligned} \tag{3.28}$$

Picking $v \in \operatorname{co} \partial\varphi(\bar{x})$ and using the classical Carathéodory theorem, we find $\eta_s \in \mathbb{R}$ and $v_s \in \mathbb{R}^n$ with $s = 1, \ldots, n+1$, such that

$$v = \sum_{s=1}^{n+1} \eta_s v_s, \quad \sum_{s=1}^{n+1} \eta_s = 1, \ \ \eta_s \geq 0, \ \ v_s \in \partial\varphi(\bar{x}) \text{ for } s = 1, \ldots, n+1. \tag{3.29}$$

Applying (3.27) to each v_s from (3.29), we thus find λ_{is}, such that

$$\lambda_{is} \geq 0, \ \ \lambda_{is} g_i(\bar{x}, \bar{y}) = 0 \text{ for } i = 1, \ldots, p; \ \ \nabla_y f(\bar{x}, \bar{y}) + \sum_{i=1}^{p} \lambda_{is} \nabla_y g_i(\bar{x}, \bar{y}) = 0,$$

$$v_s = \nabla_x f(\bar{x}, \bar{y}) + \sum_{i=1}^{p} \lambda_{is} \nabla_x g_i(\bar{x}, \bar{y}), \quad s = 1, \ldots, n+1.$$

Combining all the above via (3.25), we arrive at the necessary optimality conditions (3.14–3.20) and conclude the proof of the theorem. ■

Let us discuss certain *characteristic features* of the necessary optimality conditions derived in Theorem 3.1 and some of its natural *modifications* and *extensions*.

Remark 3.1 (Independence of optimality conditions on partial derivatives of smooth lower-level cost functions). Observe that the necessary optimality conditions for bilevel programs with smooth data obtained in Theorem 3.1 are *independent* of the *partial derivative* of the *lower-level cost function* $f(x, y)$ with respect to the *parameter* variable x. This seems to be a *new feature* in the study of bilevel programs that has not been observed before. On the other hand, it does not come as a full surprise. We refer the reader to the paper by Dempe and Lohse [9], where a similar phenomenon was observed in somewhat different setting of linear bilevel programming with parameterized lower-level constraints.

Remark 3.2 (Inner semicontinuity of the argminimum map). A crucial assumption of Theorem 3.1 is the *inner semicontinuity* of the solution/argminimum map Ψ (3.2) at the given optimal pair $(\bar{x}, \bar{y}) \in \operatorname{gph} \Psi$. Observe that it automatically holds if $\Psi(\bar{x}) = \{\bar{y}\}$ is a *singleton*, while the solution map Ψ may be multivalued at any point different from $\bar{x}$. In the books by Dempe [7], Outrata *et al.* [26], and their references the reader can find descriptions and discussions of various settings with *local uniqueness* of solution maps to lower-level problems. Let us mention some important conditions that ensure the inner semicontinuity of solution maps in parametric optimization, which thus can be applied to the *set-valued* argminimum map in Theorem 3.1:

(a) Constraint functions (in the lower-level problem) are *weakly analytic* in the sense of Klatte and Kummer; see more details and efficient conditions for this in Theorem 4.3.5 of the book [2]. Note that such functions are closely related to the so-called *LFS-functions* in the sense of Zlobec [40]. Observe that the *weak analyticity* requirement may be *far removed* from *single-valuedness* of the argminimum map, *linearity* of the lower-level problem, and/or *convexity* of the optimal value function; see more examples and discussions in the books [2,40] and the references therein. It is worth mentioning that the *inner semicontinuity* of the argminimum map is generally *not related* to the *convexity* of the optimal value function in the lower-level problem as demonstrated by the following simple example: $\min\{-y \mid y \leq x^3\}$, where $\Psi(x) = \{x^3\}$ and $\varphi(x) = -x^3$.

(b) The *second-order sufficient condition* in the sense of Robinson [27] is satisfied in the lower-level problem. This holds, in particular, if the lower-level objective is locally *strictly/strongly convex* with respect to the decision variable y.
(c) Parametric problems of *linear programming* with *additive* (right-hand side) *perturbations*; see more details in the next section.
(d) The solution map Ψ is *Lipschitz-like* around $(\bar{x}, \bar{y})$; in particular, it is *locally Lipschitzian* around $\bar{x}$. Many results ensuring Lipschitz-like behavior of solution maps are established (based on the coderivative criterion and the appropriate *second-order subdifferential* of nonsmooth functions) in the book by Mordukhovich [18, Chapter 4] and the references therein for various optimization and optimization-related problems (e.g. for variational inequalities). We particularly refer the reader to the paper by Levy and Mordukhovich [13], where efficient conditions of this type are derived and discussed for solution maps in parametric nonlinear programming.

Remark 3.3 (Smooth constraints of the equality type). The necessary optimality conditions derived above can be easily extended to the case of *additional equality constraints*

$$h_i(x, y) = 0, \quad i = p + 1, \ldots, p + q, \tag{3.30}$$
$$H_j(x) = 0, \quad j = k + 1, \ldots, k + l, \tag{3.31}$$

in (3.4) and (3.5) governed by smooth (i.e. *strictly differentiable*, as adopted in this article) functions at the lower and upper levels. This follows directly from the proof of Theorem 3.1, since the corresponding results used therein are established in fact for problems with both inequality and equality constraints. The only adjustment to mention in the formulation of Theorem 3.1 for bilevel programs involving additional equality constraints is that the lower-level and upper-level regularity should be now understood in the sense of fulfilling the *full counterparts* of the Mangasarian–Fromovitz constraints qualifications in (3.11) and (3.13) and that, of course, the *sign* conditions relate to multipliers corresponding just to *inequality* constraints. We consider here our basic bilevel model involving merely inequalities for the following two reasons: first – for simplicity, and second – to address specific features of *convex* problems studied in the next section, where convexity is appropriate for inequality *versus* equality constraints; the latter ones require linearity.

Remark 3.4 (Necessary conditions in smooth bilevel programming with no inner semicontinuity). The approach used in the proof of Theorem 3.1 and the results developed in [19,20,22] allow us to derive necessary optimality conditions for the bilevel program (3.1)–(3.5) with the smooth data *with no* inner semicontinuity assumption imposed in the theorem. It should be replaced by the *inner semicompactness* at $\bar{x}$ of the argminimum mapping Ψ from (3.2) with $K(\cdot)$ given in (3.4). The latter assumption is much more flexible and holds, in particular, when Ψ is *uniformly bounded* around $\bar{x}$ in the finite-dimensional setting under consideration. In addition, we need to assume that the strict differentiability and upper-level regularity (but *not* partial calmness) requirements of the theorem are satisfied at $(\bar{x}, y)$ *for all* $y \in \Psi(\bar{x})$. In this case, the value function φ in (3.7) is *locally Lipschitzian* around $\bar{x}$ by [20, Theorem 5.1(ii)] and – from

[Theorem 5.38(ii) and Corollary 4.35] – we have the following counterpart of the upper estimate (3.27):

$$\partial\varphi(\bar{x}) \subset \bigcup_{y\in\Psi(\bar{x})} \Big[\bigcup_{(\lambda_1,\ldots,\lambda_p)\in\Lambda(\bar{x},y)} \Big\{ \nabla_x f(\bar{x},y) + \sum_{i=1}^{p} \lambda_i \nabla_x g_i(\bar{x},y) \Big\} \Big], \tag{3.32}$$

where the set of multipliers $\Lambda(\bar{x}, y)$ is defined in (3.28) with the replacement of $\bar{y}$ by $y \in \Psi(\bar{x})$. Following now the proof of Theorem 3.1 and using the upper estimate (3.32) instead of (3.27), we find numbers $(\lambda, \lambda_{is}, \mu_i, \eta_s, \alpha_j)$ as in Theorem 3.1 and also vectors $y_s \in \Psi(\bar{x})$ with $s = 1, \ldots, n+1$, such that one has conditions (3.15), (3.18)–(3.20) along with (3.16) and (3.17) where $\bar{y}$ is replaced by y_s, and with the new optimality condition

$$\begin{aligned} &\nabla_x F(\bar{x},\bar{y}) + \lambda\nabla_x f(\bar{x},\bar{y}) - \lambda\sum_{s=1}^{n+1}\eta_s\nabla_x f(\bar{x},y_s) + \sum_{i=1}^{p}\mu_i\nabla_x g_i(\bar{x},\bar{y}) \\ &\quad - \lambda\sum_{s=1}^{n+1}\eta_s\Big(\sum_{i=1}^{p}\lambda_{is}\nabla_x g_i(\bar{x},y_s)\Big) + \sum_{j=1}^{k}\alpha_j\nabla G_j(\bar{x}) = 0 \end{aligned} \tag{3.33}$$

replacing (3.14). Observe that the new conditions just derived reduce to those in Theorem 3.1 when $\Psi(x) = \{\bar{y}\}$, while Theorem 3.1 has been proved for Ψ assumed to be merely *inner semicontinuous* at $(\bar{x}, \bar{y})$. Thus, the two results obtained are generally *independent*. Since the necessary optimality conditions presented in this remark follow from more general necessary conditions of Theorem 5.1 established for Lipschitzian bilevel programs in section 5, we do not formulate them as a separate theorem.

Remark 3.5 (Comparison with earlier work). In the case when the lower level optimal solution is *unique*, the result of Theorem 3.1 agrees with [37, Theorem 3.1] for optimistic bilevel programs with C^1 data. In the general case, under the *inner semicontinuity* assumption, the necessary optimality conditions of Theorem 3.1 are *different* from those in [37, Theorem 3.1] in several *essential points* (being more simple and providing additional information). Moreover – as seems to us – the inner semicompactness (or uniform boundedness) requirement on Ψ is just *missed*, while it is needed. If the inner semicontinuity assumption is replaced by inner semicompactness the necessary optimality conditions in (3.33) are close to the results in [37, Theorem 3.1]. It is worth mentioning besides this that the upper-level problem in [37] does not contain functional constraints as in (3.5) – since it is assumed therein that $\bar{x} \in \mathrm{int}X$.

Remark 3.6 (Partial calmness condition). Probably the most restrictive assumption of Theorem 3.1 is the *partial calmness condition*. It is actually *necessary and sufficient* for the *exactness* of the *Upenalty function approach* used in (3.23). If this exact penalty function approach cannot be used, the KKT-type necessary optimality conditions in Theorem 3.1 need to be replaced by Fritz John-type conditions containing an additional multiplier λ_0 at the upper-level objective function. The results in the papers [34,35] indicate that this assumption can be replaced by other assumptions in some situations. The papers by Ye and Zhu [37,38] give a number of sufficient conditions for partial calmness. One sufficient assumption is that the lower-level problem (1.1) has a *uniform weak sharp minimum*. Necessary conditions for weak sharp minima in optimization problems have been broadly investigated; see e.g. [4,5] and the references therein. Sufficient conditions for uniform weak sharp minima can be found in Ye [33]. Let us particularly mention the following interesting result by Burke and Ferris [5]: *optimal solutions of linear programming problems are weak sharp minima whenever the problems are solvable*; cf. also related developments by Mangasarian and Meyer [15].

Let us also mention a helpful *characterization* of partial calmness provided by [37, Proposition 3.3], which shows that it is sufficient to check whether the reference local optimal solution $(\bar{x}, \bar{y})$ is a local optimal solution of problem (3.23) for some positive λ. The following example illustrates an efficient procedure of using this characterization for the *partial calmness verification* in nonlinear settings.

Example 3.7 (Verification of partial calmness via exact penalization). Let $(\bar{x}, \bar{y})$ be a local optimal solution to the bilevel program (3.1). We have already employed in the proof of Theorem 3.1 the following fact from [37, Proposition 3.3]: problem (3.1) is *partially calm* at $(\bar{x}, \bar{y})$ *if and only if* there is $\lambda > 0$, such that $(\bar{x}, \bar{y})$ solves the *penalized* problem:

$$\text{minimize} \;\; F(x,y) + \lambda\big(f(x,y) - \varphi(x)\big) \;\; \text{subject to} \;\; x \in X, \quad g(x,y) \le 0,$$

where φ is the optimal value function (3.7) to the lower-level problem. Now we show that this characterization is a convenient tool for the partial calmness verification. Consider the *fully nonlinear*, at both lower and upper levels, bilevel program (3.1) with $(x,y) \in \mathbb{R}^2$, $X = \mathbb{R}$, and

$$F(x,y) := \frac{(x-1)^2}{2} + \frac{y^2}{2}, \quad \Phi(x) := \underset{y}{\operatorname{Argmin}}\left\{\frac{x^2}{2} + \frac{y^2}{2}\right\}.$$

It is easy to see that $\Psi(x) \equiv \{0\}$ and $\varphi(x) = x^2/2$. Furthermore, $(\bar{x}, \bar{y}) = (1, 0)$ is the only solution to the upper-level problem, and so it is an optimal solution to the bilevel program under consideration. We have $f(x,y) - \varphi(x) = y^2/2$, and hence the corresponding unconstrained penalized problem is as follows:

$$\text{minimize} \;\; \frac{(x-1)^2}{2} + \frac{y^2}{2} + \frac{\lambda y^2}{2}.$$

For any $\lambda > 0$, the latter problem is smooth, strictly convex, and has the unique optimal solution $(\bar{x}, \bar{y}) = (1, 0)$. Thus, the initial bilevel program is *partially calm* at this point.

On the other hand, it is interesting to observe – again by using [37, Proposition 3.3] – that replacing the cost function $F(x,y)$ in the above upper-level problem by

$$\frac{(x-1)^2}{2} + \frac{(y-1)^2}{2},$$

and keeping the same lower-level problem, we arrive at the optimistic bilevel program (3.1) with the optimal solution $(\bar{x}, \bar{y}) = (1, 0)$, which *fails to satisfy the partial calmness condition*. Indeed, it is easy to see that the corresponding penalized problem has the optimal solution

$$\left(1, \frac{1}{1+\lambda}\right) \neq (1, 0) \;\; \text{whenever} \;\; \lambda > 0.$$

The above example shows that partial calmness in bilevel programs *may significantly depend* on the structure of upper-level objectives. On the contrary, the next example describes a rather *general class of multidimensional bilevel programs* with nonlinear lower-level problems when partial calmness holds *independently of the upper level*.

Example 3.8 (Partial calmness in nonlinear bilevel programs via weak sharp minima). First we consider the following *quadratic*-constrained optimization problem with respect to $x=(x_1,x_2,x_3)\in\mathbb{R}^3$:

$$\text{minimize }\ \frac{x_1^2}{2}+\frac{x_2^2}{2}\ \text{ subject to }\ a_i\le x_i\le b_i,\quad i=1,2,3.$$

It is shown in [5] that any (definitely not unique) optimal solution to this problem is a *uniform weak sharp minimum* whenever either $a_i>0$ or $b_i<0$ for $i=1,2$. Therefore, *every* bilevel program having this problem at the lower level (where e.g. a_i, b_i are the parameters) is *uniformly partially calm* at its optimal solutions [37], provided that the above condition remains valid. In fact, this quadratic example generates a fairly *broad class* of *nonlinear bilevel programs* exhibiting the same partial calmness phenomenon. Indeed, the deep topological result of [12, Theorem 2.4.2] ensures that *any* sufficiently smooth constraint optimization problem on $\mathbb{R}^n$ with *nondegenerated critical points* can be equivalently reduced (by a C^1-diffeomorphism) to a quadratic optimization problem with objective function coefficients ± 1.

4. Bilevel programs with convex and linear structures

In this section, we pay the main attention to bilevel programs with certain *convex* structures of either lower-level problems or both lower-level and upper-level ones. The results obtained are significantly different for *smooth* and *nonsmooth* problems – and not only from the viewpoint of differentiation. They are also generally different from the optimality conditions for smooth bilevel programs derived in section 3 and for those governed by Lipschitzian bilevel problems that are considered in section 5. We discuss remarkable specifications of the major results in the case of *linear programming* problems at the lower level.

Let us start with *fully convex* problems in (3.1)–(3.5) defined generally by nonsmooth functions that are convex jointly with respect to *all* their variables. To proceed, we first modify lower-level and upper-level regularity requirements formulated in (3.11) and (3.13) for smooth problems. This can be naturally done by replacing the Mangasarian–Fromovitz type constraint qualifications for smooth problems as in (3.11) and (3.13) with their easier *Slater* counterparts well-recognized in convex programming.

Following this way, we say that the convex bilevel program is *lower-level regular* if for each $x\in\mathbb{R}^n$ with $K(x)\neq\emptyset$ in (3.4) there is $y_x\in\mathbb{R}^m$ such that

$$g_i(x,y_x)<0\ \text{ whenever }\ i=1,\dots,p. \tag{4.1}$$

Similarly, we say that the bilevel program is *upper-level regular* if there is $\tilde{x}\in\mathbb{R}^n$ with $\Psi(\tilde{x})\neq\emptyset$ satisfying the conditions

$$G_j(\tilde{x})<0\ \text{ for all }\ j=1,\dots,k. \tag{4.2}$$

It is well known that for smooth convex problems the regularity conditions (4.1) and (4.2) imply those in (3.11) and (3.13), respectively.

THEOREM 4.1 (necessary conditions for fully convex nonsmooth bilevel programs) *Let* $(\bar{x}, \bar{y})$ *be an optimal solution to the optimistic bilevel program* (3.1)–(3.5), *where all the cost and constraint functions are convex with respect to their variables. Assume that the bilevel program under consideration is partially calm at* $(\bar{x}, \bar{y})$, *and that it is both lower-level and upper-level regular. Suppose also that the argminimum map* Ψ *is inner semicompact at* $\bar{x}$, *which is automatic when* Ψ *is uniformly bounded around this point. Then there exist* $\lambda > 0$, $(\mu_1, \ldots, \mu_p) \in \mathbb{R}^p$ *satisfying* (3.18), $(\alpha_1, \ldots, \alpha_k) \in \mathbb{R}^k$ *satisfying* (3.20), *and also* $\tilde{y} \in \Psi(\bar{x})$ *and* $(\lambda_1, \ldots, \lambda_p) \in \mathbb{R}^p$ *such that the following relationships hold*:

$$0 \in \partial_x F(\bar{x}, \bar{y}) + \lambda\Big(\partial_x f(\bar{x}, \bar{y}) - \partial_x f(\bar{x}, \tilde{y})\Big) + \sum_{i=1}^{p} \mu_i \partial_x g(\bar{x}, \bar{y})$$
$$- \lambda \sum_{i=1}^{p} \lambda_i \partial_x g(\bar{x}, \tilde{y}) + \sum_{j=1}^{k} \alpha_j \partial G_j(\bar{x}), \tag{4.3}$$

$$0 \in \partial_y F(\bar{x}, \bar{y}) + \lambda \partial_y f(\bar{x}, \bar{y}) + \sum_{i=1}^{p} \mu_i \partial_y g_i(\bar{x}, \bar{y}), \tag{4.4}$$

$$0 \in \partial_y f(\bar{x}, \tilde{y}) + \sum_{i=1}^{p} \lambda_i \partial_y g(\bar{x}, \tilde{y}), \tag{4.5}$$

$$\lambda_i \geq 0, \quad \lambda_i g_i(\bar{x}, \tilde{y}) = 0 \ \text{ for all } \ i = 1, \ldots, p, \tag{4.6}$$

where ∂, ∂_x, and ∂_y stand, respectively, for the full and partial subdifferentials of convex analysis. If, moreover, Ψ is assumed to be inner semicontinuous at $(\bar{x}, \bar{y})$, then we can put $\tilde{y} = \bar{y}$ in the relationships (4.3), (4.5), and (4.6).

Proof Following the proof of Theorem 3.1, we conclude that $(\bar{x}, \bar{y})$ is an optimal solution to the single-level problem of mathematical programming (3.21) with the marginal function φ defined in (3.22). It is easy to check that the convexity of f and g_i implies the convexity of φ. Since the functions f and g_i are Lipschitz-continuous around the reference points (due to their convexity and continuity) and by the lower-level regularity (4.1), the value function φ is *locally Lipschitzian* around $\bar{x}$; this is a well-known fact that follows, e.g., from an essentially more general result of [20, Theorem 5.1] combined with [18, Corollary 4.43].

Using the partial calmness condition at $(\bar{x}, \bar{y})$, we conclude as in the proof of Theorem 3.1 that $(\bar{x}, \bar{y})$ is a local optimal solution to the penalized (by $\lambda > 0$) problem (3.21), which is *Lipschitzian while nonconvex*, since the cost function $F + \lambda(f - \varphi)$ therein is a *difference of convex functions*. Applying to (3.23) the generalized multiplier rule from [19, Theorem 3.21(iii)] and then the (basic) subdifferential sum rule for Lipschitzian functions given in [18, Theorem 2.23(c)], we arrive at the inclusion

$$0 \in \lambda_0 \partial F(\bar{x}, \bar{y}) + \lambda_0 \lambda \partial f(\bar{x}, \bar{y}) + \big(\lambda_0 \lambda \partial(-\varphi)(\bar{x}), 0\big)$$
$$+ \sum_{i=1}^{p} \mu_i \partial g_i(\bar{x}, \bar{y}) + \sum_{j=1}^{k} \alpha_j \big(\partial G_j(\bar{x}), 0\big), \tag{4.7}$$

where all the multipliers $\lambda_0, \mu_i, \alpha_j$ are nonnegative and not equal to zero simultaneously, with the complementary slackness conditions (3.18) and (3.20) for μ_i and α_j. Observe that in (4.7) and in what follows we use the same symbol for the basic subdifferential and the subdifferential of convex analysis, since they agree for convex functions.

As in standard convex programming (actually by definition of the subdifferential in convex analysis), we easily conclude that the combination of the lower-level and upper-level regularity conditions of the Slater type in (4.1) and (4.2) with the sign and complementary slackness conditions from (3.18) and (3.20), imply that $\lambda_0 \neq 0$ in (4.7). Thus, we arrive at the KKT-type inclusion

$$\begin{aligned} 0 \in \partial F(\bar{x}, \bar{y}) + \lambda \partial f(\bar{x}, \bar{y}) + \big(\lambda \partial(-\varphi)(\bar{x}), 0\big) \\ + \sum_{i=1}^{p} \mu_i \partial g_i(\bar{x}, \bar{y}) + \sum_{j=1}^{k} \big(\alpha_j \partial G_j(\bar{x}), 0\big). \end{aligned} \tag{4.8}$$

Then we use the following important relationship between the *full* and *partial* subdifferentials of *convex continuous* functions $\psi(x, y)$ that holds by e.g. [18, Corollary 3.44]:

$$\partial \psi(\bar{x}, \bar{y}) \subset \partial_x \psi(\bar{x}, \bar{y}) \times \partial_y \psi(\bar{x}, \bar{y}). \tag{4.9}$$

Employing (4.9) in the KKT condition (4.8), we get the inclusions (4.4) and

$$\begin{aligned} 0 \in \partial_x F(\bar{x}, \bar{y}) + \lambda \partial_x f(\bar{x}, \bar{y}) + \lambda \partial(-\varphi)(\bar{x}) \\ + \sum_{i=1}^{p} \mu_i \partial_x g(\bar{x}, \bar{y}) + \sum_{j=1}^{k} \alpha_j \partial G(\bar{x}). \end{aligned} \tag{4.10}$$

Furthermore, the symmetric property (3.26) and the *convexity* of the subgradient set $\partial\varphi(\bar{x})$ – since φ is convex – allow us to conclude that

$$\partial(-\varphi)(\bar{x}) \subset -\partial\varphi(\bar{x}). \tag{4.11}$$

The next step is to efficiently estimate the subdifferential $\partial\varphi(\bar{x})$ of the value function in the convex setting under consideration. Applying the general result for basic subgradients of marginal functions from [22, Theorem 8] (actually its *inner semicompact* counterpart) to our value function φ given in (3.22) and using the *decomposition property* (4.9), we get

$$\partial\varphi(\bar{x}) \subset \bigcup_{y \in \Psi(\bar{x})} \Big[\bigcup_{(\lambda_1, \ldots, \lambda_p) \in \Lambda(\bar{x}, y)} \Big\{ \partial_x f(\bar{x}, y) + \sum_{i=1}^{p} \lambda_i \partial_x g_i(\bar{x}, y) \Big\} \Big] \tag{4.12}$$

via the union over the argminimum set $\Psi(\bar{x})$ from (3.2) and the set of multipliers $(\lambda_1, \ldots, \lambda_p) \in \Lambda(\bar{x}, y)$ defined now by

$$\begin{aligned} \Lambda(x, y) := \Big\{ (\lambda_1, \ldots, \lambda_p) \in \mathbb{R}^p \Big| 0 \in \partial_y f(x, y) + \sum_{i=1}^{p} \lambda_i \partial_y g_i(x, y), \\ \lambda_i \geq 0, \quad \lambda_i g_i(x, y) = 0, \quad i = 1, \ldots, p \Big\}. \end{aligned} \tag{4.13}$$

Combining (4.10)–(4.13), we arrive at the necessary optimality conditions (4.3)–(4.6), in addition to (3.18) and (3.20), and thus complete the proof of the theorem in the case when Ψ is *inner semicompact* at $\bar{x}$.

Of course, this gives the simplified relationships (4.3), (4.5), and (4.6) with $\tilde{y} = \bar{y}$ if $\Psi(\bar{x}) = \{\bar{y}\}$ is a singleton. However, the above proof (cf. also the proof of Theorem 3.1) allows us to arrive at the latter conclusion when the argminimum map Ψ is assumed to be *merely inner semicontinuous* at the solution pair $(\bar{x}, \bar{y})$. ■

Remark 4.1 (Dependence of optimality conditions on partial subgradients of nonsmooth lower-level cost functions). Observe that even in the case of $\Psi(\bar{x}) = \{\bar{y}\}$ (and in the general case of inner semicontinuity of Ψ) with a *nonsmooth* lower-level cost function f, we *loose* in (4.3) the remarkable *independence phenomenon* of smooth bilevel programs discussed in Remark 3.1. Indeed, one has

$$\partial_x f(\bar{x}, \bar{y}) - \partial_x f(\bar{x}, \bar{y}) = \{0\}$$

in (4.3) *if and only if* $\partial_x f(\bar{x}, \bar{y})$ is a singleton, i.e. f is *strictly differentiable* with respect to the parameter variable x at $(\bar{x}, \bar{y})$.

Remark 4.2 (Subdifferential regularity). As follows from the proof of Theorem 4.1, the approach developed and the results employed therein allow us to extend the necessary optimality conditions obtained in (4.3)–(4.6) to more general classes of functions – with appropriate modifications. Indeed, assuming that the functions F, f, g_i, G_j, and φ are *subdifferentially/lower regular* at the corresponding points (in the sense that their basic subdifferential agrees with the Fréchet one; see section 2), we have the two key facts used in the proof of Theorem 4.1: the *decomposition property* (4.9) by [18, Corollary 3.44] and the *convexity* of the basic subdifferential $\partial\varphi(\bar{x})$. Of course, the lower-level and upper-level regularity assumptions of the Slater type should be appropriately (subdifferentially) modified in the nonconvex setting; cf. Section 5. Note that the class of subdifferentially regular functions is much broader than the collections of convex and smooth ones; it includes, e.g., all *amenable* functions; see [18,29] for more details.

Next we consider the optimistic bilevel program in (3.1)–(3.5) with *smooth* and *Uconvex* structures simultaneously. It turns out that in this setting we are able to derive new results different from the corresponding specifications of both Theorem 3.1 and Theorem 4.1. The *most remarkable feature* of the new result is that we derive an optimality condition in form (4.3) for smooth functions, where we are able to put $\tilde{y} = \bar{y}$ *without any inner semicontinuity* assumption on the argminimum map Ψ as in Theorem 4.1.

THEOREM 4.2 (Necessary conditions for smooth bilevel programs with fully convex lower-level problems). *Let* $(\bar{x}, \bar{y})$ *be a local optimal solution to the bilevel program* (3.1)–(3.5), where all the functions F, G_j, f, g_i are continuously differentiable around $(\bar{x}, \bar{y})$ and $\bar{x}$, respectively, where the value function φ defined by (3.7) is finite around $\bar{x}$, and where the lower-level functions f and g_i are fully convex. Assume furthermore that the argminimum map Ψ is inner semicompact at $\bar{x}$, that the bilevel program under consideration is partially calm at $(\bar{x}, \bar{y})$, and that $(\bar{x}, \bar{y})$ is lower-level regular in the sense of (3.11) while $\bar{x}$ is upper-level regular in the sense of (3.13). Then there are multiplies $\lambda > 0$, $(\mu_1, \ldots, \mu_p) \in \mathbb{R}^p$ satisfying (3.18), $(\alpha_1, \ldots, \alpha_k) \in \mathbb{R}^k$ satisfying (3.20), and $(\lambda_1, \ldots, \lambda_p) \in \mathbb{R}^p$ such that we have equality (3.15) and the relationships

$$\nabla_x F(\bar{x}, \bar{y}) + \sum_{i=1}^{p} \big(\mu_i - \lambda\lambda_i\big) \nabla_x g_i(\bar{x}, \bar{y}) + \sum_{j=1}^{k} \alpha_j \nabla G_j(\bar{x}) = 0, \tag{4.14}$$

$$\nabla_y f(\bar{x}, \bar{y}) + \sum_{i=1}^{p} \lambda_i \nabla_y g_i(\bar{x}, \bar{y}) = 0, \tag{4.15}$$

$$\lambda_i \geq 0, \quad \lambda_i g_i(\bar{x}, \bar{y}) = 0 \ \text{ for all } \ i = 1, \ldots, p. \tag{4.16}$$

Proof Recall first a well-known fact of finite-dimensional convex analysis ensuring that a convex function is locally Lipschitzian around any point from the interior of its effective domain; see [28, Theorem 10.4]. Since the value function φ is convex for the fully convex lower-level problem under consideration, it is *locally Lipschitzian* around $\bar{x} \in \operatorname{int}(\operatorname{dom} \varphi)$. Following then the proof of Theorem 3.1 and taking into account inclusion (4.11) by the convexity of $\partial\varphi(\bar{x})$, we arrive at the relationships (3.15) and

$$\nabla_x F(\bar{x}, \bar{y}) + \lambda \nabla_x f(\bar{x}, \bar{y}) + \sum_{i=1}^{p} \mu_i \nabla_x g(\bar{x}, \bar{y}) + \sum_{j=1}^{k} \alpha_j \nabla G(\bar{x}) \in \lambda \partial\varphi(\bar{x}), \tag{4.17}$$

where μ_i and α_j satisfy (3.18) and (3.20), respectively. It remains to employ the following subdifferential formula for the optimal value function:

$$\partial\varphi(\bar{x}) = \bigcup_{(\lambda_1, \ldots, \lambda_p) \in \Lambda(\bar{x}, \bar{y})} \Big\{ \nabla_x f(\bar{x}, \bar{y}) + \sum_{i=1}^{p} \lambda_i \nabla_x g_i(\bar{x}, \bar{y}) \Big\}, \tag{4.18}$$

with the multiplier set $\Lambda(\bar{x}, \bar{y})$ computed in (3.28), established by Tanino and Ogawa [32] (see also [31, Theorem 6.6.7]) under the assumptions made in our theorem; note that the uniform boundedness requirement on Ψ around $\bar{x}$ is imposed in [31,32] while the proof therein works under the inner semicompactness of the argminimum map. Substituting finally (4.18) into (4.17), we complete the proof of the theorem. ∎

Remark (Related results) Observe that quite recently [35, Theorem 4.1], Ye has derived by a different method necessary optimality conditions in the form of Theorem 4.2 for smooth bilevel programs under the *concavity* assumption on the function $f(x, y) - \varphi(x)$ and various types of constraint qualifications. Let us also mention the related earlier result of [34, Corollary 4.1] for the case of *jointly linear* cost and constraint functions in lower-level problems of optimistic bilevel programming.

The results obtained in Theorem 3.1 and Theorem 4.2 can be easily specified for the case of *parametric linear programs* at the lower level, provided that parameters enter *additively* in constraint and cost functions – in order to keep full convexity. Note that such problems are *always partially calm* at solution points; see [37, Proposition 5.1] in connection with [5, Theorem 3.5]. Observe also that our necessary optimality conditions for bilevel programs with linear lower-level problems – which are directly implied by Theorems 3.1 and 4.2 – are different from (significantly *simpler* than) the corresponding results presented in [37].

Let us discuss some specific features of optimality conditions in bilevel programming with linear lower-level problems.

Example 4.4 (Linear lower-level problems with additive perturbation). Consider the bilevel program (3.1)–(3.5) with the lower-level problem (3.3) given by

$$\text{minimize } \; c^\top y \; \text{ subject to } \; Ay \le x, \tag{4.19}$$

where the vector $c \in \mathbb{R}^n$ and the matrix $A \in \mathbb{R}^{n \times m}$ are fixed. A specific feature of (4.19) is that the perturbation parameter x enters only the *right-hand side* of the constraint system. Of course, the lower-level problem (4.19) is *fully convex* in (x, y). Given now an optimal solution $(\bar{x}, \bar{y})$ to the bilevel problem (3.1) with

$$\Psi(x) := \operatorname*{Argmin}_y \big\{ c^\top y \,|\, Ay \le x \big\} \tag{4.20}$$

and the set X defined in (3.5), we apply the necessary optimality conditions from Theorem 4.2 and get the following relationships:

$$\begin{aligned} \nabla_x F(\bar{x}, \bar{y}) - \sum_{i=1}^{p} (\mu_i - \lambda \beta_i) + \sum_{j=1}^{k} \alpha_j \nabla G_j(\bar{x}) = 0 \\ \nabla_y F(\bar{x}, \bar{y}) + \lambda c + \sum_{i=1}^{p} \mu_i a^i = 0 \\ c + \beta^\top A = 0 \\ \beta \ge 0, \quad \beta^\top (A\bar{y} - \bar{x}) = 0, \quad \mu \ge 0, \quad \mu^\top (A\bar{y} - \bar{x}) = 0 \\ \alpha \ge 0, \quad \alpha^\top G(\bar{x}) = 0, \end{aligned}$$

where $\lambda > 0$, $\beta = (\beta_1, \ldots, \beta_p) \in \mathbb{R}^p$, $\mu = (\mu_1, \ldots, \mu_p) \in \mathbb{R}^p$, $\alpha = (\alpha_1, \ldots, \alpha_k) \in \mathbb{R}^k$, and where a^i stands for the i-th row of the matrix A.

Remark 4.5 (Inner semicontinuous solutions maps in linear programming). It is worth indicating that the argminimum map Ψ in the lower-level linear programming problem from Example 4.4 happens to be *inner/lower semicontinuous* at the solution point $\bar{x}$ by Bank *et al.* [2, Theorem 4.3.5]. Note that this phenomenon is *lost* for more general perturbations of linear programs at the lower level, in particular, for argminimum maps in one of the forms:

$$\Psi(x) = \operatorname*{Argmin}_y \big\{ x^\top y \,|\, Ay \le b \big\} \quad \text{or} \quad \Psi(x) = \operatorname*{Argmin}_y \big\{ x^\top y \,|\, A(x)y \le b \big\},$$

which do not correspond to *fully convex* lower-level problems in (3.1–3.5). In such cases, we need to use necessary optimality conditions of type (3.33) discussed in Remark 3.4.

5. Lipschitzian bilevel programming

In the concluding section of the article, we derive necessary optimality conditions for the optimistic bilevel program described in (3.1)–(3.5), where all the functions in both lower-level and upper-level problems are *Lipschitz continuous* around the local optimal solution under consideration. The results obtained subsequently partly extend the corresponding optimality conditions established in sections 3 and 4 for smooth and fully convex bilevel programs, while certain important features of the above results are due to the specifics of smooth and convex problems and do not have any analogs in the general nonsmooth setting. The main tool of our analysis of Lipschitzian bilevel programs is the *basic subdifferential* (2.5) of locally Lipschitz functions, which enjoys *full calculus*.

To proceed, we need to formulate appropriate lower-level and upper-level regularity/qualification conditions. Given a point $(\bar{x}, \bar{y})$ satisfying the lower-level inequality constraints (3.4) with the index set $I(\bar{x}, \bar{y})$ from (3.10), we say that $(\bar{x}, \bar{y}) \in \mathbb{R}^n \times \mathbb{R}^m$ is *lower-level regular* if the following implication holds in terms of the basic subdifferential:

$$\left[\sum_{i \in I(\bar{x}, \bar{y})} \lambda_i v_i = 0,\ \lambda_i \geq 0\right] \Longrightarrow \left[\lambda_i = 0 \ \text{ for all } \ i \in I(\bar{x}, \bar{y})\right] \tag{5.1}$$

whenever $(u_i, v_i) \in \partial g_i(\bar{x}, \bar{y})$ with some $u_i \in \mathbb{R}^n$ as $i \in I(\bar{x}, \bar{y})$.

Similarly, given $\bar{x} \in \mathbb{R}^n$ satisfying the upper-level inequality constraints (3.5) with the index set $J(\bar{x})$ from (3.12), we say that $\bar{x}$ is *upper-level regular* if

$$\left[0 \in \sum_{j \in J(\bar{x})} \lambda_j \partial G_j(\bar{x}),\ \lambda_j \geq 0\right] \Longrightarrow \left[\lambda_j = 0 \ \text{ whenever } \ j \in J(\bar{x})\right]. \tag{5.2}$$

Observe that these regularity conditions developed in [18, section 4.3] are *basic nonsmooth counterparts* of the classical Mangasarian–Fromovitz constraint qualifications for the lower-level and upper-level problems, respectively. For problems with smooth data, they reduce to the lower/upper-level regularity conditions (3.11) and (3.13) used in section 3.

The next theorem presents two versions of necessary optimality conditions for bilevel programs with locally Lipschitzian data. The main difference between these mutually independent versions is in the *inner semicontinuity versus inner semicompactness* assumptions on the argminimum map (3.2); see Section 3 for more discussions.

THEOREM 5.1 (Necessary conditions for Lipschitzian bilevel programs). *Let $(\bar{x}, \bar{y})$ be a local optimal solution to the optimistic bilevel program* (3.1)–(3.5), *which is partially calm at this point. Suppose that the upper-level functions f and G_j, $j = 1, \ldots, k$, are locally Lipschitzian around $(\bar{x}, \bar{y})$ and $\bar{x}$, respectively, and that $\bar{x}$ is upper-level regular. The following assertions hold*:

(i) *Assume that the argminimum map Ψ is inner semicontinuous at $(\bar{x}, \bar{y})$, that the pair $(\bar{x}, \bar{y})$ is lower-level regular, and that the lower-level functions f and g_i, $i = 1, \ldots, p$, are locally Lipschitzian around $(\bar{x}, \bar{y})$. Then there are $\lambda > 0$, $(\mu_1, \ldots, \mu_p) \in \mathbb{R}^p$ satisfying* (3.18), $(\eta_1, \ldots, \eta_{n+1}) \in \mathbb{R}^{n+1}$ *satisfying* (3.19), $(\alpha_1, \ldots, \alpha_k) \in \mathbb{R}^k$ *satisfying* (3.20), λ_{is} *with* $i = 1, \ldots, p$ *and* $s = 1, \ldots, n+1$ *satisfying* (3.17), *and* $(u_1, \ldots, u_{n+1}) \in \mathbb{R}^{n \times (n+1)}$, *such that*

$$(u_s, 0) \in \partial f(\bar{x}, \bar{y}) + \sum_{i=1}^{p} \lambda_{is} \partial g_i(\bar{x}, \bar{y}) \quad \textit{for all } s = 1, \ldots, n+1, \tag{5.3}$$

$$\left(\lambda \sum_{s=1}^{n+1} \eta_s u_s, 0\right) \in \partial F(\bar{x}, \bar{y}) + \lambda \partial f(\bar{x}, \bar{y}) + \sum_{i=1}^{p} \mu_i \partial g_i(\bar{x}, \bar{y}) + \left(\sum_{j=1}^{k} \alpha_j \partial G_j(\bar{x}), 0\right). \tag{5.4}$$

(ii) *Assume that the argminimum map Ψ is inner semicompact at $\bar{x}$, that for each vector $y \in \Psi(\bar{x})$ the pair $(\bar{x}, y)$ is lower-level regular and the lower-level functions f and g_i are locally Lipschitzian around $(\bar{x}, y)$. Then there are $\lambda > 0$, $(\mu_1, \ldots, \mu_p) \in \mathbb{R}^p$ satisfying* (3.18), $(\eta_1, \ldots, \eta_{n+1}) \in \mathbb{R}^{n+1}$ *satisfying* (3.19), $(\alpha_1, \ldots, \alpha_k) \in \mathbb{R}^k$ *satisfying* (3.20) *as well as $y_s \in \Psi(\bar{x})$, $u_s \in \mathbb{R}^n$, and λ_{is} with $i = 1, \ldots, p$ and $s = 1, \ldots, n+1$, such that one has*

$$\lambda_{is} \geq 0, \ \lambda_{is} g_i(\bar{x}, y_s) = 0 \quad \textit{for all } i = 1, \ldots, p \textit{ and } s = 1, \ldots, n+1, \tag{5.5}$$

$$(u_s, 0) \in \partial f(\bar{x}, y_s) + \sum_{i=1}^{p} \lambda_{is} \partial g_i(\bar{x}, y_s) \quad \textit{for all } s = 1, \ldots, n+1, \tag{5.6}$$

and relationship (5.4) *is satisfied.*

Proof Consider first case (i) when the argminimum map Ψ is supposed to be inner semicontinuous at $(\bar{x}, \bar{y})$. Proceeding as in the proof of Theorem 3.1, we come up – under the partial calmness assumption – to the penalized problem (3.23) for which $(\bar{x}, \bar{y})$ is a local optimal solution. Employing [18, Corollary 4.43], we conclude that the constraint mapping $K(\cdot)$ in (3.4) is *Lipschitz-like* around $(\bar{x}, \bar{y})$ under the assumed lower-level regularity, and thus the value function φ is *locally Lipschitzian* around $\bar{x}$ by [20, Theorem 5.2(i)]. Therefore, (3.23) is a single-level problem of *Lipschitzian programming*, and we have – by [19, Theorem 5.21(ii)] and the subdifferential sum rule from [19, Theorem 2.23] – the following necessary optimality conditions for $(\bar{x}, \bar{y})$ in (3.23): there are multipliers $(\lambda_0, \mu_1, \ldots, \mu_p, \alpha_1, \ldots, \alpha_k)$, not equal to zero simultaneously, such that $\lambda_0 \geq 0$, that μ_i and α_j satisfy the sign and complementarity slackness conditions formulated in (3.18) and (3.20), respectively, and that

$$\begin{aligned} 0 \in\ & \lambda_0 \partial F(\bar{x}, \bar{y}) + \lambda_0 \lambda \partial f(\bar{x}, \bar{y}) + \big(\lambda_0 \lambda \partial(-\varphi)(\bar{x}), 0\big) \\ & + \sum_{i=1}^{p} \mu_i \partial g_i(\bar{x}, \bar{y}) + \sum_{j=1}^{k} \alpha_j \big(\partial G_j(\bar{x}), 0\big) \end{aligned} \tag{5.7}$$

via the basic subdifferential (2.5) for the Lipschitzian functions in (5.7) with respect to all their variables. Furthermore, the combination of the lower-level regularity of $(\bar{x}, \bar{y})$ and the upper-level regularity of $\bar{x}$ implies that $\lambda_0 > 0$ in (5.7), and thus we arrive at

$$\begin{aligned} 0 \in\ & \partial F(\bar{x}, \bar{y}) + \lambda \partial f(\bar{x}, \bar{y}) + \big(\lambda \partial(-\varphi)(\bar{x}), 0\big) \\ & + \sum_{i=1}^{p} \mu_i \partial g_i(\bar{x}, \bar{y}) + \sum_{j=1}^{k} \big(\alpha_j \partial G_j(\bar{x}), 0\big). \end{aligned} \tag{5.8}$$

Employing now the subdifferential formula for the value function φ from [22, Theorem 8], we get – under the inner semicontinuity and the lower-level regularity assumption made – the following inclusion in the case of locally Lipschitzian functions in the lower-level problem:

$$\begin{aligned} \partial \varphi(\bar{x}) \subset \bigcup_{(\lambda_1, \ldots, \lambda_p)} \Big\{ u \in \mathbb{R}^n \Big| \ & (u, 0) \in \partial f(\bar{x}, \bar{y}) + \sum_{i=1}^{p} \lambda_i \partial g_i(\bar{x}, \bar{y}), \\ & \lambda_i \geq 0, \ \lambda_i g_i(\bar{x}, \bar{y}) = 0, \ i = 1, \ldots, p \Big\}. \end{aligned} \tag{5.9}$$

Now taking into account the symmetric property (3.26) and substituting (5.9) into (5.8), we arrive at relationships (5.3) and (5.4), with the convexification weights η_s satisfying (3.19), similarly to the proof of Theorem 3.1. This completes the proof of assertion (i).

To prove (ii) under the inner semicompactness assumption on Ψ, we argue in the same way as above by using [20, Theorem 5.2(ii)] for the justification of the local Lipschitzian property of the value function φ around $\bar{x}$ and by employing the following *inner semicompact* counterpart of [22, Theorem 8] instead of (5.9):

$$\partial\varphi(\bar{x}) \subset \bigcup_{y\in\Psi(\bar{x})}\Bigg[\bigcup_{(\lambda_1,\ldots,\lambda_p)}\Bigg\{u \in \mathbb{R}^n \Bigg| \, (u,0) \in \partial f(\bar{x},y) + \sum_{i=1}^{p}\lambda_i\partial g_i(\bar{x},y),$$
$$\lambda_i \geq 0,\ \lambda_i g_i(\bar{x},y) = 0,\ i = 1,\ldots,p\Bigg\}\Bigg]. \tag{5.10}$$

Then we can proceed similarly to the above arguments while distributing y over the argminimum set $\Psi(\bar{x})$ in the convex combination; cf. the proof of Theorem 3.1 and also Remark 3.4. This completes the proof of the theorem. ■

Observe that the proofs of Theorem 3.1, Theorem 3.1, and Theorem 4.1 go in the same direction, while the specifics of the smooth and fully convex settings in Theorem 3.1 and Theorem 4.1 allow us to use the *decomposition* of the full derivative/ subdifferential into the partial ones as – obviously – for smooth functions

$$\nabla\psi(\bar{x},\bar{y}) = \nabla_x\psi(\bar{x},\bar{y}) \times \nabla_y\psi(\bar{x},\bar{y})$$

and – nontrivially – for fully convex functions in (4.9). Furthermore, other specific features of smooth and fully convex functions make it possible to establish particular relationship in the necessary optimality conditions of Theorem 3.1 and Theorem 4.1, which do not have any analogs in the general nonsmooth setting of Theorem 5.1.

Remark 5.1 (Avoiding convex combinations in necessary optimality conditions). An underlying feature of the necessary optimality conditions in Theorem 3.1 and Theorem 5.1 is the presence of *convex combinations* of gradients and subgradients of the constraint functions g_i in the lower-level problem; in the results of section 4, such a convexification is not needed due to the automatic convexity. The necessary of this operation in the proofs of Theorem 3.1 and Theorem 5.1 comes from the inclusion (3.26) and the classical *plus–minus symmetry* of the *convexified* (Clarke) subdifferential of Lipschitzian functions, which is not the case for the basic subdifferential. On the other hand, such a convexification is *not needed at all*, if we directly apply to $f(x,y) - \varphi(x)$ in the aforementioned proofs the *difference rule*

$$\partial(\varphi_1 - \varphi_2)(\bar{x}) \subset \partial\varphi_1(\bar{x}) - \partial\varphi_2(\bar{x}) \tag{5.11}$$

for the basic subdifferential recently obtained in [21, Corollary 3.4]. However, the subdifferential difference rule (5.11) is derived in [21] under the assumption that the Fréchet subdifferential $\widehat{\partial}\varphi_2(x)$ is *nonempty* in a neighborhood of the reference point $\bar{x}$, which seems to be a restrictive assumption for the *value function* φ in the setting (3.7) under consideration. We are going to carefully investigate this issue in our further research.

Remark 5.2 (Nonsmooth constraints of the equality type). The necessary optimality conditions obtained in Theorem 5.1 for bilevel programs with inequality constraints can be extended to the case of additional *equality* constraints (3.30) and (3.31) governed by locally *Lipschitzian* functions at both the lower and upper levels. Indeed, the results from [18–20,22] used in the proof of Theorem 5.1 are established therein for *both* equality and inequality constraints; so we restrict ourselves to the case of inequality constraints just for *simplicity*. To be able to use the aforementioned results in the case of additional equality constraints (3.30) and (3.31), we need to replace the above definitions of lower-level and upper-level regularity by the following relationships. For the *lower-level regularity*: implication (5.1) holds whenever

$$(u_i, v_i) \in \partial g_i(\bar{x}, \bar{y}) \ \text{ for } \ i \in \{1, \dots, p\} \cap I(\bar{x}, \bar{y}) \quad \text{and}$$
$$(u_i, v_i) \in \partial h_i(\bar{x}, \bar{y}) \cup \partial(-h_i)(\bar{x}, \bar{y}) \ \text{ for } \ i = p+1, \dots, p+q$$

with $u_i \in \mathbb{R}^n$ as $i \in I(\bar{x}, \bar{y})$, where the index set (3.10) is extended now to

$$I(\bar{x}, \bar{y}) := \big\{ i \in \{1, \dots, p\} \,|\, g_i(\bar{x}, \bar{y}) = 0 \big\} \cup \{p+1, \dots, p+q\}.$$

For the *upper-level regularity*: instead of (5.2) we require that

$$\Big[0 \in \sum_{j \in J(\bar{x})} \lambda_j \partial G_j(\bar{x}) + \sum_{j=k+1}^{k+l} \lambda_j (\partial H_j(\bar{x}) \cup \partial(-H_j))(\bar{x}),$$
$$\lambda_j \geq 0, \ j \in J(\bar{x}) \cup \{k+1, \dots, k+l\} \Big]$$
$$\Longrightarrow \Big[\lambda_j = 0 \ \text{ whenever } \ j \in J(\bar{x}) \cup \{k+1, \dots, k+l\} \Big],$$

here $J(\bar{x})$ is defined in (3.12); see [18, section 4.3] for more discussions. The corresponding subdifferential modifications for equality constraints, involving the terms

$$\lambda_i \big[\partial h_i(\bar{x}, \bar{y}) \cup \partial\big(-h_i\big)(\bar{x}, \bar{y}) \big] \quad \text{with } \lambda_i \geq 0, \ i = p+1, \dots, p+q,$$

apply to the subdifferential formula of [22, Theorem 8] for the value function in the lower-level problem with both equality and inequality constraints. This allows us to proceed accordingly in the proof of Theorem 5.1 for the case of general constraints.

Remark 5.3 (Infinite-dimensional extensions). The necessary optimality conditions derived in Theorem 5.1 and their specifications are satisfied *with no change* for bilevel programs in *infinite dimensions* defined on the class of *Asplund spaces* (for both decision and parameter variables), which can be equivalently described as Banach spaces whose separable subspaces have separable duals. This class is sufficiently broad particularly including every *reflexive* space; see e.g. [18] for more details and references. The above observation follows from the fact that *all* the results applied in the proof of Theorem 5.1 hold true in the case of Asplund spaces. Note that the assumptions imposed in Theorem 5.1 ensure not only the fulfillment of the required generalized differential calculus rules but also the validity of the so-called *sequential normal compactness* properties (and calculus rules for them), which are automatic in finite dimensions while playing a crucial role in infinite-dimensional variational analysis; see [18,19]. It is worth mentioning that the inner semicompactness assumption – as formulated in section 2 – stays with no change in the infinite-dimensional version of Theorem 5.1, but it does not follow anymore from the the uniform boundedness.

Acknowledgements

We are indebted to three anonymous referees for their insightful remarks that led us to the significant improvement of the original presentation. We thank Alexey Izmailov, Nguyen Mau Nam, Danny Ralph, and Stefan Scholtes for fruitful discussions about the partial calmness condition resulting in a deeper understanding of this important assumption. We also gratefully acknowledge various suggestions and remarks by Franco Giannessi and Jiři Outrata used in this article. Research of B. S. Mordukhovich was partly supported by the US National Science Foundation under grants DMS-0304989 and DMS-0603846 and by the Australian Research Council under grant DP-0451168. Finally, we thank Regina Burachik, Editor of this special issue, for her efficient handling of our article.

References

[1] Babbahadda, H. and Gadhi, N., 2006, Necessary optimality conditions for bilevel optimization problems using convexificators. *Journal of Global Optimization*, **34**, 535–549.
[2] Bank, B., Guddat, J., Klatte, D., Kummer, B. and Tammer, K., 1982, *Non-Linear Parametric Optimization* (Basel: Birkhäuser).
[3] Bao, T.Q., Gupta, P. and Mordukhovich, B.S., 2007, Necessary conditions in multiobjective optimization with equilibrium constraints. *Journal of Optimization Theory and Application*, to appear.
[4] Burke, J. V. and Deng, S., 2002, Weak sharp minima revisited, I: Basic theory. *Control Cyberneties*, **31**, 439–469.
[5] Burke, J.V. and Ferris, M.C., 1993, Weak sharp minima in mathematical programming. *SIAM Journal of Control and Optimization*, **31**, 1340–1359.
[6] Dempe, S., 1992, A necessary and a sufficient optimality condition for bilevel programming problems. *Optimization*, **25**, 341–354.
[7] Dempe, S., 2002, *Foundations of Bilevel Programming* (Dordrecht, The Netherlands: Kluwer Academic Publishers).
[8] Dempe, S., Kalashnikov, V. and Kalashnykova, N., 2006, Optimality conditions for bilevel programming problems. In: S. Dempe and V. Kalashnikov (Eds) *Optimization with Multivalued Mappings: Theory, Applications and Algorithms* (Berlin: Springer-Verlag), pp. 3–28.
[9] Dempe, S. and Lohse, S., 2006, Inverse linear programming. In: A. Seeger (Ed.) *RecentAdvances in Optimization*, Lecture Notes in Economic and Mathematical System, Vol. 563 (Berlin: Springer-Verlag), pp. 19–28.
[10] Dutta, J. and Dempe, S., 2006, Bilevel programming with convex lower-level problems. In: S. Dempe and V. Kalashnikov (Eds) *Optimization with Multivalued Mappings: Theory, Applications and Algorithms* (Berlin: Springer-Verlag).
[11] Gauvin, J. and Dubeau, F., 1982, Differential properties of the marginal function in mathematical programming. *Mathematical Programming Study*, **19**, 101–119.
[12] Guddat, J., Guerra Vasquez, F. and Jongen, H.T., 1990, *Parametric Optimization: Singularities, Pathfollowing and Jumps* (Stuttgart, Chichester: John Wiley & Sons, B.G. Teubner).
[13] Levy, A.B. and Mordukhovich, B.S., 2004, Coderivative analysis in parametric optimization. *Mathematical Programming*, **99**, 311–327.
[14] Luo, Z.Q., Pang, J.-S. and Ralph, D., 1996, *Mathematical Programs with Equilibrium Constraints* (Cambridge, UK: Cambridge University Press).
[15] Mangasarian, O.L. and Meyer, R.R., 1979, Nonlinear perturbations of linear programs. *SIAM Journal of Control and Optimization*, **17**, 745–752.
[16] Mordukhovich, B.S., 1976, Maximum principle in problems of time optimal control with nonsmooth constraints. *Journal of Applied Mathematics and Mechanism*, **40**, 960–969.
[17] Mordukhovich, B.S., 2004, Necessary conditions in nonsmooth minimization via lower and upper subgradients. *Set-Valued Analysis*, **12**, 163–193.
[18] Mordukhovich, B.S., 2006, Variational Analysis and Generalized Differentiation, I: Basic Theory. *Grundlehren Series (Fundamental Principles of Mathematical Sciences)*, Vol. 330 (Berlin: Springer-Verlag).
[19] Mordukhovich, B.S., 2006, Variational Analysis and Generalized Differentiation, II: Applications. *Grundlehren Series (Fundamental Principles of Mathematical Sciences)*, Vol. 331 (Berlin: Springer-Verlag).

[20] Mordukhovich, B.S. and Nam, N.M., 2005, Variational stability and marginal functions via generalized differentiation. *Mathematics of Operations Research*, **30**, 800–816.
[21] Mordukhovich, B.S., Nam, N.M. and Yen, N.D., 2006, Fréchet subdifferential calculus and optimality conditions in nondifferentiable programming. *Optimization*, **55**, 685–708.
[22] Mordukhovich, B.S., Nam, N.M. and Yen, N.D. 2007, Subgradients of marginal functions in parametric mathematical programming. *Mathematical Programming*, to appear.
[23] Outrata, J.V., 1990, On the numerical solution of a class of Stackelberg problems. *ZOR–Methods and Models of Operations Research*, **34**, 255–277.
[24] Outrata, J.V., 2000, A generalized mathematical program with equilibrium constraints. *SIAM Journal of Control and Optimization*, **38**, 1623–1638.
[25] Outrata, J.V. 2006. Personal communication.
[26] Outrata, J.V., Kočvara, M. and Zowe, J., 1998, *Nonsmooth Approach to Optimization Problems with Equilibrium Constraints* (Dordrecht, The Netherlands: Kluwer Academic Publishers).
[27] Robinson, S.M., 1982, Generalized equations and their solutions, part II: Applications to nonlinear programming. *Mathematical Programming Study*, **19**, 200–221.
[28] Rockafellar, R.T., 1970, *Convex Analysis* (Princeton, New Jersey: Princeton University Press).
[29] Rockafellar, R.T. and Wets, R.J.-B., 1998, Variational Analysis. *Grundlehren Series (Fundamental Principles of Mathematical Sciences)* (Berlin: Springer-Verlag).
[30] Scholtes, S., 1994, Introduction to piecewise differentiable equations. *Technical report No. 53/1994. Universität Karlsruhe, Institut für Statistik und Mathematische Wirtschaftstheorie* (Karlsruhe, Germany).
[31] Shimizu, K., Ishizuka, Y. and Bard, J.F., 1997, *Nondifferentiable and Two-Level Mathematical Programming* (Dordrecht: Kluwer Academic Publishers).
[32] Tanino, T. and Ogawa, T., 1984, An algorithm for solving two-level convex optimization problems. *International Journal of Systems Science*, **15**, 163–174.
[33] Ye, J.J., 1998, New uniform parametric error bounds. *Journal of Optimization Theory and Applications*, **98**, 197–219.
[34] Ye, J.J., 2004, Nondifferentiable multiplier rules for optimization and bilevel optimization problems. *SIAM Journal on Optimization*, **15**, 252–274.
[35] Ye, J.J. 2006. Constraint qualifications and KKT conditions for bilevel programming problems. *Mathematics of Operations Research*, **31**, 811–824.
[36] Ye, J.J. and Ye, X.Y., 1997, Necessary optimality conditions for optimization problems with variational inequality constraints. *Mathematics of Operations Research*, **22**, 977–997.
[37] Ye, J.J. and Zhu, D.L., 1995, Optimality conditions for bilevel programming problems. *Optimization*, **33**, 9–27.
[38] Ye, J.J. and Zhu, D.L., 1997, A note on optimality conditions for bilevel programming problems. *Optimization*, **39**, 361–366.
[39] Zhang, R., 1994, Problems of hierarchical optimization in finite dimensions. *SIAM Journal on Optimization*, **4**, 521–536.
[40] Zlobec, S., 2001, Stable Parametric Programming. *Applied Optimization*, Vol. 57 (Dordrecht: Kluwer Academic Publishers).