TWO-LEVEL VALUE FUNCTION APPROACH TO NONSMOOTH OPTIMISTIC AND PESSIMISTIC BILEVEL PROGRAMS

STEPHAN DEMPE, BORIS S. MORDUKHOVICH, AND ALAIN B. ZEMKOHO

The paper is dedicated to Jiří Outrata in honor of his 70th birthday

ABSTRACT. The authors' paper in *Optimization 63 (2014), 505–533*, see Ref. [5], was the first one to provide detailed optimality conditions for pessimistic bilevel optimization. The results there were based on the concept of the *two-level optimal value function* introduced and analyzed in *SIAM J. Optim. 22 (2012), 1309–1343*; see Ref. [4], for the case of optimistic bilevel programs. One of the basic assumptions in both of these papers is that the functions involved in the problems are at least continuously differentiable. Motivated by the fact that many real-world applications of optimization involve functions that are nondifferentiable at some points of their domain, the main goal of the current paper is extending the two-level value function approach to deriving new necessary optimality conditions for both optimistic and pessimistic versions in bilevel programming with nonsmooth data.

CONTENTS

1. INTRODUCTION

The original framework of *bilevel programming*, or *bilevel optimization*, goes back to von Stackelberg [23] and can be written in the form

$$\min_{x \in X} F(x,y) \text{ subject to } y \in S(x) := \arg \min_{y \in K(x)} f(x,y), \qquad (1.1)$$

where the leader is only in control of the upper level variable x, while the follower controls the lower level variable y. Here $F \colon \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ and $f \colon \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ are called the *upper-level* and *lower-level* objective functions, respectively, X denotes the upper-level feasible set, while the set-valued mapping $K \colon \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ describes the lower-level feasible set. Note nonetheless that the main focus in the literature (as discussed, i.e., in [2] and [4]), has been not on the original version (1.1) but on its following conventional simplification:

$$\min_{x,y} F(x,y) \ \text{ subject to } \ x \in X, \ y \in S(x), \tag{1.2}$$

were the leader appears to be in control of both variables. The latter formulation is *globally* equivalent to the following *optimistic* bilevel optimization problem, which is denoted by (P_o) and is written in the form

$$\min_{x \in X} \ \varphi_o(x) := \min_y \{ F(x,y) \mid y \in S(x) \}, \tag{1.3}$$

which indicates a certain cooperation between the upper and lower level players. However, (1.2) and (1.3) are *not* equivalent *locally* as demonstrated and discussed in [4]. To this end, observe also that the leader simply may not be able to cooperate with the follower in practice. To limit damages that may result from negative choices of the follower, the leader needs to solve the following *pessimistic* bilevel program denoted by (P_p):

$$\min_{x \in X} \ \varphi_p(x) := \max_y \{ F(x,y) \mid y \in S(x) \}, \tag{1.4}$$

where the *min* operation on the right-hand side of (1.3) is replaced by the *max* one; both minimum and maximum are achieved in (1.4) and in what follows under the assumptions made. It has been well recognized that problems (P_o) and (P_p) are essentially different from each other; see, e.g., [2] for more discussions. In fact, problem (P_p) belongs the most challenging class in bilevel optimization. This is the reason why very little has been known in the literature for solving this problem. There are few publications (see, e.g., [1, 24]) that propose global optimization techniques to solve (P_p). Taking into account serious limitations in global optimization, it is important to develop local optimization methods to solve pessimistic bilevel models. Pursuing this goal, the authors developed in [5], for the first time in the literature, detailed necessary conditions applied to local optimal solutions of pessimistic bilevel programs. Our approach there is based on the study and application of subdifferential properties of the *two-level value function*

$$\varphi_o(x) := \min_y \{ F(x,y) \mid y \in S(x) \}, \quad x \in \mathbb{R}^n, \tag{1.5}$$

defined in [4] and its symmetric *maximizing* version considered in [5] in the form

$$\varphi_p(x) := \max_y \{ F(x,y) \mid y \in S(x) \}, \quad x \in \mathbb{R}^n. \tag{1.6}$$

We refer the reader to [4, 5] for more details and the reasons behind the names of (1.5) and (1.6). As shown in [4], the subdifferential and related Lipschitz continuity properties of the function φ_o are instrumental for deriving optimality and stationarity conditions for optimistic bilevel programs while the form (1.6) is crucial to obtain similar results for the pessimistic model with smooth data.

Note that the two-level value functions (1.5) and (1.6) are clearly different from the *lower-level optimal value function* defined by

$$\varphi(x) := \min_y \left\{ f(x,y) \middle| y \in K(x) \right\} \tag{1.7}$$

that is widely spread in variational analysis under the name of "marginal function." Jiří Outrata was the first [20] to use (1.7) in the framework of optimistic bilevel programming via the so-called *value function approach* introduced by him and then largely developed in the literature; see, e.g., [2, 3, 14, 16, 25] and the references therein. Observe also that both functions (1.5) and (1.7) are *intrinsically nonsmooth* even when the initial data of problems (1.3) and (1.4) are infinitely differentiable.

Developing here the two-level value function approach to bilevel optimization, recall that in both papers [4, 5] we assumed that the functions involved in the descriptions of optimistic and pessimistic problems are at least *continuously differentiable*. As discussed, e.g., in [7], many real-world applications of bilevel optimization are described by nondifferentiable functions. In order to widen the scope of applications of the results in [4, 5], we extend them in this paper to functions that are *not necessarily differentiable*. Before proceeding with optimality conditions for (P_o) and (P_p) in the nonsmooth settings of Section 4 and Section 5, respectively, we derive the corresponding sensitivity results for the two-level value functions φ_o in Section 3 and its symmetric maximizing version φ_p in Section 5. The concluding Section 6 contains some discussions on further research.

The tools of variational analysis and generalized differentiation needed in what follows are summarized in the next section. Throughout the paper we use the standard notation and terminology in the area; see, e.g., [13, 22].

2. Preliminaries from Variational Analysis

In this section we recall and briefly overview for the reader's convenience some standard tools of generalized differentiation, which are broadly used in the subsequent variational analysis of both optimistic and pessimistic bilevel programs. More details can be found in the books [13, 22] and their bibliographies.

Starting with sets, consider $\Theta \subset \mathbb{R}^n$ locally closed around $\bar{x} \in \Theta$ and define the (basic, limiting, Mordukhovich) *normal cone* to Θ at $\bar{x}$ by

$$N_\Theta(\bar{x}) := \Big\{ v \in \mathbb{R}^n \Big| \ \exists\, x_k \to \bar{x},\ w_k \in \Pi_\Theta(x_k),\ \alpha \geq 0 \\ \text{such that}\ \alpha(x_k - w_k) \to v\ \text{as}\ k \to \infty \Big\}, \tag{2.1}$$

where $\Pi_\Theta(x)$ stands for the *Euclidean projector* of x onto Θ. There are various equivalent descriptions of (2.1), which are not used in this paper. Note that the construction (2.1) is *robust* (i.e., the set-valued mapping $x \mapsto N_\Theta(x)$ has closed graph), while $N_\Theta(\bar{x})$ is often *nonconvex* for nonconvex sets Θ as, e.g., for $\Theta := \mathrm{gph}\,|x| \subset \mathbb{R}^2$, the graph of the simplest nonsmooth convex function, at $\bar{x} = 0$. Nevertheless, the normal cone (2.1) and associated with it subdifferential and coderivative notions for extended-real-valued functions and set-valued mappings/multifunctions enjoy *full calculus* based on the *variational* and *extremal principles* of variational analysis.

The convex closure of the normal cone (2.1) denoted by

$$\bar{N}_\Theta(\bar{x}) := \mathrm{cl\,co}\, N_\Theta(\bar{x}), \quad \bar{x} \in \Theta, \tag{2.2}$$

is known as the (Clarke) *convexified normal cone* to Θ at $\bar{x}$. Note that the convexification in (2.2) may dramatically enlarge the sets of normals (e.g.,$\bar{N}_\Theta(o) = \mathbb{R}^2$ for $\Theta = \text{gph}|x|$), worsen calculus rules, and lead to the lost of robustness in non-Lipschitzian setting. On the other hand, it has computational advantages over (2.1) and allows to use for its study and applications powerful tools of convex analysis.

Given an extended-real-valued lower semicontinuous (l.s.c.) function $\psi\colon (-\infty,\infty]$ finite at $\bar{x}$, we define its *subdifferential* at this point geometrically via the normal cone (2.1) to the epigraph epi ψ by

$$\partial\psi(\bar{x}) := \left\{\xi \in \mathbb{R}^n|\, (\xi,-1) \in N_{\text{epi}\psi}\big(\bar{x},\psi(\bar{x})\big)\right\} \tag{2.3}$$

while noting that (2.3) admits several equivalent analytical representations. When ψ is locally Lipschitzian around $\bar{x}$, we get from (2.3) the (Clarke) *generalized gradient*

$$\bar{\partial}\psi(\bar{x}) := \text{co}\,\partial\psi(\bar{x}), \tag{2.4}$$

which can also be defined by different equivalent ways. Both constructions (2.3) and (2.4) reduce to the classical gradient and subdifferential of convex analysis when ψ is continuous differentiable around $\bar{x}$ and convex, respectively. Observe the convex hull symmetry property for locally Lipschitz functions

$$\text{co}\,\partial\big(-\psi\big)(\bar{x}) = -\text{co}\,\partial\psi(\bar{x}) \tag{2.5}$$

that is largely used in what follows.

Next we recall some notions to (single-valued and set-valued) mappings. Given a closed-graph multifunction $\Psi\colon \mathbb{R}^n \rightrightarrows \mathbb{R}^m$, its *coderivative* at the graph point $(\bar{x},\bar{y}) \in \text{gph}\,\Psi := \{(x,y) \in \mathbb{R}^n \times \mathbb{R}^m|\, y \in \Psi(x)\}$ is defined as a set-valued mapping generated by the normal cone (2.1) to the graph of Ψ by

$$D^*\Psi(\bar{x},\bar{y})(y^*) := \left\{x^* \in \mathbb{R}^n|\, (x^*,-y^*) \in N_{\text{gph}\,\Psi}(\bar{x},\bar{y})\right\},\ y^* \in \mathbb{R}^m, \tag{2.6}$$

where $\bar{y} = \Psi(\bar{x})$ is dropped in notation if $\Psi\colon \mathbb{R}^n \to \mathbb{R}^m$ is single-valued. If in the latter case Ψ is continuously differentiable at $\bar{x}$, the coderivative (2.6) is single-valued as well and reduced to the adjoint/transposed Jacobian linearly applied to y^*:

$$D^*\Psi(\bar{x})(y^*) = \left\{\nabla\Psi(\bar{x})^*y^*\right\}\ \text{for all}\ y^* \in \mathbb{R}^m.$$

In general the coderivative (2.6) is a positively homogeneous set-valued mapping satisfying comprehensive calculus rules. Observe the scalarization formula

$$D^*\Psi(\bar{x})(y^*) = \partial\big\langle y^*,\Psi\big\rangle(\bar{x}),\quad y^* \in \mathbb{R}^m,$$

relating the coderivative and subdifferential of single-valued Lipschitzian mappings.

Proceeding further with set-valued mappings, recall that $\Psi\colon \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ is *inner semicompact* at some point $\bar{x}$ with $\Psi(\bar{x}) \neq \emptyset$ if for every sequence $x_k \to \bar{x}$ with $\Psi(x_k) \neq \emptyset$ there is a sequence of $y_k \in \Psi(x_k)$ that contains a convergent subsequence as $k \to \infty$. It follows that the inner semicompactness holds in the finite-dimensional setting under consideration whenever Ψ is uniformly bounded around $\bar{x}$, i.e., there exists a neighborhood U of $\bar{x}$ and a bounded set $\Theta \subset \mathbb{R}^m$ such that $\Psi(x) \subset \Theta$ for all $x \in U$.

The following property is more restrictive while bring us to more precise results of the coderivative calculus and its applications. The mapping $\Psi\colon \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ is said to be *inner semicontinuous* at $(\bar{x},\bar{y}) \in \text{gph}\,\Psi$ if for every sequence $x_k \to \bar{x}$ there is a sequence of $y_k \in \Psi(x_k)$ that converges to $\bar{y}$ as $k \to \infty$. For single-valued mappings $\Psi\colon \mathbb{R}^n \to \mathbb{R}^m$ this

property obviously reduces to the continuity of Ψ at $\bar{x}$. Furthermore, if $\Psi \colon \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ is inner semicompact at $\bar{x}$ with $\Psi(\bar{x}) = \{\bar{y}\}$, then Ψ is inner semicontinuous at $(\bar{x}, \bar{y})$.

Finally in this section, we review some Lipschitzian properties of set-valued mappings used in what follows. The mapping $\Psi \colon \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ satisfies the *Lipschitz-like* (Aubin or pseudo-Lipschitz) property around $(\bar{x}, \bar{y}) \in \mathrm{gph}\,\Psi$ if there are neighborhoods U of $\bar{x}$ and V of $\bar{y}$ together with a constant $\ell \geq 0$ such that

$$\Psi(x) \cap V \subset \Psi(u) + \ell \|x - u\| \mathbb{B} \ \text{ for all } \ x, u \in U, \tag{2.7}$$

where $\mathbb{B}$ stands for the closed unit ball of the space in question. This property is a graphical localization of the classical (Hausdorff) local Lipschitz continuity of Ψ around $\bar{x}$, which corresponds to the case of $V = \mathbb{R}^m$ in (2.7). A complete characterization of the Lipschitz-like property (2.7), with precise computing the infimum of Lipschitzian moduli ℓ in (2.7), is given by the *coderivative/Mordukhovich criterion* in the form

$$D^*\Psi(\bar{x}, \bar{y})(0) = \{0\}. \tag{2.8}$$

We refer the reader to [12, 13, 14, 22] for more details and numerous applications of (2.8) in variational and nonlinear analysis, optimization, and related areas.

The *calmness* property of Ψ at $(\bar{x}, \bar{y})$ is defined via (2.7) with the fixed vector $u = \bar{x}$ therein. In the case where $V = \mathbb{R}^m$, this property goes back to Robinson [21] who called it the "upper Lipschitz property" of Ψ at $\bar{x}$. It is proved in [21] that the upper Lipschitz (and hence calmness) property holds at every point if the graph of Ψ is *piecewise polyhedral*, i.e., expressible as the union of finitely many polyhedral sets. Efficient conditions for the validity of the calmness property and its broad applications to variational analysis and optimization were strongly developed by Jiří Outrata and his collaborators; see, e.g., [9, 10, 19] among many other publications.

3. Sensitivity Analysis of the Two-Level Value Function

This section is devoted to a local sensitivity/stability analysis of the two-level value function φ_0 from (1.5). Besides deriving efficient pointbased upper estimate of the basic subdifferential (2.3), we establish here verifiable conditions for the local *Lipschitz continuity* of φ_0 around the reference point. From now on unless otherwise stated, the structure of the lower-level feasible solution map is considered to be

$$K(x) := \left\{ y \in \mathbb{R}^m \middle| (x, y) \in \Theta_2 \cap \vartheta_2^{-1}(\Lambda_2) \right\}, \tag{3.1}$$

where the sets $\Theta_2 \subset \mathbb{R}^n \times \mathbb{R}^m$ and $\Lambda_2 \subset \mathbb{R}^p$ are closed, and where the mapping $\vartheta_2 \colon \mathbb{R}^n \to \mathbb{R}^p$ is Lipschitz continuous.

As major constraint qualifications, we impose the *calmness* property at the reference points of the following set-valued mappings:

$$\begin{aligned}
\Phi^K(u) &:= \left\{ (x, y) \in \Theta_2 \middle| \vartheta_2(x, y) + u \in \Lambda_2 \right\}, \\
\Phi^S(u) &:= \left\{ (x, y) \in \mathrm{gph}\,K \middle| f(x, y) - \varphi(x) + u \leq 0 \right\},
\end{aligned} \tag{3.2}$$

where φ is the lower-level optimal value function defined in (1.7).

The aforementioned result by Robinson [21] ensures the calmness of Φ^K) (resp. Φ^S) at any point of the graph provided that the sets Θ_2 and Λ_2 (resp. $\mathrm{gph}\,K$) are polyhedral and the functions ϑ_2 (resp. f and φ) are linear with respect to their variables. Note that φ is piecewise linear if, e.g., the set-valued mapping K is defined by a system of linear equalities and/or inequalities while the cost function f is also linear.

Some results obtained below to verify the local Lipschitz continuity of, respectively, the lower-level and two-level value functions in the bilevel programs under consideration impose the following pointbased constraint qualifications:

$$\Big[(x^*, o) \in \partial \langle u, \vartheta_2 \rangle (\bar{x}, \bar{y}) + N_{\Theta_2}(\bar{x}, \bar{y}) \ \text{ with } \ u \in N_{\Lambda_2}(\vartheta_2(\bar{x}, \bar{y}))\Big] \Longrightarrow x^* = o, \qquad (3.3)$$

$$\left. \begin{array}{c} (x^*, o) \in r\partial(f - \varphi)(\bar{x}, \bar{y}) + \partial \langle u, \vartheta_2 \rangle (\bar{x}, \bar{y}) + N_{\Theta_2}(\bar{x}, \bar{y}) \\ r \geq o, \ u \in N_{\Lambda_2}\big(\vartheta_2(\bar{x}, \bar{y})\big) \end{array} \right\} \Longrightarrow x^* = o. \qquad (3.4)$$

These conditions are related to implementing the coderivative criterion (2.8) for the set-valued mappings K and S, respectively. It is worth mentioning that standard constraint qualifications as, e.g., the Mangasarian-Fromovitz one, fail to hold for the constraint structure $y \in S(x)$ with S taken from (1.1) and K described by simple smooth functions via inequality constraints in the lower-level optimal value function reformulation of bilevel programs; see [6] for more discussions. However, condition (3.4) can be satisfied to provide therefore a valuable constraint qualification in bilevel programming. To illustrate it, we recall the following typical statement in this direction taken from [4].

Proposition 3.1 (validity of the coderivative-based constraint qualification). Consider φ in (1.7) with $K(x) := \{y \in \mathbb{R}^m | \ g(y) \leq o\}$, where the functions f and g are convex and smooth. Take $(\bar{x}, \bar{y})$ such that $\varphi(\bar{x}) < \infty$ and the mapping $M(u) := \{y \in \mathbb{R}^m | \ g(y) + u \leq o\}$ is calm at $(o, \bar{y})$. Then the constraint qualification (3.4) holds at $(\bar{x}, \bar{y})$.

The next theorem is the first result in the literature that provides a local sensitivity analysis of the two-level value function (1.5) in bilevel programs described by Lipschitzian functions. In its formulation we use a mild assumption about inner semicompactness of the solution map associated with φ_o and defined by

$$S_o(x) := \arg\min_y \{F(x, y) | \ y \in S(x)\} = \{y \in S(x) | \ F(x, y) \leq \varphi_o(x)\}. \qquad (3.5)$$

Theorem 3.2 (subdifferential estimate and Lipschitz continuity of the two-level value function under inner semicompactness). Let in the framework of (1.5) the functions f, ϑ_2, and F be Lipschitz continuous around $(\bar{x}, y)$ with $y \in S(\bar{x})$ and $(\bar{x}, y)$ with $y \in S_o(\bar{x})$, respectively. Assume that the sets Θ_2 and Λ_2 are closed and the set-valued mapping Φ^K (resp. Φ^S) is calm at $(o, \bar{x}, y)$ with $y \in S(\bar{x})$ (resp. at $(o, \bar{x}, y)$ with $y \in S_o(\bar{x})$). Suppose also that S_o is inner semicompact at $\bar{x}$ and that the constraint qualifications (3.3) and (3.4) are satisfied at $(\bar{x}, y)$ with $y \in S(\bar{x})$ and $(\bar{x}, y)$ with $y \in S_o(\bar{x})$, respectively. Then the two-level value function φ_o is Lipschitz continuous around $\bar{x}$ and its basic subdifferential (2.3) at $\bar{x}$ satisfies the following upper estimate:

$$\partial\varphi_o(\bar{x}) \subset \bigcup_{y \in S_o(\bar{x})} \bigcup_{u \in N_{\Lambda_2}(\vartheta_2(\bar{x}, y))} \bigcup_{r \geq o} \Big\{ x^* \Big| \sum_{s=1}^{n+1} v_s = 1, $$
$$v_s \geq o, \ u_s \in N_{\Lambda_2}(\vartheta_2(\bar{x}, y_s)), \ y_s \in S(\bar{x}), \ s = 1, \ldots, n+1,$$
$$(x_s^*, o) \in \partial f(\bar{x}, y_s) + \partial \langle u_s, \vartheta_2 \rangle (\bar{x}, y_s) + N_{\Theta_2}(\bar{x}, y_s), \ s = 1, \ldots, n+1,$$
$$\Big(x^* + r \sum_{s=1}^{n+1} v_s x_s^*, o\Big) \in \partial F(\bar{x}, y) + r\partial f(\bar{x}, y) + \partial \langle u, \vartheta_2 \rangle (\bar{x}, y) + N_{\Theta_2}(\bar{x}, y)\Big\}.$$

Proof. Pick $\bar{x}^* \in \partial\varphi_o(\bar{x})$ and deduce from [18, Theorem 7(ii)], by taking into account the inner semicompactness of S_o at $\bar{x}$ and the Lipschitz continuity of F around $(\bar{x}, y)$ with $y \in$

$S_\circ(\bar{x})$, the existence of $y \in S_\circ(\bar{x})$ with $\bar{x}^* \in x^* + D^*S(\bar{x},y)(y^*)$ for some $(x^*, y^*) \in \partial F(\bar{x},y)$. Noting then that $\mathrm{gph}\, K = \Phi^K(\circ)$ and applying [9, Theorem 4.1], we get that for any $(x^*, y^*) \in N_{\mathrm{gph}\, K}(\bar{x},y)$ there exists $u \in N_{\Lambda_2}(\vartheta_2(\bar{x},y))$ such that $(x^*, y^*) \in \partial\langle u, \vartheta_2\rangle(\bar{x},y) + N_{\Theta_2}(\bar{x},y)$ with $y \in S(\bar{x})$. This uses the assumed closedness of Θ_2 and Λ_2, Lipschitz continuity of ϑ_2, and calmness of Φ^K. It follows from this estimate and the coderivative definition (2.6) that for $x^* \in D^*K(\bar{x},y)(y^*)$ there exist $u \in N_{\Lambda_2}(\vartheta_2(\bar{x},y))$ satisfying

$$(x^*, -y^*) \in \partial\langle u, \vartheta_2\rangle(\bar{x},y) + N_{\Theta_2}(\bar{x},y).$$

A clear consequence of this implication tells us that the constraint qualification (3.3) is a sufficient condition for the coderivative criterion (2.8) to hold for the set-valued mapping K at $(\bar{x}, y)$ with $y \in S(\bar{x})$. Combining the latter fact with the inner semicompactness of S (obtained from that of $S_\circ$ since $S_\circ(x) \subseteq S(x)$ for all $x \in X$), we derive from [15, Theorem 5.2 (ii)] the Lipschitz continuity of the lower-level value function φ around $\bar{x}$.

Repeating now the above process in the case of S and observing that $\mathrm{gph}\, S = \Phi^S(\circ)$, we get that for any $(x^*, y^*) \in N_{\mathrm{gph}\, S}(\bar{x},y)$ with $y \in S_\circ(\bar{x})$ there is $r \geq \circ$ such that

$$(x^*, y^*) \in r\partial(f - \varphi)(\bar{x},y) + N_{\mathrm{gph}\, K}(\bar{x},y)$$

under the imposed calmness assumption on Φ^S at $(\bar{x}, y)$ and the established local Lipschitz continuity of the lower-level value function φ around $\bar{x}$. Combining the latter expression with the one obtained above in the case of K, it follows from the coderivative definition that for any $x^* \in D^*S(\bar{x},y)(y^*)$ there are $r \geq \circ$ and $u \in N_{\Lambda_2}(\vartheta_2(\bar{x},y))$ such that

$$(x^*, -y^*) \in r\partial(f - \varphi)(\bar{x},y) + \partial\langle u, \vartheta_2\rangle(\bar{x},y) + N_{\Theta_2}(\bar{x},y). \tag{3.6}$$

This allows us to conclude under the assumptions of the theorem that the imposed constraint qualification condition (3.4) at $(\bar{x}, y)$ implies the validity of the coderivative criterion (2.8) for S at the same point and therefore the Lipschitz-like property of S around it. Invoking again [15, Theorem 5.2(ii)] tells us that the two-level value function $\varphi_\circ$ is locally Lipschitz continuous around $\bar{x}$.

The presence of the lower-level value function φ in (3.6) gives us room for a further description of the upper bound for $D^*S(\bar{x},y)(y^*)$ in terms of problem data. Since φ is a marginal function [13], it follows from [13, Theorem 3.38], the structure of S with K in (3.1), and the normal cone calculus rules in [13, Theorems 3.4 and 3.8] that for any $x^* \in \partial\varphi(\bar{x})$ we can find $y \in S(\bar{x})$ and $u \in N_{\Lambda_2}(\vartheta_2(\bar{x},y))$ satisfying the inclusion

$$(x^*, \circ) \in \partial f(\bar{x},y) + \partial\langle u, \vartheta_2\rangle(\bar{x},y) + N_{\Theta_2}(\bar{x},y).$$

Furthermore, the verified local Lipschitz continuity of φ around $\bar{x}$ ensures that

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

where the equality holds due to the convex hull symmetry property (2.5). Applying now the classical Carathéodory's convex hull theorem in $\mathbb{R}^n$ leads us to the following upper estimate of the basic subdifferential of $-\varphi$: for any $x^* \in \partial(-\varphi)(\bar{x})$ there are $v_s \geq \circ$, $u_s \in N_{\Lambda_2}(\vartheta_2(\bar{x},y_s))$, $y_s \in S(\bar{x})$ as $s = 1,\ldots,n+1$, and

$$(x_s^*, \circ) \in \partial f(\bar{x},y_s) + \partial\langle u_s, \vartheta_2\rangle(\bar{x},y_s) + N_{\Theta_2}(\bar{x},y_s) \text{ with } \sum_{s=1}^{n+1} v_s = 1$$

such that $x^* = -\sum_{s=1}^{n+1} v_s x_s^*$. Employing this together with the subdifferential sum rule [13, Theorem 2.33] in inclusion (3.6), we get that for any $x^* \in D^* S(\bar{x}, y)(y^*)$ there are $v_s \geq 0$, $u_s \in N_{\Lambda_2}(\vartheta_2(\bar{x}, y_s))$, $y_s \in S(\bar{x})$, and

$$(x_s^*, 0) \in \partial f(\bar{x}, y_s) + \partial \langle u_s, \vartheta_2 \rangle (\bar{x}, y_s) + N_{\Theta_2}(\bar{x}, y_s)$$

with $s = 1, \ldots, n+1$ and $\sum_{s=1}^{n+1} v_s = 1$ such that

$$\left(x^* + r \sum_{s=1}^{n+1} v_s x_s^*, -y^* \right) \in r \partial f(\bar{x}, y) + \partial \langle u, \vartheta_2 \rangle (\bar{x}, y) + N_{\Theta_2}(\bar{x}, y).$$

This verifies the claimed estimate for $\partial \varphi_0$ and completes the proof of the theorem. $\square$

It is definitely appealing to simplify the subdifferential upper estimate for the two-level value function φ_0 in Theorem 3.2 by avoiding the convex hull therein. There are three approaches in the bilevel programming literature to proceed in this direction for the lower-level value function φ in optimistic programs with smooth data. One way is to assume the *full convexity* of the lower-level functions as in [3, 8, 14, 26]. The second one is imposing the *inner semicontinuity* property of the solution map instead of its inner semicompactness (cf. [3, 14, 15]), and the third approach exploits a *difference rule* for regular/Fréchet subgradients as in [14, 17]. In what follows we extend the first two approaches to the case of the two-level value function φ_0.

The next theorem deals with the fully convex case in the lower-level problem. To simplify the presentation, we use the following multipliers sets:

$$\Lambda(\bar{x}, y) := \left\{ \gamma \,\middle|\, 0 \in \partial_y f(\bar{x}, y) + \sum_{i=1}^{p} \gamma_i \partial_y g_i(\bar{x}, y), \ \gamma \geq 0, \ \langle \gamma, g(\bar{x}, y) \rangle = 0 \right\}, \tag{3.7}$$

$$\Lambda^\circ(\bar{x}, y) := \left\{ (r, \beta) \,\middle|\, r \geq 0, \ \beta \geq 0, \ \langle \beta, g(\bar{x}, y) \rangle = 0, \right.$$
$$\left. 0 \in \partial_y F(\bar{x}, y) + r \partial_y f(\bar{x}, y) + \sum_{i=1}^{p} \beta_i \partial_y g_i(\bar{x}, y) \right\}. \tag{3.8}$$

Theorem 3.3 (subdifferential estimate and Lipschitz continuity of the two-level value function under full convexity). Let the bilevel program (1.1), (3.1) be fully convex, i.e., the functions $f, \vartheta_2 := g$, and F are convex in (x, y) with $\Theta_2 := \mathbb{R}^n \times \mathbb{R}^m$ and $\Lambda_2 := \mathbb{R}^p_-$. Assume that the set-valued mapping Φ^K (resp. Φ^S) is calm at $(0, \bar{x}, y)$ with $y \in S(\bar{x})$ (resp. $(0, \bar{x}, y)$ with $y \in S_0(\bar{x})$), that S_0 is inner semicompact $\bar{x} \in \operatorname{dom} \varphi$, and that the constraint qualification (3.4) is satisfied at $(\bar{x}, y)$ with $y \in S_0(\bar{x})$. Then the two-level value function φ_0 is Lipschitz continuous around $\bar{x}$ with the subdifferential estimate

$$\partial \varphi_0(\bar{x}) \subset \bigcup_{y \in S_0(\bar{x})} \bigcup_{(r, \beta) \in \Lambda^\circ(\bar{x}, y)} \bigcup_{\gamma \in \Lambda(\bar{x}, y)} \left\{ \partial_x F(\bar{x}, y) + r \Big(\partial_x f(\bar{x}, y) - \partial_x f(\bar{x}, y) \Big) \right.$$
$$\left. + \sum_{i=1}^{p} \beta_i \partial_x g_i(\bar{x}, y) - r \sum_{i=1}^{p} \gamma_i \partial_x g_i(\bar{x}, y) \right\}.$$

Proof. It largely takes the same pattern as that of Theorem 3.2 with the following essential changes. We can easily check that the lower-level value function φ is convex under the assumptions made. Given $x^* \in \partial \varphi(\bar{x})$, this allows us–by taking into account the calmness

of the set-valued mapping Φ^K at $(0, \bar{x}, y)$ and basic convex analysis as in the proof of [26, Theorem 4.1]–to find a vector $\gamma \in \Lambda(\bar{x}, y)$ satisfying

$$x^* \in \partial_x f(\bar{x}, y) + \sum_{i=1}^{p} \gamma_i \partial_x g_i(\bar{x}, y) \ \text{ whenever } \ y \in S(\bar{x}).$$

The second observation is that, by the convexity of φ, the symmetry property (2.5) reads

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

Combining it with the previous facts and using the decomposition property

$$\partial\psi(x, y) \subset \partial_x \psi(x, y) \times \partial_y \psi(x, y)$$

valid for the convex functions F, f, and g, we complete the proof. $\square$

The following statement also requires the convexity of the bilevel program data while drops some other assumptions of the previous one in a particular setting.

Proposition 3.4 (sensitivity of the two-level value function for fully convex programs with parameter-independent lower-level solution sets). Consider the bilevel program (1.3) with $S(x) := S := \arg\min\{f(y) \mid g(y) \leq 0\}$ as $x \in X$, where the functions f, g, and F convex in y and (x, y), respectively. Then the two-level value function $\varphi_\circ$ is locally Lipschitzian around $\bar{x} \in \text{dom}\,\varphi_\circ$, and we have

$$\partial\varphi_\circ(\bar{x}) \subset \partial_x F(\bar{x}, y) \ \text{ for any } \ y \in S_\circ(\bar{x}).$$

Proof. The assumptions made ensure the local Lipschitz continuity of the convex function $\varphi_\circ$ around $\bar{x}$. Arguing as in the proof of [26, Theorem 4.1] with taking into account Proposition 3.1, pick $\bar{y} \in S_\circ(\bar{x})$ and observe that for any $u \in \partial\varphi_\circ(\bar{x})$ the pair $(\bar{x}, \bar{y})$ is an optimal solution to the nonlinear program

$$\min_{x, y} \{F(x, y) - \langle u, x \rangle\} \ \text{ subject to } \ (x, y) \in \mathbb{R}^n \times S. \tag{3.9}$$

Applying now [13, Proposition 5.3] gives us the necessary optimality condition in problem (3.9) written in the form

$$(u, 0) \in \partial F(\bar{x}, \bar{y}) + N_{\mathbb{R}^n \times S}(\bar{x}, \bar{y}).$$

The decomposition property for the fully convex function F yields the inclusions

$$u \in \partial_x F(\bar{x}, \bar{y}) \ \text{ and } \ 0 \in N_S(\bar{y}),$$

which complete the proof, since the latter one is obvious. $\square$

The last result of this section concerns again the general two-level value function (1.5) and provides, besides its local Lipschitz property, a tighter upper estimate of the basic subdifferential (2.3) with the replacement of the inner semicompactness of the solution map $S_\circ$ in (3.5) by its inner semicontinuity. Furthermore, we replace the calmness of the set-valued mapping K from (3.2) by the following generalized form of the Mangasarian-Fromovitz constraint qualification expressed via the generalized gradient (2.4):

$$\left[0 \in \sum_{i=1}^{p} \gamma_i \bar{\partial} g_i(\bar{x}, \bar{y}), \ \gamma_i \geq 0, \gamma_i g_i(\bar{x}, \bar{y}) = 0, \ i = 1, \ldots, p\right] \Longrightarrow \gamma_i = 0, \ i = 1, \ldots, p. \tag{3.10}$$

Theorem 3.5 (subdifferential estimate and Lipschitz continuity of the two-level value function under inner semicontinuity). Consider the bilevel program (1.3), where the functions f, $\vartheta_2 := g$, and F are Lipschitz continuous near $(\bar{x}, \bar{y})$, where $\Theta_2 := \mathbb{R}^n \times \mathbb{R}^m$ and $\Lambda_2 := \mathbb{R}^p_-$, and where the set-valued mapping Φ^S is calm at $(o, \bar{x}, \bar{y})$. Suppose also that S_o is inner semicontinuous at $(\bar{x}, \bar{y})$, and that the constraint qualifications (3.4) and (3.10) are satisfied at $(\bar{x}, \bar{y})$. Then φ_o is Lipschitz continuous around $\bar{x}$ and

$$\partial \varphi_o(\bar{x}) \subset \bigcup_{r \geq o} \left\{ x^* \middle| (x^* + r x^*_\varphi, o) \in \partial F(\bar{x}, \bar{y}) + r \partial f(\bar{x}, \bar{y}) + \sum_{i=1}^p \beta_i \partial g_i(\bar{x}, \bar{y}), \right.$$

$$\left. (x^*_\varphi, o) \in \bar{\partial} f(\bar{x}, \bar{y}) + \sum_{i=1}^p \gamma_i \bar{\partial} g_i(\bar{x}, \bar{y}), \beta \geq o, \langle \beta, g \rangle(\bar{x}, \bar{y}) = o, \gamma \geq o, \langle \gamma, g \rangle(\bar{x}, \bar{y}) = o \right\}.$$

Proof. By $S_o(x) \subset S(x)$ for all $x \in X$ we have that S is also inner semicontinuous at $(\bar{x}, \bar{y})$. In addition to constraint qualification (3.10), it follows from [16, Theorem 5.9] that for any $u \in \bar{\partial} \varphi(\bar{x})$ there is a vector γ with $\gamma \geq o$ (componentwise) and $\langle \gamma, g(\bar{x}, \bar{y}) = o$ such that

$$(u, o) \in \bar{\partial} f(\bar{x}, \bar{y}) + \sum_{i=1}^p \gamma_i \bar{\partial} g_i(\bar{x}, \bar{y}).$$

As discussed above, the coderivative qualification condition (3.3) ensures that the lower-level value function φ is Lipschitz continuous around $\bar{x}$. Then it follows from [18, Theorem 7(i)], by the Lipschitz continuity of F around $(\bar{x}, \bar{y})$ and the inner semicontinuity of S_o at the same point, that for any $\bar{x}^* \in \partial \varphi_o(\bar{x})$ there is a pair $(x^*, y^*) \in \partial F(\bar{x}, \bar{y})$ with $\bar{x}^* \in x^* + D^* S(\bar{x}, \bar{y})(y^*)$. Remembering the Lipschitz continuity of φ around $\bar{x}$, we get as in the proof of Theorem 3.2 that for any $x^* \in D^* S(\bar{x}, \bar{y})(y^*)$ there are $r \geq o$, γ, and β with

$$\gamma \geq o, \ \langle \gamma, g(\bar{x}, \bar{y}) \rangle = o, \ \beta \geq o, \ \text{and} \ \langle \beta, g(\bar{x}, \bar{y}) \rangle = o$$

satisfying the following inclusions, by taking into account the convex hull property (2.5):

$$(x^*_\varphi, o) \in \bar{\partial} f(\bar{x}, \bar{y}) + \sum_{i=1}^p \gamma_i \bar{\partial} g_i(\bar{x}, y),$$

$$(x^* + r x^*_\varphi, -y^*) \in r \partial f(\bar{x}, y) + \sum_{i=1}^p \beta_i \partial g_i(\bar{x}, y).$$

Combining these facts, we arrive at the upper estimate of $\partial \varphi_o(\bar{x})$ claimed in the theorem. The Lipschitz continuity of the two-level value function φ_o is obtained as in the proof of Theorem 3.2; cf. also the proof of [15, Theorem 5.2(i)].   □

Remark 3.6 (**discussions on sensitivity analysis of the two-level value function**). The following observations are in order to make:

  (i) Assuming in the framework of Theorem 3.2 that $S_o(\bar{x}) = S(\bar{x}) = \{y\}$ and that the multiplier u in the upper estimate of $\partial \varphi(\bar{x})$ is unique, we see that all the upper bounds of $\partial \varphi_o(\bar{x})$ obtained in Theorems 3.2, 3.3, and 3.5 coincide provided that the functions involved are subdifferentially regular (i.e., $\partial \psi(\bar{x}) = \bar{\partial} \psi(\bar{x})$ in the setting under consideration) and the sets Θ_2 and Λ_2 are adjusted accordingly.

(ii) If all the functions involved in (1.5) are continuously differentiable around the reference points with $\Theta_2 := \mathbb{R}^n \times \mathbb{R}^m$ and $\Lambda_2 := \mathbb{R}^p_-$, then the upper estimates that we get for $\partial \varphi_o$ in Theorems 3.2 and 3.5 reduce to those obtained in [4, Theorem 5.9]. This will induce the same observation concerning the necessary optimality conditions for the optimistic bilevel program (P_o) obtained in Theorem 4.1.

(iii) Note that for the upper estimates of the basic subdifferential of φ_o derived in Theorems 3.2, 3.3, and 3.5, the coderivative constraint qualification (3.4) is in fact not necessary. The latter condition comes into play only to obtain the Lipschitz continuity of the aforementioned two-level value function.

4. Necessary Optimality Conditions for Optimistic Bilevel Programs

Our focus in this section is on deriving necessary optimality conditions for the original optimistic formulation of the bilevel programming problem (P_o) by using the two-level value function (1.5) and the sensitivity results obtained for it in Section 3. In addition to the expression of the lower-level feasible set-valued mapping K in (3.1), we consider now the representation of the upper-level feasible set given by

$$X := \Theta_1 \cap \vartheta_1^{-1}(\Lambda_1), \tag{4.1}$$

where the sets $\Theta_1 \subset \mathbb{R}^n$ and $\Lambda_1 \subset \mathbb{R}^k$ are always assumed to be closed, and where the mapping $\vartheta_1 : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^k$ is locally Lipschitz continuous. Similarly to the lower-level setting (3.2) the *calmness* assumption is imposed on the set-valued mapping

$$\Phi^X(u) := \left\{ x \in \Theta_1 \,\middle|\, \vartheta_1(x) + u \in \Lambda_1 \right\}. \tag{4.2}$$

The following necessary optimality conditions for the optimistic bilevel program (P_o) from (1.3) are based on sensitivity analysis of the two-level value function φ_o and implement all the independent results for such an analysis obtained in Theorems 3.2, 3.3, and 3.5 while combining them in the three statements of the next theorem.

Theorem 4.1 (necessary optimality conditions for the original version in optimistic bilevel programming). Let $\bar{x}$ be a local optimal solution to (P_o), where ϑ_1 is Lipschitz continuous around this point, and where the set-valued mapping Φ^X from (4.2) is calm at $(o, \bar{x})$. Then the following assertions hold:

(i) Under the assumptions of Theorem 3.2 there exist $y \in S_o(\bar{x})$, $(\alpha, u, r) \in \mathbb{R}^{k+p+1}$, $(u_s, v_s) \in \mathbb{R}^{p+1}$, $y_s \in S(\bar{x})$, and $x_s^* \in \mathbb{R}^n$ as $s = 1, \ldots, n+1$ such that

$$\left(r \sum_{s=1}^{n+1} v_s x_s^*, o \right) \in \partial F(\bar{x}, y) + r \partial f(\bar{x}, y) \tag{4.3}$$

$$+ \left[\partial \langle \alpha, \vartheta_1 \rangle(\bar{x}) + N_{\Theta_1}(\bar{x}) \right] \times \{o\} + \partial \langle u, \vartheta_2 \rangle(\bar{x}, y) + N_{\Theta_2}(\bar{x}, y),$$

$$\text{for } s = 1, \ldots, n+1, \ (x_s^*, o) \in \partial f(\bar{x}, y_s) + \partial \langle u_s, \vartheta_2 \rangle(\bar{x}, y_s) + N_{\Theta_2}(\bar{x}, y_s), \tag{4.4}$$

$$\text{for } s = 1, \ldots, n+1, \ u_s \in N_{\Lambda_2}\big(\vartheta_2(\bar{x}, y_s)\big), \ v_s \geq o, \ \sum_{s=1}^{n+1} v_s = 1, \tag{4.5}$$

$$u \in N_{\Lambda_2}\big(\vartheta_2(\bar{x}, y)\big), \tag{4.6}$$

$$r \geq o, \ \alpha \in N_{\Lambda_1}\big(\vartheta_1(\bar{x})\big). \tag{4.7}$$

(ii) Under the assumptions of Theorem 3.3 there exist $y \in S_o(\bar{x})$, α, β, γ and r such that condition (4.7) holds together with

$$o \in \partial_x F(\bar{x}, y) + r\Big(\partial_x f(\bar{x}, y) - \partial_x f(\bar{x}, y)\Big) + \sum_{i=1}^{p} \beta_i \partial_x g_i(\bar{x}, y) \tag{4.8}$$

$$- r \sum_{i=1}^{p} \gamma_i \partial_x g_i(\bar{x}, y) + \partial \langle \alpha, \vartheta_1 \rangle(\bar{x}) + N_{\Theta_1}(\bar{x}),$$

$$o \in \partial_y F(\bar{x}, y) + r \partial_y f(\bar{x}, y) + \sum_{i=1}^{p} \beta_i \partial_y g_i(\bar{x}, y), \tag{4.9}$$

$$o \in \partial_y f(\bar{x}, y) + \sum_{i=1}^{p} \gamma_i \partial_y g_i(\bar{x}, y), \tag{4.10}$$

$$\beta \geq o, \ \langle \beta, g(\bar{x}, y) \rangle = o, \tag{4.11}$$

$$\gamma \geq o, \ \langle \gamma, g(\bar{x}, y) \rangle = o. \tag{4.12}$$

(iii) Under the assumptions of Theorem 3.5 there exist x^*, α, β, γ, and r such that we have conditions (4.7) and (4.11)–(4.12) (where $y := \bar{y}$) together with

$$(rx^*, o) \in \partial F(\bar{x}, \bar{y}) + r \partial f(\bar{x}, \bar{y}) + \sum_{i=1}^{p} \beta_i \partial g_i(\bar{x}, \bar{y})$$
$$+ \Big[\partial \langle \alpha, \vartheta_1 \rangle(\bar{x}) + N_{\Theta_1}(\bar{x}) \Big] \times \{o\}, \tag{4.13}$$

$$(x^*, o) \in \bar{\partial} f(\bar{x}, \bar{y}) + \sum_{i=1}^{p} \gamma_i \bar{\partial} g_i(\bar{x}, \bar{y}). \tag{4.14}$$

Proof. Under the assumptions of either (i), or (ii), or (iii) the two-level value function φ_o is Lipschitz continuous near $\bar{x}$ as proved in the aforementioned theorems. Then remembering the closedness of the set $X = \Theta_1 \cap \vartheta^{-1}(\Lambda_1)$, we get by applying the necessary optimality condition from [13, Proposition 5.3] to problem (P_o) due to its structure in (1.3) with the feasible set (4.1) that

$$o \in \partial \varphi_o(\bar{x}) + N_X(\bar{x}). \tag{4.15}$$

The imposed calmness constraint qualification on the set-valued mapping Φ^X at $(o, \bar{x})$ allows us to deduce from [9, Theorem 4.1] that for any $x^* \in N_X(\bar{x})$ there is a vector $\alpha \in N_{\Lambda_1}(\vartheta_1(\bar{x}))$ satisfying the inclusion

$$x^* \in \partial \langle \alpha \vartheta_1 \rangle(\bar{x}) + N_{\Theta_1}(\bar{x}).$$

Combining the latter with (4.15) and the upper estimate for $\partial \varphi_o(\bar{x})$ from Theorems 3.2, 3.3, and 3.5, we arrive at the conclusions in assertions (i)–(iii) of the theorem. $\square$

Remark 4.2 (**discussion on necessary optimality conditions in optimistic bilevel programming**). The following remarks discuss some important features of the obtained necessary optimality conditions for the original optimistic model (P_o) and their comparison with known results for optimistic bilevel programs.

(i) As follows from the proof of Theorem 4.1, the Lipschitz continuity of the two-level value function $\varphi_\circ$ is used to obtain the intermediate necessary optimality condition in form (4.15). It is worth mentioning that the necessary optimality condition (4.15) holds under the more general constraint qualification

$$\partial^\infty \varphi_\circ(\bar{x}) \cap \left(-N_X(\bar{x})\right) = \{0\}, \tag{4.16}$$

where $\partial^\infty \psi(\bar{x})$ is the *singular subdifferential* of $\psi \colon \mathbb{R}^n \to (-\infty, \infty]$ at $\bar{x} \in \operatorname{dom} \psi$ given by

$$\partial^\infty \psi(\bar{x}) := \left\{ \xi \in \mathbb{R}^n \mid (\xi, 0) \in N_{\operatorname{epi}\psi}\left(\bar{x}, \psi(\bar{x})\right) \right\}. \tag{4.17}$$

It is well known in variational analysis [13, 22] that the singular subdifferential (4.17) of an extended-real-valued l.s.c. function in finite dimensions completely characterized its locally Lipschitzian property around the point in question via $\partial^\infty \psi(\bar{x}) = \{0\}$. On the other hand, the local Lipschitz continuity of $\varphi_\circ$ is by far not necessary for the validity of the qualification requirement (4.16), while the upper subdifferential estimates obtained in Theorems 3.2, 3.3, and 3.5 do not use the Lipschitz continuity of the two-level value function (1.5). In particular, under the validity of (4.16) the coderivative constraint qualification (3.4) needed just for the Lipschitz continuity of $\varphi_\circ$ can be dropped. It happens, e.g., when $X := \mathbb{R}^n$ in the upper-level problem.

(ii) To compare Theorem 4.1 with previous work on the subject, consider the original optimistic bilevel optimization problem $(\mathrm{P}_\circ)$ in the particular case where $\Theta_1 := \mathbb{R}^n$, $\Lambda_1 := \mathbb{R}^k_-$, $\Theta_2 := \mathbb{R}^n \times \mathbb{R}^m$, and $\Lambda_2 := \mathbb{R}^p_-$. Then the necessary optimality conditions obtained in Theorem 4.1(i,ii,iii) reduce to those derived in [3, Theorem 5.1], [3, Theorem 4.1], and [16, Theorem 6.4], respectively, for the now conventional form (1.2) in optimistic bilevel programming with the additional statement on $y \in S_\circ(\bar{x})$. In the framework of smooth data in optimistic bilevel programs, the results somewhat similar to (while different from) Theorem 4.1(i) can also be found in [6, 25]. Observe that the aforementioned papers employ the so-called *partial calmness* constraint qualification introduced in Ye and Zhu [25], which is replaced here by the calmness of the set-valued mapping Φ^S. In fact, these constraint qualifications are largely related; cf. the proof of [6, Theorem 4.10]. Note also that the (Mangasarian-Fromovitz type) upper-level and lower-level regularity conditions used in [3, 16] yield (being significantly stronger) the fulfilment of the calmness conditions for the set-valued mappings Φ^X and Φ^K imposed in Theorem 4.1. Furthermore, the coderivative constraint qualification (3.3) readily follows from the lower-level regularity. The reader is referred to [3, 6, 16, 25] and the bibliographies therein for the definitions and more detailed discussions on the partial calmness, upper-level and lower-level regularity, and related constraint qualifications. We particularly emphasize remarkable developments by Henrion and Surowiec [11] who strongly investigated relationships between calmness, partial calmness as well as novel constraint qualifications introduced and implemented by them to derive new necessary optimality conditions for the conventional optimistic bilevel model with a convex lower-level problem.

5. Sensitivity and Optimality Conditions for Pessimistic Bilevel Programs

This section is devoted to the study of the pessimistic bilevel program (P_p) formulated in (1.4) with its initial data X, S, φ, and K defined as in (4.1), (1.1), (1.7), and (3.1),

respectively. Developing the two-level value function approach, we obtain here the results on sensitivity analysis and necessary optimality conditions for (P_p) similar to those established for the original optimistic model (P_o) in Sections 3 and 4. In fact, a strong advantage of our two-level value function approach to bilevel programming is the possibility to simultaneously apply it to sensitivity and optimality issues in both optimistic and pessimistic versions. In this section we demonstrate it for the case of the pessimistic model (1.4) by involving the maximizing two-level value function φ_p from (1.6) instead of φ_o from (1.5) used above. Taking into account the relationship

$$\varphi_p(x) = -\varphi_p^o(x) \text{ with } \varphi_p^o(x) := \min_y \left\{ -F(x,y) \big| y \in S(x) \right\} \tag{5.1}$$

and that the local Lipschitzian property is invariant with respect to the negative sign, we are able to deduce sensitivity results for the pessimistic program from those obtained in Theorems 3.2, 3.3, and 3.5. To proceed in this way, define the solution set

$$S_o^p(x) := \left\{ y \in S(x) \big| F(x,y) + \varphi_p^o(x) \geq 0 \right\}. \tag{5.2}$$

Theorem 5.1 (sensitivity analysis for pessimistic bilevel programs). Considering the maximizing two-level value function φ_p in (1.6), the following assertions hold:

(i) Let the assumptions of Theorem 3.2 be satisfied for the function φ_p^o from (5.1) with the set S_o (3.5) replaced by S_o^p (5.2). Then φ_p is Lipschitz continuous around $\bar{x}$ and we have the subdifferential upper estimate

$$\begin{aligned}
\partial \varphi_p(\bar{x}) \subset \bigg\{ &-\sum_{t=1}^{n+1} \eta_t x_t^* \bigg| \sum_{s=1}^{n+1} v_s = 1, \; \sum_{t=1}^{n+1} \eta_t = 1, \\
&v_s \geq 0, \; u_s \in N_{\Lambda_2}\big(\vartheta_2(\bar{x}, y_s)\big), \; y_s \in S(\bar{x}), \; s = 1, \ldots, n+1, \\
&(x_s^*, 0) \in \partial f(\bar{x}, y_s) + \partial \langle u_s, \vartheta_2 \rangle(\bar{x}, y_s) + N_{\Theta_2}(\bar{x}, y_s), \; s = 1, \ldots, n+1, \\
&\eta_t \geq 0, \; r_t \geq 0, \; u_t \in N_{\Lambda_2}\big(\vartheta_2(\bar{x}, y_t)\big), \; y_t \in S_o^p(\bar{x}), \; t = 1, \ldots, n+1, \\
&(x_t^* + r_t \sum_{s=1}^{n+1} v_s x_s^*, 0) \in \partial(-F)(\bar{x}, y_t) + r_t \partial f(\bar{x}, y_t) \\
&\qquad\qquad + \partial \langle u_t, \vartheta_2 \rangle(\bar{x}, y_t) + N_{\Theta_2}(\bar{x}, y_t), \; t = 1, \ldots, n+1 \bigg\}.
\end{aligned}$$

(ii) Let the assumptions of Theorem 3.3 be satisfied for φ_p^o with S_o replaced by S_o^p. Then φ_p is Lipschitz continuous around $\bar{x}$ and we have the following inclusion, where $\Lambda(\bar{x}, y_t)$ and $\Lambda^o(\bar{x}, y_t)$ are given in (3.7) and (3.8), respectively:

$$\begin{aligned}
\partial \varphi_p(\bar{x}) \subset \bigg\{ &-\sum_{t=1}^{n+1} \eta_t x_t^* \bigg| \sum_{t=1}^{n+1} \eta_t = 1, \\
&\eta_t \geq 0, \; (r_t, \beta^t) \in \Lambda^o(\bar{x}, y_t), \; \gamma \in \Lambda(\bar{x}, y_t). \; y_t \in S_o^p(\bar{x}), \; t = 1, \ldots, n+1, \\
&x_t^* \in \partial_x(-F)(\bar{x}, y_t) + r_t\big(\partial_x f(\bar{x}, y_t) - \partial_x f(\bar{x}, y_t)\big) \\
&\qquad + \sum_{i=1}^{p} \beta_i^t \partial_x g_i(\bar{x}, y_t) - r_t \sum_{i=1}^{p} \gamma_i \partial_x g_i(\bar{x}, y_t), \; t = 1, \ldots, n+1 \bigg\}.
\end{aligned}$$

(iii) Let the assumptions of Theorem 3.5 be satisfied for φ_p^o with S_o replaced by S_o^p. Then φ_p is Lipschitz continuous near $\bar{x}$ and we have the inclusion

$$\partial \varphi_p(\bar{x}) \subset \Big\{ -\sum_{t=1}^{n+1} \eta_t x_t^* \Big| \sum_{t=1}^{n+1} \eta_t = 1,$$
$$\eta_t \geq 0,\ r_t \geq 0,\ \beta^t \geq 0,\ \langle \beta^t \rangle, g(\bar{x},\bar{y}) = 0,\ t = 1,\ldots,n+1,$$
$$(x_\varphi^*, 0) \in \bar{\partial} f(\bar{x},\bar{y}) + \sum_{i=1}^{p} \gamma_i \bar{\partial} g_i(\bar{x},\bar{y}),\ \gamma \geq 0,\ \langle \gamma, g(\bar{x},\bar{y}) \rangle = 0,$$
$$(x_t^* + r_t x_\varphi^*, 0) \in \partial(-F)(\bar{x},\bar{y}) + r_t \partial f(\bar{x},\bar{y}) + \sum_{i=1}^{p} \beta_i^t \partial g_i(\bar{x},\bar{y}),\ t = 1,\ldots,n+1 \Big\}.$$

Proof. Due to (5.1), the Lipschitz continuity of φ_p follows from Theorems 3.2, 3.5, and 3.3, respectively. To verify the upper estimate of $\partial \varphi_p$ in (i), observe from the proof of Theorem 3.2 that for any $x^* \in \partial \varphi_p^o(\bar{x})$ there exist a number $r \geq 0$ and vectors $y \in S_o^p(\bar{x})$, $u \in N_{\Lambda_2}(\vartheta_2(\bar{x},y))$ such that $(x^*, 0) = (x^{*1}, y^{*1}) + (x^{*2}, y^{*2}) + (x^{*3}, y^{*3})$ with the couples

$$(x^{*1}, y^{*1}) \in (-F)(\bar{x},y) + r\partial f(\bar{x},y),\ (x^{*2}, y^{*2}) \in r\partial(-\varphi)(\bar{x}) \times \{0\},$$
$$\text{and}\ (x^{*3}, y^{*3}) \in \partial \langle u, \vartheta_2 \rangle(\bar{x},y) + N_{\Theta_2}(\bar{x},y).$$

Furthermore, it follows from (2.5) that $\partial \varphi_p(\bar{x}) \subset -\text{co}\,\partial \varphi_p^o(\bar{x})$, and that for any $x^* \in \text{co}\,\partial \varphi_p^o(\bar{x})$ there exist x_t^* as $t = 1,\ldots,n+1$ satisfying $x^* = \sum_{t=1}^{n+1} \eta_t x_t^*$, where

$$\sum_{t=1}^{n+1} \eta_t = 1,\ y_t \in S_o^p(\bar{x}),\ \eta_t \geq 0,\ r_t \geq 0,$$
$$u_t \in N_{\Lambda_2}\big(\vartheta_2(\bar{x},y_t)\big),\ (x_t^*, 0) = (x_t^{*1}, y_t^{*1}) + (x_t^{*2}, y_t^{*2}) + (x_t^{*3}, y_t^{*3}),$$
$$(x_t^{*1}, y_t^{*1}) \in (-F)(\bar{x},y_t) + r\partial f(\bar{x},y_t),\ (x_t^{*2}, y_t^{*2}) \in r\partial(-\varphi)(\bar{x}) \times \{0\},$$
$$\text{and}\ (x_t^{*3}, y_t^{*3}) \in \partial \langle u_t, \vartheta_2 \rangle(\bar{x},y_t) + N_{\Theta_2}(\bar{x},y_t)\ \text{as}\ t = 1,\ldots,n+1.$$

This yields the existence of $x_\varphi^* \in \text{co}\,\partial \varphi(\bar{x})$ with $(x_t^* + r_t x_\varphi^*, 0) = (x_t^{*1}, y_t^{*1}) + (x_t^{*3}, y_t^{*3})$,

$$(x_t^{*1}, y_t^{*1}) \in (-F)(\bar{x},y_t) + r\partial f(\bar{x},y_t),\ \text{and}\ (x_t^{*3}, y_t^{*3}) \in \partial \langle u_t, \vartheta_2 \rangle(\bar{x},y_t) + N_{\Theta_2}(\bar{x},y_t).$$

Invoking now the upper estimate of $\partial \varphi(\bar{x})$ from the proof of Theorem 3.2 verifies assertion (i). The proofs of assertions (ii) and (iii) are similar with the usage of the results obtained in Theorems 3.3 and 3.5, respectively. $\square$

Proceeding as in the case of the original optimistic bilevel program (1.3), we derive now necessary optimality conditions for nonsmooth pessimistic bilevel programs.

Theorem 5.2 (necessary optimality conditions in nonsmooth pessimistic programming). Let $\bar{x}$ be a local optimal solution to problem (P_p) with the constraints defined at the beginning of this section, where the function ϑ_1 is locally Lipschitz continuous around $\bar{x}$, and where the sets Θ_1 and Λ_1 are closed. If the set-valued mapping Φ^X from (4.2) is calm at $(0,\bar{x})$, then the following assertions hold:

(i) Under the assumptions of Theorem 5.1(i) there exist $y_t \in S_o^p(\bar{x})$, η_t, r_t, u_t with $t = 1,\ldots,n+1$, $y_s \in S(\bar{x})$, v_s, and u_s with $s = 1,\ldots,n+1$ such that conditions (4.4)–(4.5) and (4.7) (where $r := r_t$) are satisfied together with

$$\sum_{t=1}^{n+1} \eta_t x_t^* \in \partial\langle \alpha, \vartheta_1\rangle(\bar{x}) + N_{\Theta_1}(\bar{x}) \text{ for all } t = 1,\ldots,n+1, \tag{5.3}$$

$$\left(x_t^* + r_t \sum_{s=1}^{n+1} v_s x_s^*, o\right) \in \partial(-F)(\bar{x}, y_t) + r_t \partial f(\bar{x}, y_t) + \tag{5.4}$$

$$\partial\langle u_t, \vartheta_2\rangle(\bar{x}, y_t) + N_{\Theta_2}(\bar{x}, y_t),$$

$$\text{for } t = 1,\ldots,n+1, \, \eta_t \geq o, \, \sum_{t=1}^{n+1} \eta_t = 1, \tag{5.5}$$

$$\text{for } t = 1,\ldots,n+1, \, u_t \in N_{\Lambda_2}\big(\vartheta_2(\bar{x}, y_t)\big). \tag{5.6}$$

(ii) Under the assumptions of Theorem 5.1(ii) for any $y \in S(\bar{x})$ there exist γ, $y_t \in S_o^p(\bar{x})$, η_t, r_t, and β^t with $t = 1,\ldots,n+1$ such that the relationships in (4.7) (with $r := r_t$), (4.10) (with $y := y_t$), (4.12) (with $y := y_t$), (5.3), and (5.5) hold together with the following conditions satisfying for all $t = 1,\ldots,n+1$:

$$x_t^* \in \partial_x(-F)(\bar{x}, y_t) + r_t\big(\partial_x f(\bar{x}, y_t) - \partial_x f(\bar{x}, y)\big)$$

$$+ \sum_{i=1}^{p} \beta_i^t \partial_x g_i(\bar{x}, y_t) - r_t \sum_{i=1}^{p} \gamma_i \partial_x g_i(\bar{x}, y), \tag{5.7}$$

$$o \in \partial_y F(\bar{x}, y_t) + r_t \partial_y f(\bar{x}, y_t) + \sum_{i=1}^{p} \beta_i^t \partial_y g_i(\bar{x}, y_t), \tag{5.8}$$

$$\beta^t \geq o, \, \langle \beta^t, g(\bar{x}, y_t)\rangle = o. \tag{5.9}$$

(iii) Under the assumptions of Theorem 5.1(iii) there exist η_t, r_t, u_t as $t = 1,\ldots,n+1$ such that the relationships in (4.7) (with $r := r_t$), (4.12) (with $y := \bar{y}$), (4.14), (5.3), (5.5), and (5.9) are satisfied together with the condition

$$\left(x_t^* + r_t x_\varphi^*, o\right) \in \partial(-F)(\bar{x}, \bar{y}) + r_t \partial f(\bar{x}, \bar{y}) + \sum_{i=1}^{p} \beta_i^t \partial g_i(\bar{x}, \bar{y}), \, t = 1,\ldots,n+1.$$

Proof. It follows the lines in the proof of Theorem 4.1 with replacing φ_o by φ_p therein and applying the subdifferential estimates for φ_p obtained in assertions (i), (ii), and (iii) of Theorem 5.1, respectively. $\square$

Remark 5.3 (**comparison and further discussions**). Let us mention here some other relations and consequences of the developed approach and obtained results.

(i) If all the functions involved in (P_p) are smooth and $\Theta_1 := \mathbb{R}^n$, $\Lambda_1 := \mathbb{R}_-^k$, $\Theta_2 := \mathbb{R}^n \times \mathbb{R}^m$, and $\Lambda_2 := \mathbb{R}_-^p$ therein, the necessary optimality conditions for pessimistic bilevel program established in Theorem 5.2 agree with those obtained in [5]. Note that the latter paper explores also other approaches to pessimistic bilevel programs with smooth

data that are related to the generalized equation and Karush-Kuhn-Tucker reformulations. These two approaches lead us to necessary optimality conditions, which are largely different from those obtained here by using the two-level value function approach.

(ii) Theorem 5.1 sheds new light on the local Lipschitz continuity and upper estimates for the basic subdifferential of the standard maximum functions corresponding to the replacement of the optimal solution map S in (1.6) by the constraint mapping K of type (3.1). Proceeding as in the proof of Theorem 5.2 with $\Theta_1 := \mathbb{R}^n$, $\Lambda_1 := \mathbb{R}^k_-$, $\Theta_2 := \mathbb{R}^n \times \mathbb{R}^m$, and $\Lambda_2 := \mathbb{R}^p_-$ allows us to recover, in particular, the necessary optimality conditions for minimax problem obtained in [26].

6. Concluding Remarks

This paper shows that the two-level value function approach is instrumental to handle both original optimistic and pessimistic models of bilevel programming with nonsmooth data. Developing this approach allows us to carry out a local sensitivity analysis (Lipschitz continuity and subdifferential estimates) of the two-level value function and derive necessary optimality conditions for local solutions to optimistic and pessimistic bilevel programs without any differentiability assumptions.

Our future research aims to implement the obtained stability and optimality results to designing appropriate numerical algorithms for solving bilevel programs, which belong to the class of ill-posed optimization problems; see [27] for more discussions and some developments in this direction for optimistic bilevel models with smooth data. Eventually we plan to proceed with developing generalized versions of Newton's method by involving the construction of the second-order subdifferential/generalized Hessian for extended-real-valued functions [13].

Another goal of our future research is combine advantages of the two-level value function approach with the one initiated in [11] and implemented there for conventional optimistic bilevel programs with twice continuously differential initial data and convex lower-level problems. The latter approach produced in [11] first-order and second-order necessary optimality conditions under certain calmness constraint qualifications. We plan to extend this approach, by marring it to the two-level value function one under consideration, to the settings of the original optimistic and pessimistic bilevel programs with smooth and nonsmooth data. The aforementioned notion of the second-order subdifferential seems to be instrumental in the absence of twice continuous differentiability.

References

1. M. Červinka, C. Matonoha and J. V. Outrata, *On the computation of relaxed pessimistic solutions to MPECs*, Optim. Methods Softw. **28** (2013), 186–206.
2. S. Dempe, *Foundations of Bilevel Programming*, Kluwer Academic Publishers, Dordrecht, 2002.
3. S. Dempe, J. Dutta and B.S. Mordukhovich, *New necessary optimality conditions in optimistic bilevel programming*, Optimization **56** (2007), 577–604.
4. S. Dempe, B.S. Mordukhovich and A.B. Zemkoho, *Sensitivity analysis for two-level value functions with applications to bilevel programming*, SIAM J. Optim. **22** (2012), 1309–1343.
5. ———, *Necessary optimality conditions in pessimistic bilevel programming*, Optimization **63** (2014), 505–533.
6. S. Dempe and A.B. Zemkoho, *The generalized Mangasarian-Fromowitz constraint qualification and optimality conditions for bilevel programs*, J. Optim. Theory Appl. **148** (2011), 433–441.

7. _____, *On the Karush-Kuhn-Tucker reformulation of the bilevel optimization problem*, Nonlinear Anal. **75** (2014) 1202–1218.

8. N. Dinh, B.S. Mordukhovich and T.T.A. Nghia, *Subdifferentials of value functions and optimality conditions for DC and bilevel infinite and semi-infinite programs*, Math. Program. **123** (2010), 101-138.

9. R. Henrion, A. Jourani and J.V. Outrata, *On the calmness of a class of multifunctions*, SIAM J. Optim. **13** (2002), 603–618.

10. R. Henrion and J.V. Outrata, *A subdifferential condition for calmness of multifunctions*, J. Math. Anal. Appl. **258** (2001), 110–130.

11. R. Henrion and T. Surowiec, *On calmness conditions in convex bilevel programming*, Applic. Anal. **90** (2011), 951–970.

12. B.S. Mordukhovich, *Complete characterizations of openness, metric regularity, and Lipschitzian properties of multifunctions*, Trans. Amer. Math. Soc. **340** (1993), 1–35.

13. _____, *Variational Analysis and Generalized Differentiation, I: Basic Theory, II: Applications*, Springer, Berlin, 2006.

14. _____, *Basics of Variational Analysis and Applications*, Springer, New York, 2018.

15. B.S. Mordukhovich and N.M. Nam, *Variational stability and marginal functions via generalized differentiation*, Math. Oper. Res. **30**, 1–18.

16. B.S. Mordukhovich, N.M. Nam and H.M. Phan, *Variational analysis of marginal function with applications to bilevel programming problems*, J. Optim. Theory Appl. **152** (2012), 557–586.

17. B.S. Mordukhovich, N.M. Nam and N.D. Yen, *Fréchet subdifferential calculus and optimality conditions in nondifferentiable programming*, Optimization **55** (2006), 685–708.

18. _____, *Subgradients of marginal functions in parametric mathematical programming*, Math. Program. **116** (2009), 369–396.

19. B.S. Mordukhovich and J.V. Outrata, *Coderivative analysis of quasi-variational inclusions with applications to stability and optimization*, SIAM J. Optim. **18** (2007), 389–412.

20. J.V. Outrata, *On the numerical solution of a class of Stackelberg problems*, ZOR–Math. Methods of Oper. Res. **34** (1990), 255–277.

21. S.M. Robinson, *Some continuity properties of polyhedral multifunctions*, Math. Program. Study **14** (1981), 206–214.

22. R.T. Rockafellar and R.J-B. Wets, *Variational Analysis*, Springer, Berlin, 1998.

23. H.F. von Stackelberg, *Marktform und Gleichgewicht*, Springer, Berlin, 1934.

24. W. Wiesemann, A. Tsoukalas, P. Kleniati and B. Rustem, *Pessimistic bilevel optimization*, SIAM J. Optim. **23** (2013), 353–380.

25. J.J. Ye and D.L. Zhu, *Optimality conditions for bilevel programming problems*, Optimization **33** (1995), 9–27; Erratum in *Optimization* **39** (1997), 361–366.

26. A.B. Zemkoho, *A simple approach to optimality conditions in minmax programming*, Optimization **63** (2014), 385–401.

27. A.B. Zemkoho, *Solving ill-posed bilevel programs*, Set-Valued Var. Anal. **24** (2016), 423–448.

S. Dempe: Institut für Numerische Mathematik und Optimierung, TU Freiberg, Germany
E-mail address: dempe@math.tu-freiberg.de

B.S. Mordukhovich: Department of Mathematics, Wayne State University Detroit, USA
E-mail address: boris@math.wayne.edu

A.B. Zemkoho: School of Mathematics, University of Southampton, Southampton, UK
E-mail address: a.b.zemkoho@soton.ac.uk