

KKT REFORMULATION AND NECESSARY CONDITIONS FOR OPTIMALITY IN NONSMOOTH BILEVEL OPTIMIZATION

STEPHAN DEMPE* AND ALAIN B. ZEMKOHO†

Abstract. For a long time, the bilevel programming problem has essentially been considered as a special case of mathematical programs with equilibrium constraints (MPECs), in particular when the so-called KKT reformulation is in question. Recently though, this widespread believe was shown to be false in general. In this paper, other aspects of the difference between both problems are revealed as we consider the KKT approach for the nonsmooth bilevel program. It turns out that the new inclusion (constraint) which appears as a consequence of the partial subdifferential of the lower-level Lagrangian (PSLLL) places the KKT reformulation of the nonsmooth bilevel program in a new class of mathematical program with both set-valued and complementarity constraints. While highlighting some new features of this problem, we attempt here to establish close links with the standard optimistic bilevel program. Moreover, we discuss possible natural extensions for C-, M-, and S-stationarity concepts. Most of the results rely on a coderivative estimate for the PSLLL that we also provide in this paper.

Key words. nonsmooth bilevel optimization, parametric optimization, coderivative, variational analysis, constraint qualifications, stationarity conditions

AMS subject classifications. 90C26, 90C30, 90C31, 90C33, 90C46, 49M05

1. Introduction. Our basic interest in this paper is the following class of the standard optimistic bilevel programming problem that we denote by (P) :

$$\begin{aligned} & \min_{x,y} \{F(x,y) \mid y \in S(x), G_j(x) \leq 0, j = 1, \dots, k\} \\ & \text{with } S(x) := \arg \min_y \{f(x,y) \mid g_i(x,y) \leq 0, i = 1, \dots, p\}, \end{aligned} \quad (1.1)$$

where the functions $G_j [\mathbb{R}^n \rightarrow \mathbb{R}]$ for $j = 1, \dots, k$, define the upper-level constraints, while $g_i [\mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}]$ for $i = 1, \dots, p$, describe the lower-level constraints. On the other hand, $F [\mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}]$ and $f [\mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}]$ denote the upper- and lower-level objective/cost functions, respectively. The set-valued mapping $S : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ represents the solution/argminimum mapping of the so-called lower-level problem. Further recall that problem (1.1) as a whole is often called upper-level problem. All the functions involved in (P) are assumed to be locally Lipschitz continuous and not necessary continuously differentiable as it is often the case in the literature.

It is important to recall that functions used to model real situations are often not differentiable which was an essential initial point to investigate nonsmooth optimization problems. Recent applications of nonsmooth bilevel optimization include image denoising [18] and variational inequality problems [11]. Bilevel optimization problems are really hard problems, known to be NP-hard, see, e.g. [10] and others. Nevertheless, investigating them in a nonsmooth setting is interesting and can be helpful if the problem in a real situation cannot be formulated using only smooth functions. In that case it arises that the combination of the complementarity conditions and the new inclusion constraint replacing the lower-level problem lead to new challenges and thus to a new insight into the bilevel optimization problem; cf. Sections 3–7.

*Department of Mathematics and Computer Science, TU Bergakademie Freiberg, Akademiestraße 6, D-09596 Freiberg, Germany (dempe@math.tu-freiberg.de)

†School of Mathematics, University of Birmingham, Edgbaston, Birmingham, B15 2TT, UK (a.zemkoho@bham.ac.uk). The first version of the paper was completed while this author was a Research Associate at TU Bergakademie Freiberg, Akademiestraße 6, D-09596 Freiberg, Germany.

Nonsmooth bilevel optimization problems have been investigated before, see, for example, [19, 31], where solution methods are suggested for some special classes. Necessary optimality conditions are derived in [5, 6, 23, 36] while using the so-called *lower-level value function* (LLVF) approach

$$\min_{x,y} \{F(x,y) \mid G_j(x) \leq 0, j = 1, \dots, k, \quad (1.2)$$

$$f(x,y) \leq \varphi(x), g_i(x,y) \leq 0, i = 1, \dots, p\},$$

with $\varphi(x) := \min_y \{f(x,y) \mid g_i(x,y) \leq 0, i = 1, \dots, p\}$ denoting the optimal value function of the lower-level problem. It happens, however, that in some particular settings (see, e.g., [15]), optimality conditions obtained via the KKT reformulation provide a richer set of information than their LLVF-counterpart. Moreover, the literature on mathematical programs with equilibrium/complementarity constraints (i.e., MPECs/MPCCs, for short) provides an important number of algorithmic schemes that could well be extended to the nonsmooth case. Further note that the LLVF reformulation being defined by an implicit constraint makes it quite complicate to construct viable CQs and algorithms. Hence, we are interested in this paper to extend some results about the KKT reformulation of the bilevel optimization problem in the smooth case (see, e.g., [9, 35, 38]) to the nonsmooth setting.

To proceed, we assume throughout the paper that *the lower-level problem is convex*, i.e., the functions $f(x, \cdot)$ and $g_i(x, \cdot)$ for $i = 1, \dots, p$, are convex for all x satisfying the upper-level constraints: $G_j(x) \leq 0$ for $j = 1, \dots, k$. If, additionally, we assume for a moment that all the functions involved in (P) are \mathcal{C}^1 (with f and g being \mathcal{C}^2), then it can take the form of a classical KKT reformulation

$$\min_{x,y,u} \{F(x,y) \mid \mathcal{L}(x,y,u) = 0, G_j(x) \leq 0, j = 1, \dots, k, \quad (1.3)$$

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

where $\mathcal{L}(x,y,u) := \nabla_y f(x,y) + \sum_{i=1}^p u_i \nabla_y g_i(x,y)$ stands for the derivative of the lower-level Lagrangian w.r.t. y . Next, we recall the link between (1.3) and (P) . For the remainder of this result, recall that for some \bar{x} , the Slater constraint qualification (CQ) is said to be satisfied at this point if it holds that

$$\{y \in \mathbb{R}^m \mid g_i(\bar{x}, y) < 0, i = 1, \dots, p\} \neq \emptyset. \quad (1.4)$$

To easily refer to the upper-level feasible points in what follows, we collect them in the set $X := \{x \in \mathbb{R}^n \mid G_j(x) \leq 0, j = 1, \dots, k\}$.

THEOREM 1.1 (relation of (P) to the KKT reformulation in the smooth case [4]). *Let (\bar{x}, \bar{y}) be a global (resp. local) optimal solution of (P) and assume that CQ (1.4) is satisfied at \bar{x} . Then, for each $\bar{u} \in \Lambda(\bar{x}, \bar{y})$, the point $(\bar{x}, \bar{y}, \bar{u})$ is a global (resp. local) optimal solution of problem (1.3). Conversely, let CQ (1.4) hold at all $x \in X$ (resp. at \bar{x}). Further assume that $(\bar{x}, \bar{y}, \bar{u})$ is a global optimal solution (resp. local optimal solution for all $\bar{u} \in \Lambda(\bar{x}, \bar{y})$) of problem (1.3). Then, (\bar{x}, \bar{y}) is a global (resp. local) optimal solution of (P) .*

Here, $\Lambda(\bar{x}, \bar{y})$ stands for the set of vectors u satisfying $\mathcal{L}(\bar{x}, \bar{y}, u) = 0$, $u \geq 0$, $g(x,y) \leq 0$ and $u^\top g(x,y) = 0$, i.e., the set of Lagrange multipliers for the lower-level program. To recall the definition of the stationarity concepts of (P) resulting from (1.3), we now introduce the following partition of the indices of the functions involved in the associated complementarity constraints:

$$\begin{aligned} \eta &:= \eta(\bar{x}, \bar{y}, \bar{u}) := \{i = 1, \dots, p \mid \bar{u}_i = 0, g_i(\bar{x}, \bar{y}) < 0\}, \\ \theta &:= \theta(\bar{x}, \bar{y}, \bar{u}) := \{i = 1, \dots, p \mid \bar{u}_i = 0, g_i(\bar{x}, \bar{y}) = 0\}, \\ \nu &:= \nu(\bar{x}, \bar{y}, \bar{u}) := \{i = 1, \dots, p \mid \bar{u}_i > 0, g_i(\bar{x}, \bar{y}) = 0\}. \end{aligned} \quad (1.5)$$

The middle set θ is known as the *biactive* or *degenerate index set*. The difference between the concepts is materialized by the structure of some components corresponding to θ . To further simplify the presentation, consider the following set of conditions which remain unchanged:

$$\nabla_x F(\bar{x}, \bar{y}) + \sum_{j=1}^k \alpha_j \nabla G_j(\bar{x}) + \sum_{i=1}^p \beta_i \nabla_x g_i(\bar{x}, \bar{y}) + \sum_{l=1}^m \gamma_l \nabla_x \mathcal{L}_l(\bar{x}, \bar{y}, \bar{u}) = 0, \quad (1.6)$$

$$\nabla_y F(\bar{x}, \bar{y}) + \sum_{i=1}^p \beta_i \nabla_y g_i(\bar{x}, \bar{y}) + \sum_{l=1}^m \gamma_l \nabla_y \mathcal{L}_l(\bar{x}, \bar{y}, \bar{u}) = 0, \quad (1.7)$$

$$\forall j = 1, \dots, k : \alpha_j \geq 0, \alpha_j G_j(\bar{x}) = 0, \quad (1.8)$$

$$\nabla_y g_\nu(\bar{x}, \bar{y}) \gamma = 0, \beta_\eta = 0. \quad (1.9)$$

Observe that the derivative of the function \mathcal{L} induces second order terms for functions involved in the lower-level problem:

$$\nabla \mathcal{L}(\bar{x}, \bar{y}, \bar{u})^\top \gamma = \begin{bmatrix} \sum_{l=1}^m \gamma_l (\nabla_{x y_l}^2 f(\bar{x}, \bar{y}) + \sum_{i=1}^p u_i \nabla_{x y_l}^2 g_i(\bar{x}, \bar{y})) \\ \sum_{l=1}^m \gamma_l (\nabla_{y y_l}^2 f(\bar{x}, \bar{y}) + \sum_{i=1}^p u_i \nabla_{y y_l}^2 g_i(\bar{x}, \bar{y})) \\ (\sum_{l=1}^m \gamma_l \nabla_{y_l} g_1(\bar{x}, \bar{y}), \dots, \sum_{l=1}^m \gamma_l \nabla_{y_l} g_p(\bar{x}, \bar{y}))^\top \end{bmatrix}. \quad (1.10)$$

Further note that the vector $\nabla_y g_\nu(\bar{x}, \bar{y}) \gamma$ in (1.9) represents the components of the last line of the right-hand-side of (1.10) for which $i \in \nu \subseteq \{1, \dots, p\}$. The reason for fully understanding the formula in (1.10) will become clear as from Section 3, when we develop the coderivative estimates of \mathcal{L} in the nonsmooth case.

DEFINITION 1.2 (C-, M-, and S-stationarity concepts in the smooth case). *For problem (P), a feasible point (\bar{x}, \bar{y}) is said to be:*

(i) **SP-C-STATIONARY** (resp. **P-C-STATIONARY**) *if for every $\bar{u} \in \Lambda(\bar{x}, \bar{y})$ (resp. for some $\bar{u} \in \Lambda(\bar{x}, \bar{y})$) we can find a triple $(\alpha, \beta, \gamma) \in \mathbb{R}^{k+p+m}$ such that the conditions (1.6)–(1.9) together with the following one are satisfied:*

$$\forall i \in \theta : \beta_i \sum_{l=1}^m \gamma_l \nabla_{y_l} g_i(\bar{x}, \bar{y}) \geq 0. \quad (1.11)$$

(ii) **SP-M-STATIONARY** (resp. **P-M-STATIONARY**) *if for every $\bar{u} \in \Lambda(\bar{x}, \bar{y})$ (resp. for some $\bar{u} \in \Lambda(\bar{x}, \bar{y})$) we can find a triple $(\alpha, \beta, \gamma) \in \mathbb{R}^{k+p+m}$ such that the conditions (1.6)–(1.9) together with the following one are satisfied:*

$$\forall i \in \theta : (\beta_i > 0 \wedge \sum_{l=1}^m \gamma_l \nabla_{y_l} g_i(\bar{x}, \bar{y}) > 0) \vee \beta_i \sum_{l=1}^m \gamma_l \nabla_{y_l} g_i(\bar{x}, \bar{y}) = 0. \quad (1.12)$$

(iii) **SP-S-STATIONARY** (resp. **P-S-STATIONARY**) *if for every $\bar{u} \in \Lambda(\bar{x}, \bar{y})$ (resp. for some $\bar{u} \in \Lambda(\bar{x}, \bar{y})$) we can find a triple $(\alpha, \beta, \gamma) \in \mathbb{R}^{k+p+m}$ such that the conditions (1.6)–(1.9) together with the following one are satisfied:*

$$\forall i \in \theta : \beta_i \geq 0 \wedge \sum_{l=1}^m \gamma_l \nabla_{y_l} g_i(\bar{x}, \bar{y}) \geq 0. \quad (1.13)$$

Note for instance, that the term “SP-C-STATIONARY” stands for STRONG P-C-STATIONARY. The “P” refers to the stationarity concepts of problem (P) in (1.1) as

oppose to “ P_o ” and “ P_p ” which are used in [7, 8, 38] to symbolize the counterparts of these conditions for the *original optimistic* and *pessimistic* bilevel programs, respectively. Similar statements can be made for the other stationarity concepts. Obviously, we have the following relationships:

$$\begin{array}{ccccc} \text{SP-S-stationary} & \implies & \text{SP-M-stationary} & \implies & \text{SP-C-stationary} \\ \downarrow & & \downarrow & & \downarrow \\ \text{P-S-stationary} & \implies & \text{P-M-stationary} & \implies & \text{P-C-stationary} \end{array}$$

These concepts were introduced and justified in [9] (also see [38] for more details) under appropriate CQs. For stationarity concepts of related MPECs, the interested reader is referred, for example, to [13, 30, 33], where many other classes of stationarity conditions are discussed.

The main aim of the current paper is to extend Theorem 1.1 and the stationarity concepts of Definition 1.2 to the case where the functions involved in (P) are nonsmooth. To proceed, we use notions from variational analysis that are introduced in the next section. In Section 3, we develop tools in the framework of nonsmooth parametric optimization, including the upper semicontinuity of the nonsmooth counterpart of Λ , as well as coderivative estimates for the partial subdifferential Lagrangian of lower-level problem (PSLLL). The latter point can essentially be formalized as the extension of the formula (1.10) to the nonsmooth framework. In Section 4, we discussed the nonsmooth version of Theorem 1.1, while using developments from the previous section. The remaining sections 5, 6 and 7 are devoted to the introduction and justification of nonsmooth counterparts of the C-, M- and S-stationarity concepts. We mainly use the basic/Mordukhovich subdifferential, as it allows good calculus rules and generates sharper optimality conditions. The Clarke subdifferential is partly involved just when the C-stationarity is in consideration or when the plus/minus symmetry is needed. Final comments and extensions of the results developed in the paper are discussed in Section 8.

Throughout the paper, we may use 0_n for the origin of \mathbb{R}^n in situations where some confusion may be possible. For any vector a , we could use a_b (with $b = \nu, \eta$ or θ) to symbolize $(a_i)_{i \in b}$. Finally, for two vectors a and b , we may also write (a, b) instead of $(a, b)^\top$ to simplify notations.

2. Basic definitions and concepts from variational analysis. For a closed subset C of \mathbb{R}^n , the basic (or limiting/Mordukhovich) normal cone to C at one of its points \bar{x} is the set

$$N_C(\bar{x}) := \{v \in \mathbb{R}^n \mid \exists v_k \rightarrow v, x_k \rightarrow \bar{x} (x_k \in C) : v_k \in \widehat{N}_C(x_k)\}, \quad (2.1)$$

where \widehat{N}_C denotes the dual of the contingent/Bouligand tangent cone to C . Note that if $C := \psi^{-1}(\Xi)$, where $\Xi \subseteq \mathbb{R}^m$ is a closed set and $\psi: \mathbb{R}^n \rightarrow \mathbb{R}^m$ a Lipschitz continuous function around \bar{x} , then we have

$$N_C(\bar{x}) \subseteq \bigcup \{\partial \langle v, \psi \rangle(\bar{x}) \mid v \in N_\Xi(\psi(\bar{x}))\}, \quad (2.2)$$

provided the following basic-type qualification condition (QC) is satisfied at \bar{x} :

$$[0 \in \partial \langle v, \psi \rangle(\bar{x}), v \in N_\Xi(\psi(\bar{x}))] \implies v = 0, \quad (2.3)$$

cf. [20] or [29]. Equality holds in (2.2), provided that the set Ξ is normally regular at $\psi(\bar{x})$, i.e., $N_\Xi(\psi(\bar{x})) = \widehat{N}_\Xi(\psi(\bar{x}))$. This is obviously the case if Ξ is a convex set.

In (2.2) and (2.3), the term $\partial\langle v, \psi \rangle(\bar{x})$ refers to the basic (or limiting/Mordukhovich) subdifferential of the function $x \mapsto \sum_{i=1}^m v_i \psi_i(x)$ at the point \bar{x} . Generally speaking, if $\psi: \mathbb{R}^n \rightarrow \mathbb{R}$, then the basic subdifferential of ψ at \bar{x} can be defined by

$$\partial\psi(\bar{x}) := \{\xi \in \mathbb{R}^n \mid (\xi, -1) \in N_{\text{epi}\psi}(\bar{x}, \psi(\bar{x}))\}.$$

Here, $\text{epi}\psi$ stands for the epigraph of ψ . If $\psi(x) := d_C(x)$, i.e., the distance function from x to the nonempty closed set $C \subseteq \mathbb{R}^n$, then we have

$$\partial\psi(\bar{x}) = N_C(\bar{x}) \cap \mathbb{B} \quad (2.4)$$

with \mathbb{B} denoting the unit ball centered at the origin of \mathbb{R}^n , cf. [29, Example 8.53]. Furthermore, in a more general framework, if ψ is a Lipschitz continuous function around \bar{x} , then we can also define the convexified (or Clarke) subdifferential of ψ at \bar{x}

$$\bar{\partial}\psi(\bar{x}) := \text{co } \partial\psi(\bar{x}).$$

In the case where ψ is a convex function, then $\partial\psi(\bar{x})$ and $\bar{\partial}\psi(\bar{x})$ coincide with the subdifferential in the sense of convex analysis.

It is worth mentioning here that the inclusion in (2.2) remains valid if the weaker *calmness property* holds for the set-valued map $\Psi(v) := \{x \in \mathbb{R}^n \mid \psi(x) + v \in \Xi\}$, cf. [14, Theorem 4.1]. A set-valued map $\Psi[\mathbb{R}^n \rightrightarrows \mathbb{R}^m]$ will be said to be calm at some point $(\bar{x}, \bar{y}) \in \text{gph } \Psi := \{(x, y) \in \mathbb{R}^n \times \mathbb{R}^m \mid y \in \Psi(x)\}$, if there exist neighborhoods U of \bar{x} , V of \bar{y} , and a constant $\kappa > 0$ such that

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

Another continuity property of set-valued maps useful in this paper is the *inner semicompactness*. $\Psi[\mathbb{R}^n \rightrightarrows \mathbb{R}^m]$ will be said to be inner semicompact at a point \bar{x} , if for every sequence $x_k \rightarrow \bar{x}$, there is a sequence of $y_k \in \Psi(x_k)$ that contains a convergent subsequence as $k \rightarrow \infty$. Observe that this property automatically holds at \bar{x} , if the map Ψ is *uniformly bounded* around this point, i.e., there exists a neighborhood U of \bar{x} such that $\Psi(U)$ is bounded. If the set $\Psi(\bar{x})$ is closed, then we say that Ψ is *upper semicontinuous* at \bar{x} , if for every sequence $x_k \rightarrow \bar{x}$, each sequence of $y_k \in \Psi(x_k)$ has an accumulation point contained in $\Psi(\bar{x})$. Obviously, Ψ is inner semicompact at \bar{x} if it is upper semicontinuous at this point with $\Psi(\bar{x})$ closed.

Finally, we introduce the notion of *coderivative* that will play a central role in this paper. For a set-valued map $\Psi[\mathbb{R}^n \rightrightarrows \mathbb{R}^m]$, the coderivative of Ψ at some point $(\bar{x}, \bar{y}) \in \text{gph } \Psi$ is a positively homogeneous set-valued mapping $D^*\Psi(\bar{x}|\bar{y}): \mathbb{R}^m \rightrightarrows \mathbb{R}^n$, defined by

$$D^*\Psi(\bar{x}|\bar{y})(v) := \{u \in \mathbb{R}^n \mid (u, -v) \in N_{\text{gph } \Psi}(\bar{x}, \bar{y})\}, \quad (2.5)$$

for all $y \in \mathbb{R}^m$. Here, $N_{\text{gph } \Psi}$ denotes the basic normal cone (2.1) to $\text{gph } \Psi$. It is worth mentioning that this concept was first introduced in the paper [22]. Further note that more details on the material briefly discussed in this section can be found in the books [21, 29] and references therein.

3. Parametric nonsmooth optimization. In this section, we are interested in the parametric optimization problem

$$\min_y \{f(x, y) \mid g_i(x, y) \leq 0, i = 1, \dots, p\} \quad (3.1)$$

defining our lower-level problem in the bilevel optimization problem (P) . The functions $f[\mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}]$ and $g_i[\mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}]$ for $i = 1, \dots, p$, are assumed to be locally Lipschitz continuous and not necessarily differentiable. Moreover, we assume throughout the section that *problem (3.1) is convex*, i.e., the functions $f(x, \cdot)$ and $g_i(x, \cdot)$, $i = 1, \dots, p$, are convex for all $x \in \mathbb{R}^n$. Our aim here is to provide some properties of problem (3.1) which are useful in the analysis of nonsmooth bilevel programs via the KKT reformulation.

Considering the fact that problem (3.1) is convex, and denoting by $S(x)$ its optimal solution set for a given x , we have from [28, Corollary 28.3.1] that

$$y \in S(x) \text{ if and only if there exists } u \text{ such that:} \quad (3.2)$$

$$\begin{cases} 0 \in \partial_y f(x, y) + \sum_{i=1}^p u_i \partial_y g_i(x, y), \\ u_i \geq 0, g_i(x, y) \leq 0, u_i g_i(x, y) = 0, i = 1, \dots, p, \end{cases}$$

provided CQ (1.4) holds at x . Here, $\partial_y \psi(x, y)$ stands for the subdifferential in the sense of convex analysis of the function $\psi(x, \cdot)$ at y . From here on, the set-valued map

$$\Lambda(x, y) := \left\{ u \in \mathbb{R}^p \mid \begin{aligned} &0 \in \partial_y f(x, y) + \sum_{i=1}^p u_i \partial_y g_i(x, y), \\ &u_i \geq 0, g_i(x, y) \leq 0, u_i g_i(x, y) = 0, i = 1, \dots, p \end{aligned} \right\} \quad (3.3)$$

denotes the nonsmooth counterpart of the set of Lagrange multipliers discussed in Section 1. Next, we establish that this map is closed and upper semicontinuous. It is well-known that the set-valued map Λ (3.3) is upper semicontinuous under a regularity condition, see, e.g., [27, Theorem 3.2]. But since we were unable to find a reference where it is shown in the nonsmooth case, we include a proof here.

To proceed, recall that a function $\psi[\mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}]$ defined by $(x, y) \mapsto \psi(x, y)$ is said to be locally Lipschitz continuous around \bar{y} uniformly in x if there exist a number $\ell > 0$ (independent of x) and a neighborhood V of \bar{y} in \mathbb{R}^m such that we have $|\psi(x, y) - \psi(x, y')| \leq \ell \|y - y'\|$ for all $y, y' \in V$, $x \in \mathbb{R}^n$.

THEOREM 3.1 (closedness and upper semicontinuity of Λ). *Let the functions f and g_i , $i = 1, \dots, p$ be Lipschitz continuous around \bar{y} uniformly in x . Then, the set-valued mapping Λ (3.3) is closed. If, in addition, CQ (1.4) holds at \bar{x} , then, Λ is upper semicontinuous at (\bar{x}, \bar{y}) , for all $y \in \mathbb{R}^m$.*

Proof. Consider a sequence $(x^k, y^k, u^k) \in \text{gph } \Lambda$ such that $(x^k, y^k, u^k) \rightarrow (\bar{x}, \bar{y}, \bar{u})$. Then, by the definition of Λ , it holds that

$$\begin{aligned} &0 \in \partial_y f(x^k, y^k) + \sum_{i=1}^p u_i^k \partial_y g_i(x^k, y^k) \quad \text{for all } k \in \mathbb{N}, \\ &u_i^k \geq 0, g_i(x^k, y^k) \leq 0, u_i^k g_i(x^k, y^k) = 0 \quad \text{for all } k \in \mathbb{N}, i = 1, \dots, p. \end{aligned} \quad (3.4)$$

The first line of (3.4) can be equivalently replaced by

$$0 \in \partial_y f(x^k, y^k) + \sum_{i=1}^p (u_i^k - \bar{u}_i) \partial_y g_i(x^k, y^k) + \sum_{i=1}^p \bar{u}_i \partial_y g_i(x^k, y^k). \quad (3.5)$$

Since the functions $g_i(x, \cdot)$, $i = 1, \dots, p$, are Lipschitz continuous around \bar{y} uniformly in x , it holds that

$$(u_i^k - \bar{u}_i) \partial_y g_i(x^k, y^k) \subseteq \ell_i |u_i^k - \bar{u}_i| \mathbb{B}_m \quad \text{for all } k \in \mathbb{N}, i = 1, \dots, p, \quad (3.6)$$

where ℓ_i , $i = 1, \dots, p$, denote the uniformly Lipschitz constants of $g_i(x, \cdot)$, $i = 1, \dots, p$, respectively. Also note that \mathbb{B}_m stands for the unit ball of \mathbb{R}^m centered at the origin. Passing to the limit in (3.5) and in the second line of (3.4), we arrive at

$$\begin{aligned} &0 \in \partial_y f(\bar{x}, \bar{y}) + \sum_{i=1}^p \bar{u}_i \partial_y g_i(\bar{x}, \bar{y}), \\ &\bar{u}_i \geq 0, g_i(\bar{x}, \bar{y}) \leq 0, \bar{u}_i g_i(\bar{x}, \bar{y}) = 0, i = 1, \dots, p, \end{aligned}$$

while taking into account that $\partial_y f$ and $\partial_y g_i$, $i = 1, \dots, p$ are upper semicontinuous, as the functions f and g_i , $i = 1, \dots, p$ are uniformly Lipschitz continuous around \bar{y} , cf. [3, Chapter 2]. This means that $(\bar{x}, \bar{y}, \bar{u}) \in \text{gph } \Lambda$. Hence, Λ is closed.

For the upper semicontinuity of Λ at (\bar{x}, \bar{y}) , suppose that, there are sequences $(x^k, y^k) \rightarrow (\bar{x}, \bar{y})$ and $u^k \in \Lambda(x^k, y^k)$ with $\|u^k\| \rightarrow \infty$. Now consider the sequence $v_i^k := u_i^k / \|u^k\|$ for all $k \in \mathbb{N}$ and $i = 1, \dots, p$. Obviously, we have $\|v^k\| = 1$ for all k . Hence, we can find a subsequence of v^k that we denote similarly (provided there is no confusion) which converges to some v with $\|v\| = 1$. On the other hand, note that

$$\begin{aligned} 0 &\in \frac{1}{\|u^k\|} \partial_y f(x^k, y^k) + \sum_{i=1}^p v_i^k \partial_y g_i(x^k, y^k) \quad \text{for all } k \in \mathbb{N}, \\ v_i^k &\geq 0, g_i(x^k, y^k) \leq 0, v_i^k g_i(x^k, y^k) = 0 \quad \text{for all } k \in \mathbb{N}; i = 1, \dots, p. \end{aligned} \quad (3.7)$$

Similarly to the previous proof, the first line of this system can be rewritten as

$$0 \in \frac{1}{\|u^k\|} \partial_y f(x^k, y^k) + \sum_{i=1}^p (v_i^k - v_i) \partial_y g_i(x^k, y^k) + \sum_{i=1}^p v_i \partial_y g_i(x^k, y^k). \quad (3.8)$$

The functions $f(x, \cdot)$ and $g_i(x, \cdot)$, $i = 1, \dots, p$, being Lipschitz continuous around \bar{y} uniformly in x , it holds that

$$\begin{aligned} \frac{1}{\|u^k\|} \partial_y f(x^k, y^k) &\subseteq \frac{\ell_0}{\|u^k\|} \mathbb{B}_m \quad \text{for all } k \in \mathbb{N}, \\ (v_i^k - v_i) \partial_y g_i(x^k, y^k) &\subseteq \ell_i |v_i^k - v_i| \mathbb{B}_m \quad \text{for all } k \in \mathbb{N}, i = 1, \dots, p, \end{aligned}$$

where ℓ_0 and ℓ_i , $i = 1, \dots, p$ denote the uniformly Lipschitz constants of $f(x, \cdot)$ and $g_i(x, \cdot)$, $i = 1, \dots, p$, respectively. Hence, passing to the limit in (3.8) and in the second line of (3.7), we have

$$\begin{aligned} 0 &\in \sum_{i=1}^p v_i \partial_y g_i(\bar{x}, \bar{y}), \\ v_i &\geq 0, g_i(\bar{x}, \bar{y}) \leq 0, v_i g_i(\bar{x}, \bar{y}) = 0, i = 1, \dots, p. \end{aligned} \quad (3.9)$$

Thus, we have $v = 0$ (since CQ (1.4) holds at \bar{x}), which is a contradiction to the fact that $\|v\| = 1$. In addition to the closedness of Λ , it follows that this map is upper semicontinuous at (\bar{x}, \bar{y}) . \square

It appears from the proof of this theorem that the uniform boundedness of the mappings $\partial_y f$ and $\partial_y g_i$ for $i = 1, \dots, p$ is enough to guaranty that inclusion (3.6) is satisfied. Hence, this assumption will be sufficient for many subsequent results in the next sections. Further recall that Λ is also inner semicompact under the assumptions made in the above theorem.

For the rest of the section, we are mainly interested in estimating the *coderivative* (2.5) of the set-valued mapping \mathcal{L} that we label as the *partial subdifferential of the lower-level Lagrangian* (PSLLL) and which is defined by

$$\mathcal{L}(x, y, u) := \partial_y f(x, y) + \sum_{i=1}^p u_i \partial_y g_i(x, y). \quad (3.10)$$

PROPOSITION 3.2 (coderivative estimate of a Cartesian product set-valued map). *Consider the set-valued mappings $\Psi_i : \mathbb{R}^n \rightrightarrows \mathbb{R}^q$ for $i = 1, \dots, p$, and define a Cartesian product mapping $\Psi : \mathbb{R}^n \rightrightarrows \mathbb{R}^{q \times p}$ by*

$$\Psi(x) := \prod_{i=1}^p \Psi_i(x) = \Psi_1(x) \times \dots \times \Psi_p(x).$$

Assume that $\text{gph } \Psi_i$, $i = 1, \dots, p$, is closed and the following qualification condition

$$\left[\sum_{i=1}^p v^i = 0, v^i \in D^* \Psi_i(\bar{x}|\bar{y}^i)(0), i = 1, \dots, p \right] \implies v^1 = \dots = v^p = 0 \quad (3.11)$$

is satisfied at (\bar{x}, \bar{y}) with $\bar{y} := (\bar{y}^i)_{i=1}^p \in \Psi(\bar{x})$. Then, for any $v := (v^i)_{i=1}^p \in \prod_{i=1}^p \mathbb{R}^q$,

$$D^* \Psi(\bar{x}|\bar{y})(v) \subseteq \sum_{i=1}^p D^* \Psi_i(\bar{x}|\bar{y}^i)(v^i). \quad (3.12)$$

Equality holds in (3.12), if $\text{gph } \Psi_i$ is normally regular at (\bar{x}, \bar{y}^i) , for $i = 1, \dots, p$.

Proof. Observe that the graph of Ψ can take the form $\text{gph } \Psi = \psi^{-1}(\Xi)$ where

$$\psi(x, y) := \prod_{i=1}^p \psi_i(x, y) \text{ and } \Xi := \prod_{i=1}^p \text{gph } \Psi_i. \quad (3.13)$$

with $\psi_i(x, y) := (x, y^i)$ for $i = 1, \dots, p$. The set Ξ is closed given that for each $i = 1, \dots, p$, $\text{gph } \Psi_i$ is assumed to be closed. Now consider a vector w such that

$$w^\top = \prod_{i=1}^p (u^i, v^i) \in \prod_{i=1}^p N_{\text{gph } \Psi_i}(\bar{x}, \bar{y}^i) = N_{\Xi}(\psi(\bar{x}, \bar{y})), \quad (3.14)$$

then we have the following calculations

$$\nabla \psi(\bar{x}, \bar{y})^\top w = \sum_{i=1}^p \nabla \psi_i(\bar{x}, \bar{y})^\top (u^i, v^i)^\top = \left[\sum_{i=1}^p u^i, v^1, \dots, v^p \right]^\top. \quad (3.15)$$

Thus the qualification condition (2.3) in the framework of (3.13) reduces to (3.11), while considering the definition of the coderivative in (2.5). Furthermore, combining (2.2), (3.14) and (3.15), it holds that

$$N_{\text{gph } \Psi}(\bar{x}, \bar{y}) \subseteq \left\{ \left[\sum_{i=1}^p u^i, v^1, \dots, v^p \right]^\top \mid (u^i, v^i) \in N_{\text{gph } \Psi_i}(\bar{x}, \bar{y}^i), i = 1, \dots, p \right\}.$$

Considering once more the interplay in (2.5) between the coderivative and the normal cone, (3.12) follows from the latter inclusion. As for the equality, note that Ξ in (3.13) is regular at $\psi(\bar{x}, \bar{y})$ provided each $\text{gph } \Psi_i$ is regular at (\bar{x}, \bar{y}^i) for $i = 1, \dots, p$. \square

It is worth mentioning that the normal regularity assumption required to get equality in (3.11) is very restrictive and does not hold for important classes of mappings including subdifferentials and normal cone maps. Further details on this topic can be found in the book [21, Chapter 1].

We are now ready to provide an upper bound for the coderivative of the multifunction (3.10) in terms of the functions involved in the parametric problem (3.1).

THEOREM 3.3 (coderivative estimate for the PSLLL set-valued mapping). *Assume that the set-valued mappings $\partial_y f$ and $\partial_y g_i$ for $i = 1, \dots, p$, are closed and uniformly bounded around (\bar{x}, \bar{y}) . Furthermore, let $\bar{v} \in \mathcal{L}(\bar{x}, \bar{y}, \bar{u})$ and assume that for all $t := (t^0, t^1, \dots, t^p)$ with $t^0 \in \partial_y f(\bar{x}, \bar{y})$, $t^i \in \partial_y g_i(\bar{x}, \bar{y})$ for $i = 1, \dots, p$, and $t^0 + \sum_{i=1}^p \bar{u}_i t^i = \bar{v}$, the following qualification condition is satisfied:*

$$\left[v^0 \in D^*(\partial_y f)((\bar{x}, \bar{y})|t^0)(0), v^i \in D^*(\partial_y g_i)((\bar{x}, \bar{y})|t^i)(0), i = 1, \dots, p \mid v^0 + v^1 + \dots + v^p = 0 \right] \implies v^0 = v^1 = \dots = v^p = 0. \quad (3.16)$$

Then for all $v \in \mathbb{R}^m$, the coderivative of the mapping from (3.10) is estimated by

$$D^* \mathcal{L}((\bar{x}, \bar{y}, \bar{u})|\bar{v})(v) \subseteq \bigcup_{\substack{t: t^0 + \sum_{i=1}^p \bar{u}_i t^i = \bar{v} \\ t^0 \in \partial_y f(\bar{x}, \bar{y}), t^i \in \partial_y g_i(\bar{x}, \bar{y})}} \left\{ \left[D^*(\partial_y f)((\bar{x}, \bar{y})|t^0)(v) \right. \right. \\ \left. \left. + \sum_{i=1}^p D^*(\partial_y g_i)((\bar{x}, \bar{y})|t^i)(\bar{u}_i v) \right] \times \left\{ \left(\sum_{l=1}^m t_l^1 v_l, \dots, \sum_{l=1}^m t_l^p v_l \right)^\top \right\} \right\}.$$

Proof. Start by observing that the set-valued mapping \mathcal{L} from (3.10) can be represented as the composition of a \mathcal{C}^1 function $\varphi [\mathbb{R}^{m(1+p)} \times \mathbb{R}^p \rightarrow \mathbb{R}^m]$ and the multifunction $\Psi [\mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p \rightrightarrows \mathbb{R}^{m(1+p)} \times \mathbb{R}^p]$:

$$\begin{aligned} \mathcal{L}(x, y, u) &= \varphi \circ \Psi(x, y, u) \\ \text{with } \begin{cases} \varphi(t, u) &:= t^0 + \sum_{i=1}^p u_i t^i, \\ \Psi(x, y, u) &:= [\Psi_o(x, y), u] := \{(t, u) \mid t \in \Psi_o(x, y)\}, \\ \Psi_o(x, y) &:= \partial_y f(x, y) \times \partial_y g_1(x, y) \times \dots \times \partial_y g_p(x, y). \end{cases} \end{aligned} \quad (3.17)$$

Note that the set-valued map $\Psi_o [\mathbb{R}^n \times \mathbb{R}^m \rightrightarrows \mathbb{R}^{m(1+p)}]$. To apply the chain rule from [20, Corollary 5.3] to the above expression of \mathcal{L} , also note that the set-valued mapping $\Psi \cap \varphi^{-1} [\mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p \times \mathbb{R}^m \rightrightarrows \mathbb{R}^{m(1+p)} \times \mathbb{R}^p]$ defined by

$$\Psi(x, y, u) \cap \varphi^{-1}(v) := \{(t, u) \mid t \in \Psi_o(x, y), \varphi(t, u) = v\}$$

is uniformly bounded around (\bar{x}, \bar{y}) since the set-valued mappings $\partial_y f$ and $\partial_y g_i$ for $i = 1, \dots, p$, are all uniformly bounded around the same point. Furthermore, the multifunction Ψ is closed, since $\partial_y f$ and $\partial_y g_i$ for $i = 1, \dots, p$, are closed. Thus we have by the aforementioned chain rule that

$$D^* \mathcal{L}((\bar{x}, \bar{y}, \bar{u})|\bar{v})(v) \subseteq \bigcup_{(t, u) \in \Psi(\bar{x}, \bar{y}, \bar{u}) \cap \varphi^{-1}(\bar{v})} \left[D^* \Psi((\bar{x}, \bar{y}, \bar{u})|(t, u))(\nabla \varphi(t, u)^\top v) \right]. \quad (3.18)$$

Basic calculations generate the following expression for $\nabla \varphi(t, u)^\top v$:

$$\nabla \varphi(t, u)^\top v = \left[v, u_1 v, \dots, u_p v, \sum_{i=1}^m t_i^1 v_i, \dots, \sum_{i=1}^m t_i^p v_i \right]^\top. \quad (3.19)$$

In the next step, we estimate the coderivative of Ψ in terms of that of Ψ_o . To proceed, note that $\text{gph } \Psi = \psi^{-1}(\Xi)$ with ψ and Ξ respectively defined by

$$\psi(x, y, u, t, v) := (x, y, t, u - v)^\top \text{ and } \Xi := \text{gph } \Psi_o \times \{0_p\}. \quad (3.20)$$

Consider a quadruple $(a, b, c, d) \in N_\Xi(\psi(\bar{x}, \bar{y}, \bar{u}, t, v)) = N_{\text{gph } \Psi_o}(\bar{x}, \bar{y}, t) \times \mathbb{R}^p$, then one can easily check that

$$\nabla \psi(\bar{x}, \bar{y}, \bar{u}, t, v)^\top (a, b, c, d) = [a, b, d, c, -d]^\top. \quad (3.21)$$

Thus the qualification condition (2.3) holds in the framework of (3.20). Hence, for $(x^*, y^*, u^*) \in D^* \Psi((\bar{x}, \bar{y}, \bar{u})|(t, u))(s, w)$, we have by (2.2) while considering equality (2.5) that there exists $(a, b, c, d) \in N_{\text{gph } \Psi_o}(\bar{x}, \bar{y}, t) \times \mathbb{R}^p$ such that $x^* = a$, $y^* = b$, $u^* = d$, $s = -c$ and $w = d$, cf. equality (3.21). Clearly, this means that we have the following inclusion:

$$D^* \Psi((\bar{x}, \bar{y}, \bar{u})|(t, u))(s, w) \subseteq D^* \Psi_o((\bar{x}, \bar{y})|t)(s) \times \{w\}. \quad (3.22)$$

If we insert the value of (3.19) and the upper estimate of (3.22) in inclusion (3.18), we arrive at the following upper bound for the coderivative of \mathcal{L} :

$$\bigcup_{t \in \Psi_o(\bar{x}, \bar{y}), \varphi(t, u) = \bar{v}} \left[D^* \Psi_o((\bar{x}, \bar{y})|t)(v, u_1 v, \dots, u_p v) \right. \\ \left. \times \left\{ \left(\sum_{i=1}^m t_i^1 v_i, \dots, \sum_{i=1}^m t_i^p v_i \right)^\top \right\} \right]. \quad (3.23)$$

Since the graphs of $\partial_y f$ and $\partial_y g_i$, $i = 1, \dots, p$ are assumed to be closed, we apply Proposition 3.2 to Ψ_o , and it follows under the qualification condition (3.16) that

$$D^* \Psi_o((\bar{x}, \bar{y})|t)(v, u_1 v, \dots, u_p v) \subseteq D^*(\partial_y f)((\bar{x}, \bar{y})|t^0)(v) \\ + \sum_{i=1}^p D^*(\partial_y g_i)((\bar{x}, \bar{y})|t^i)(u_i v).$$

The inclusion in the theorem is obtained by inserting the latter one in (3.23). \square

REMARK 3.4 (estimate of the coderivative of \mathcal{L} via the sum rule). The coderivative sum rule (see [20]) could also be used to compute $D^* \mathcal{L}$ while considering \mathcal{L} as the sum of $\partial_y f$ and $(x, y, u) \mapsto \sum_{i=1}^p u_i \partial_y g_i(x, y)$. In this case, however, the chain rule may still be invoked to estimate the coderivative of the latter map. Thus the approach in Theorem 3.3 is more efficient as it allows us to avoid such a lengthy process of combining the sum and chain rules successively.

Considering the structure of the above coderivative estimate of \mathcal{L} and the definition of the notion of *second order* basic subdifferential (also known as Mordukhovich or generalized Hessian) of a function $\psi [\mathbb{R}^n \rightarrow \mathbb{R}]$,

$$\partial^2 \psi(\bar{x}|\bar{y})(v) := D^*(\partial \psi)(\bar{x}|\bar{y})(v) \text{ for } v \in \mathbb{R}^n,$$

it would be interesting to write $D^* \mathcal{L}$ in terms of second order subdifferentials of the functions involved in (3.1). Thus applying [24, Theorem 3.1], we get the following expression of $D^*(\partial_y f)$ in terms of the second order subdifferential of the function f :

$$D^*(\partial_y f)((\bar{x}, \bar{y})|t^0)(v) = \partial^2 f((\bar{x}, \bar{y})|(0, t^0))(0, v).$$

Similar formulae can be written for $D^*(\partial_y g_i)(\bar{x}, \bar{y})$, $i = 1, \dots, p$. Hence, we obtain

$$D^* \mathcal{L}((\bar{x}, \bar{y}, \bar{u})|\bar{v})(v) \subseteq \bigcup_{\substack{t: t^0 + \sum_{i=1}^p \bar{u}_i t^i = \bar{v} \\ t^0 \in \partial_y f(\bar{x}, \bar{y}), t^i \in \partial_y g_i(\bar{x}, \bar{y})}} \left\{ \left[\partial^2 f((\bar{x}, \bar{y})|(0, t^0))(0, v) \right. \right. \\ \left. \left. + \sum_{i=1}^p \partial^2 g_i((\bar{x}, \bar{y})|(0, t^i))(0, \bar{u}_i v) \right] \times \left\{ \left(\sum_{i=1}^m t_i^1 v_i, \dots, \sum_{i=1}^m t_i^p v_i \right)^\top \right\} \right\}, \quad (3.24)$$

provided all the assumptions of Theorem 3.3 are satisfied. In the same vein, the sets $\partial^2 f((\bar{x}, \bar{y})|(0, t^0))(0)$ and $\partial^2 g_i((\bar{x}, \bar{y})|(0, t^i))(0)$, $i = 1, \dots, p$, can replace the coderivative terms in QC (3.16). The resulting QC and its original form in (3.16) are automatically satisfied when the functions f and g_i , $i = 1, \dots, p$, are \mathcal{C}^2 . Moreover, the inclusion in (3.3) reduces to formula (1.10), in the latter situation.

4. KKT reformulation in nonsmooth bilevel programming. If the lower-level problem in the bilevel optimization problem (1.1) is replaced with its KKT conditions (3.2), we get the following natural extension of the KKT reformulation (1.3) to the framework of the nonsmooth bilevel program (P) , where the functions involved are Lipschitz continuous and not necessarily differentiable:

$$\min_{x, y, u} \{ F(x, y) \mid 0 \in \mathcal{L}(x, y, u), G_j(x) \leq 0, j = 1, \dots, k, \\ u_i \geq 0, g_i(x, y) \leq 0, u_i g_i(x, y) = 0, i = 1, \dots, p \}. \quad (4.1)$$

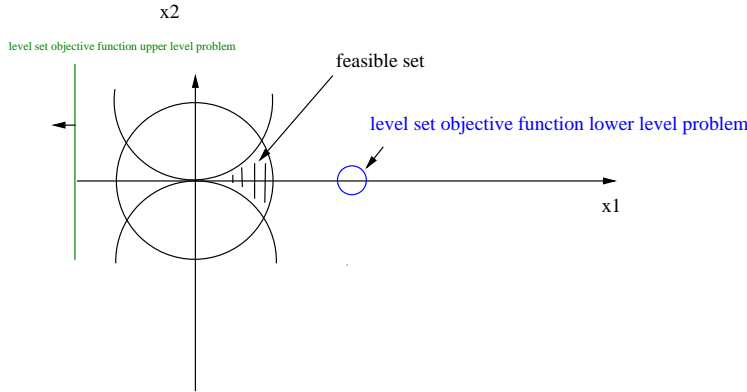


FIG. 4.1. Fritz-John versus KKT reformulation

Note the presence of the inclusion $0 \in \mathcal{L}(x, y, u)$ (with the set-valued mapping \mathcal{L} defined in (3.10)) stressing that unlike in the smooth case, the KKT reformulation of the nonsmooth bilevel program is a special class of mathematical programs with set-valued inclusion constraint.

It is well-known that regularity conditions as the MFCQ are violated at every feasible point of the KKT reformulation [30]. The generic structure of the feasible set of the bilevel optimization problem with only one upper-level variable is investigated in [17]. Moreover, the linear independence constraint qualification is not a generic regularity condition in the lower-level problem [4], at least in the case when the lower-level constraints depend on the parameter. In [1], the more general problem where the KKT conditions of the lower-level problem are replaced by the Fritz-John conditions is considered and the generic structure of the feasible region is studied. It is important to mention that the resulting problem extremely modifies the bilevel optimization problem. Consider, for example, the simple bilevel problem

$$\min \{x_1 : x \in \operatorname{argmin}\{\|x - (2 \ 0)^\top\| : \|x\| \leq 1, x_2 \leq x_1^2, x_2 \geq -x_1^2\}\}.$$

The problem is sketched in Figure 4.1. Here, if we use the KKT conditions to replace the lower-level problem, we get the unique feasible point $x = (1 \ 0)^\top$. If, instead, we replace the lower-level problem with the Fritz-John conditions, the point $x = (0, 0)^\top$ becomes feasible and, hence, optimal. Note that there are two Fritz-John points in this problem.

Optimization problems of the form (4.1) have been investigated for example in [2] under very general settings. In this paper, we are interested in developing necessary optimality conditions tailored to (4.1) while also taking into account the nature of the other constraints which are of the complementarity type. For results on mathematical programs with complementarity constraints (MPCCs) with smooth data, the reader is referred for example to [13, 33]. Nonsmooth MPCCs were recently considered in [25, 26] while using generalized differentiation tools by Clarke and Michel-Penot. Not only the model in the latter papers does not encompass our problem in (4.1) which contains both set-valued and complementarity constraints, but we rather provide the most natural extensions of the stationarity conditions of (4.1) from a completely different perspective.

Next, we first establish relationships in terms of optimal solutions between the bilevel program from (1.1) and its KKT reformulation (4.1). Note that the LLVF reformulation (1.2) is completely equivalent to the initial problem. This is unfortunately not the case for the KKT reformulation, as recently observed in [4] in the smooth case. It turns out that for KKT reformulation we need additional assumptions to establish a workable relationship with (P). This is even more true for the nonsmooth case as we need even more conditions to obtain a local optimal solution of (P) from (4.1). The following theorems are extensions of results from [4] (see Theorem 1.1) obtained there under the smooth setting.

THEOREM 4.1 (local relation of (P) to its KKT reformulation in the nonsmooth case). *Let (\bar{x}, \bar{y}) be a local optimal solution of (P) and assume that CQ (1.4) holds at \bar{x} . Then, for all $\bar{u} \in \Lambda(\bar{x}, \bar{y})$, the triple $(\bar{x}, \bar{y}, \bar{u})$ is also a local optimal solution of (4.1). Conversely, let $(\bar{x}, \bar{y}, \bar{u})$ be a local optimal solution of problem (4.1) for all $\bar{u} \in \Lambda(\bar{x}, \bar{y})$. Further assume that CQ (1.4) holds at \bar{x} , while the functions f and g_i , $i = 1, \dots, p$ are Lipschitz continuous around \bar{y} uniformly in x . Then, (\bar{x}, \bar{y}) is a local optimal solution of problem (P).*

Proof. For the first implication (\Rightarrow), assume that there exists $\tilde{u} \in \Lambda(\bar{x}, \bar{y})$ such that $(\bar{x}, \bar{y}, \tilde{u})$ is not a local optimal solution of problem (4.1). Then, we can find a sequence (x^k, y^k, u^k) with $x^k \rightarrow \bar{x}$, $y^k \rightarrow \bar{y}$, and $u^k \rightarrow \tilde{u}$ such that

$$F(\bar{x}, \bar{y}) > F(x^k, y^k), x^k \in X, u^k \in \Lambda(x^k, y^k) \text{ for all } k \in \mathbb{N}.$$

Since CQ (1.4) holds at \bar{x} and is persistent around this point, there exists a number K such that this CQ is satisfied at x^k for all $k > K$. Thus in addition to the convexity of the lower-level problem, we have $y^k \in S(x^k)$ for all $k > K$. Similarly, note that with $\tilde{u} \in \Lambda(\bar{x}, \bar{y})$ and the fulfilment of the Slater CQ at \bar{x} , it holds that $\bar{y} \in S(\bar{x})$, while taking into account the convexity of the lower-level problem. In conclusion,

$$F(\bar{x}, \bar{y}) > F(x^k, y^k), x^k \in X, y^k \in S(x^k) \text{ for all } k > K \quad (4.2)$$

with $x^k \rightarrow \bar{x}$ and $y^k \rightarrow \bar{y}$. Thus, (\bar{x}, \bar{y}) is not a local optimal solution for (P).

For the reverse implication (\Leftarrow), suppose that (\bar{x}, \bar{y}) is not a local optimal solution of (P). Then there exists (x^k, y^k) with $x^k \rightarrow \bar{x}$ and $y^k \rightarrow \bar{y}$ such that (4.2) holds for all $k \in \mathbb{N}$. Since the CQ (1.4) holds at \bar{x} , it follows from the proof of the previous implication that there exists a number K such this CQ is satisfied at x^k for all $k > K$. Thus in addition to the convexity of the lower-level problem, we have $\Lambda(x^k, y^k) \neq \emptyset$ for all $k > K$. Further note that from Theorem 3.1, the set-valued map Λ (3.3) is inner semicompact at (\bar{x}, \bar{y}) under the assumptions made. Hence, there exists a sequence $u^k \in \Lambda(x^k, y^k)$ admitting a subsequence converging to some \tilde{u} . The mapping Λ being also closed according to the same theorem, we have $\tilde{u} \in \Lambda(\bar{x}, \bar{y})$. Now observe that the upper-level objective function F is independent of u . Hence, the inequality $F(\bar{x}, \bar{y}) > F(x^k, y^k)$ raises a contradiction to $(\bar{x}, \bar{y}, \tilde{u})$ being a local optimal solution of problem (4.1). This concludes the proof. \square

REMARK 4.2 (on the reverse implication (\Leftarrow) of Theorem 4.1). We have consciously proceeded by using the weaker inner semicompactness argument rather the upper semicontinuity which is also ensured by the assumptions made, cf. Theorem 3.1. This highlights the fact that the latter property, also used in [4] is more than what is needed to conclude the proof. Of course, the closedness of Λ is absolutely required in this case. What is interesting in this implication is that when there is more than one Lagrange multiplier, one must check local optimality of (4.1) for all of them, in order to generate a local optimal point for (P). If this property holds for all

but one element of $\Lambda(\bar{x}, \bar{y})$, the result may not hold [4]. Obviously, in the nonsmooth case, we need even more assumptions as shown in the above Theorem 4.1. In the smooth case, it was further observed in [4] that under the constant rank CQ, one can reduce the points to be checked only to the vertices of $\Lambda(\bar{x}, \bar{y})$. Such a property still has to be investigated in the nonsmooth framework.

It is also important to recall that the uniform local Lipschitz continuity of the functions involved in the lower-level problem (3.1) can be replaced by the uniform boundedness of the corresponding partial subdifferential set-valued mappings. The next result is the global counterpart of the previous one. We do not include the proof here, since it follows easily on the lines of its smooth counterpart from [4].

THEOREM 4.3 (global relation of (P) to its KKT reformulation in the nonsmooth case). *Let (\bar{x}, \bar{y}) be a global optimal solution of (P) and assume that CQ (1.4) holds at \bar{x} . Then, for each $\bar{u} \in \Lambda(\bar{x}, \bar{y})$, the point $(\bar{x}, \bar{y}, \bar{u})$ is also a global optimal solution of (4.1). Conversely, assume that $(\bar{x}, \bar{y}, \bar{u})$ is a global optimal solution of problem (4.1) and CQ (1.4) holds at all $x \in X$. Then, (\bar{x}, \bar{y}) is a global optimal solution of (P) .*

5. M-stationarity in the nonsmooth case. In this section, we introduce and justify a nonsmooth extension of the M-stationarity concept from Definition 1.2 (ii), in terms of subdifferentials and coderivatives.

DEFINITION 5.1 (M-stationarity concepts for nonsmooth bilevel programs). *A point (\bar{x}, \bar{y}) is said to be SP-M-STATIONARY (resp. P-M-STATIONARY) if for every $\bar{u} \in \Lambda(\bar{x}, \bar{y})$ (resp. for some $\bar{u} \in \Lambda(\bar{x}, \bar{y})$) we can find $t := (t^0, t^1, \dots, t^p)$ and $(\alpha, \beta, \gamma, \lambda)$ with $\lambda \in \mathbb{R}_+$ such that (1.8) holds together with the following conditions:*

$$0 \in \partial F(\bar{x}, \bar{y}) + \sum_{j=1}^k \alpha_j (\partial G_j(\bar{x}), 0_m) + \partial \langle \beta, g \rangle(\bar{x}, \bar{y}) \\ + D^*(\partial_y f)((\bar{x}, \bar{y})|t^0)(\gamma) + \sum_{i=1}^p D^*(\partial_y g_i)((\bar{x}, \bar{y})|t^i)(\bar{u}_i \gamma), \quad (5.1)$$

$$t^0 \in \partial_y f(\bar{x}, \bar{y}), \quad t^i \in \partial_y g_i(\bar{x}, \bar{y}), \quad i = 1, \dots, p, \quad t^0 + \sum_{i=1}^p \bar{u}_i t^i = 0, \quad (5.2)$$

$$\forall i \in \nu : \sum_{l=1}^m t_l^i \gamma_l = 0, \quad \beta_\eta = 0, \quad (5.3)$$

$$\forall i \in \theta : (\beta_i > 0 \wedge \sum_{l=1}^m t_l^i \gamma_l > 0) \vee \beta_i (\sum_{l=1}^m t_l^i \gamma_l) = 0. \quad (5.4)$$

The relationships (1.8) and (5.1)–(5.4) will be called M-STATIONARITY CONDITIONS.

Note that if the functions involved in (P) become \mathcal{C}^1 and \mathcal{C}^2 (for those involved in the lower-level problem), these conditions coincide with their smooth counterpart in Definition 1.2 (ii). In this situation, equation $t^0 + \sum_{i=1}^p \bar{u}_i t^i = 0$, from (5.2) reduces to $\mathcal{L}(x, y, u) := \nabla_y f(x, y) + \sum_{i=1}^p u_i \nabla_y g_i(x, y) = 0$.

To simplify the justification of the concepts in Definition 5.1, we first derive the M-stationarity conditions for the problem in (4.1), i.e. the KKT reformulation itself.

THEOREM 5.2 (M-stationarity conditions for the KKT reformulation (4.1)). *Let $(\bar{x}, \bar{y}, \bar{u})$ be a local optimal solution of (4.1) and assume that $\partial_y f$ and $\partial_y g_i$ for $i = 1, \dots, p$, are closed and uniformly bounded around (\bar{x}, \bar{y}) . Furthermore, suppose that*

for all $t := (t^0, t^1, \dots, t^p)$ satisfying (5.2), the QC (3.16) holds together with CQ:

$$\left. \begin{array}{l} 0 \in \sum_{j=1}^k \alpha_j (\partial G_j(\bar{x}), 0_m) + \partial \langle \beta, g \rangle(\bar{x}, \bar{y}) \\ + D^*(\partial_y f)((\bar{x}, \bar{y})|t^0)(\gamma) + \sum_{i=1}^p D^*(\partial_y g_i)((\bar{x}, \bar{y})|t^i)(\bar{u}_i \gamma) \\ \text{with (1.8), (5.2)–(5.4) and } \|\gamma\| \leq \lambda, \lambda \geq 0 \end{array} \right\} \implies \begin{cases} \alpha = 0, \\ \beta = 0, \\ \lambda = 0. \end{cases} \quad (5.5)$$

Then, there exist $t := (t^0, t^1, \dots, t^p)$ and $(\alpha, \beta, \gamma, \lambda, r)$ with $\lambda \in \mathbb{R}_+$, $r \in \mathbb{R}_+ \setminus \{0\}$, $\|\gamma\| \leq \lambda$ and $\|(\alpha, \beta, \lambda)\| \leq r$ such that (1.8) and (5.1)–(5.4) are satisfied.

Proof. Recall that since the set-valued mappings $\partial_y f$ and $\partial_y g_i$ for $i = 1, \dots, p$, are closed and uniformly bounded around (\bar{x}, \bar{y}) , the multifunction \mathcal{L} is closed around this point. Thus the constraint $0 \in \mathcal{L}(x, y, u)$ can be reformulated as $d_{\text{gph } \mathcal{L}}(x, y, u, 0) \leq 0$ with $d_{\text{gph } \mathcal{L}}$ denoting the distance function on $\mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p$. Hence, problem (4.1) can take the following operator constraint form:

$$\begin{array}{l} \min_{x, y, u} \{F(x, y) \mid \psi(x, y, u) \in \Xi\} \\ \text{with } \begin{cases} \psi(x, y, u) := [G(x), h(x, y, u), (u, -g(x, y))], \\ h(x, y, u) := d_{\text{gph } \mathcal{L}}(x, y, u, 0), \\ \Xi := \mathbb{R}^k \times \mathbb{R}_- \times \Pi, \\ \Pi := \{(a, b) \in \mathbb{R}^{2p} \mid a \geq 0, b \geq 0, a^\top b = 0\}. \end{cases} \end{array} \quad (5.6)$$

Note that $h[\mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p \rightarrow \mathbb{R}]$ and $\psi[\mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p \rightarrow \mathbb{R}^k \times \mathbb{R} \times \mathbb{R}^p \times \mathbb{R}^p]$. Applying [9, Proposition 3.1] to (5.6), it follows that there exists v with $\|v\| \leq r$ (for some $r > 0$) and $v \in N_\Lambda(\psi(\bar{x}, \bar{y}, \bar{u}))$ such that we have the optimality condition

$$0 \in \partial_{x, y, u} F(\bar{x}, \bar{y}) + \partial \langle v, \psi \rangle(\bar{x}, \bar{y}, \bar{u}) \quad (5.7)$$

provided the following qualification condition holds at $(\bar{x}, \bar{y}, \bar{u})$:

$$[0 \in \partial \langle v, \psi \rangle(\bar{x}, \bar{y}, \bar{u}), v \in N_\Xi(\psi(\bar{x}, \bar{y}, \bar{u}))] \implies v = 0. \quad (5.8)$$

In what follows, we provide detailed forms for conditions (5.7) and (5.8) in terms of the problem data in (4.1). By the product rule of normal cones, we have

$$\begin{aligned} N_\Xi(\psi(\bar{x}, \bar{y}, \bar{u})) &= N_{\mathbb{R}^k}(G(\bar{x})) \times N_{\mathbb{R}_-}(h(\bar{x}, \bar{y}, \bar{u})) \times N_\Pi(\bar{u}, -g(\bar{x}, \bar{y})) \\ &= \{(\alpha, \lambda, \zeta, \beta) \in \mathbb{R}^{k+1+2p} \mid \alpha \geq 0, \alpha^\top G(\bar{x}) = 0, \lambda \geq 0, \\ &\quad \zeta_\nu = 0, \beta_\eta = 0, \forall i \in \theta : (\zeta_i < 0 \wedge \beta_i > 0) \vee \zeta_i \beta_i = 0\}, \end{aligned} \quad (5.9)$$

where the second equality is due to the expression of the normal cone to Π given, for example in [34]. Now let $v := (\alpha, \lambda, \zeta, \beta) \in N_\Xi(\psi(x, y, u))$, then we have

$$\langle v, \psi \rangle(x, y, u) = \sum_{j=1}^k \alpha_j G_j(x) + \lambda h(x, y, u) + \sum_{i=1}^p \zeta_i u_i - \sum_{i=1}^p \beta_i g_i(x, y).$$

Applying the basic subdifferential sum rule on this equality we arrive at the inclusion

$$\begin{aligned} \partial \langle v, \psi \rangle(\bar{x}, \bar{y}, \bar{u}) &\subseteq \sum_{j=1}^k \alpha_j \partial G_j(\bar{x}) \times \{(0_m, 0_p)\} \\ &\quad + \partial \langle -\beta, g \rangle(\bar{x}, \bar{y}) \times \{\zeta\} + \lambda \partial h(\bar{x}, \bar{y}, \bar{u}), \end{aligned} \quad (5.10)$$

since all the functions involved are locally Lipschitz continuous and the multipliers λ and α_j for $j = 1, \dots, k$, are nonnegative. Moreover, the Lipschitz continuity of the distance function $d_{\text{gph } \mathcal{L}}$ implies that

$$\begin{aligned} \partial h(\bar{x}, \bar{y}, \bar{u}) &\subseteq \{(a, b, c) \in \mathbb{R}^{n+m+p} \mid (a, b, c, d) \in \partial d_{\text{gph } \mathcal{L}}(\bar{x}, \bar{y}, \bar{u}, 0)\} \\ &= \{(a, b, c) \in \mathbb{R}^{n+m+p} \mid (a, b, c, d) \in N_{\text{gph } \mathcal{L}}(\bar{x}, \bar{y}, \bar{u}, 0) \cap \mathbb{B}_{q+m}\} \\ &\subseteq \bigcup_{\gamma \in \mathbb{B}_m} D^* \mathcal{L}((\bar{x}, \bar{y}, \bar{u})|0)(\gamma) \end{aligned} \quad (5.11)$$

with $q := n + m + p$. Here, the second line is obtained from (2.4), while taking in account that the graph of \mathcal{L} is closed. The last inclusion in (5.11) follows from the definition in (2.5). Inserting the coderivative estimate of Theorem 3.3 in (5.11) and substituting the outcome in (5.10), we arrive at the following upper bound

$$\begin{aligned} & \bigcup_{\gamma \in \mathbb{B}_m(0, \lambda)} \bigcup_{\substack{t: t^0 + \sum_{i=1}^p u_i t^i = 0 \\ t^0 \in \partial_y f(\bar{x}, \bar{y}), t^i \in \partial_y g_i(\bar{x}, \bar{y})}} \left\{ \left[\sum_{j=1}^k \alpha_j (\partial G_j(\bar{x}), 0_m) + \partial \langle -\beta, g \rangle(\bar{x}, \bar{y}) \right. \right. \\ & \quad \left. \left. + D^*(\partial_y f)((\bar{x}, \bar{y})|t^0)(\gamma) + \sum_{i=1}^p D^*(\partial_y g_i)((\bar{x}, \bar{y})|t^i)(\bar{u}_i \gamma) \right] \right. \\ & \quad \left. \times \left\{ \zeta + \left(\sum_{l=1}^m t_l^1 \gamma_l, \dots, \sum_{l=1}^m t_l^p \gamma_l \right)^\top \right\} \right\} \end{aligned} \quad (5.12)$$

for $\partial \langle v, \psi \rangle(\bar{x}, \bar{y}, \bar{u})$. Inserting (5.9) and the above estimate of $\partial \langle v, \psi \rangle(\bar{x}, \bar{y}, \bar{u})$ in (5.8), we get CQ (5.5) which implies the fulfilment of (5.8). Proceeding similarly on the optimality condition (5.7), we have (5.1)–(5.4). \square

It is worth mentioning that the bound r on the multipliers α , β and λ can exactly be chosen as $r := \ell_W \ell_F + 1$ (see [9]), where ℓ_F denotes the Lipschitz modulus of the upper-level objective function F , whereas ℓ_W stands for the Lipschitz modulus of $\Psi_W(v) := \{(x, y, u) \in W \mid \psi(x, y, u) + v \in \Xi\}$ with ψ and Ξ given in (5.6). Note that W denotes the neighborhood of $(\bar{x}, \bar{y}, \bar{u})$ where this point is locally optimal for (4.1). It is also important to recall that under the CQ (5.8), the above multifunction is Lipschitz-like [21]. In the same vein, note that the result in Theorem 5.2 remains valid if the weaker calmness assumption is imposed on the mapping Ψ_W . Finally, observe that to verify CQ (5.5), it might be useful, in some cases, to replace γ by $\lambda \gamma$.

CQ (5.5) is closely related to the no nonzero abnormal multiplier constraint qualification (NNAMCQ) employed in [33] for the smooth MPCC. In the literature on MPCCs/MPECs the so-called MPEC-MFCQ also plays an important role in the development of a variety of optimality conditions. We are now interested in the extension of this CQ to the KKT reformulation (4.1) of the nonsmooth version of the bilevel program. Considering the structure of the inclusion constraint $0 \in \mathcal{L}(x, y, u)$, it appears to be difficult to efficiently do this, as one may need to introduce a notion of *membership* for the coderivative in the sense that $t \in D^*M(\bar{x}, \bar{y})$ if and only if $\langle y^*, t \rangle \in D^*M(\bar{x}, \bar{y})(y^*)$ for all y^* . Consequently, one may have to prove or assume that for $x^* \in D^*M(\bar{x}, \bar{y})(y^*)$, there exists some $t \in D^*M(\bar{x}, \bar{y})$ such that $x^* = \langle y^*, t \rangle$. To avoid this difficulty, we introduce the following *second order lower-level constraint qualification* (SOLLCQ)

$$0 \in D^*(\partial_y f)((\bar{x}, \bar{y})|t^0)(\gamma) + \sum_{i=1}^p D^*(\partial_y g_i)((\bar{x}, \bar{y})|t^i)(\bar{u}_i \gamma) \Big\} \implies \gamma = 0 \quad (5.13)$$

$$\forall i = 1, \dots, p: \sum_{l=1}^m t_l^i \gamma_l = 0 \text{ and (5.2) satisfied}$$

in order to reasonably move $0 \in \mathcal{L}(x, y, u)$ to the upper-level objective function by a partial exact penalization via the distance function. This then paves the way for a M-type MPEC-MFCQ tailored to (4.1), which emerges from a combination of the nonsmooth MFCQ [16] and the smooth MPEC-MFCQ employed in [30]. To proceed, we denote by $J := J(\bar{x}) := \{j \in \{1, \dots, k\} \mid G_j(\bar{x}) = 0\}$.

DEFINITION 5.3 (M-type MPEC-Mangasarian-Fromowitz CQ). *The M-MPEC-MFCQ holds at $(\bar{x}, \bar{y}, \bar{u})$ if for all $a_j^G \in \partial G_j(\bar{x})$ with $j \in J$ and for all $a_i^g \in \partial g_i(\bar{x}, \bar{y})$ with $i \in \nu \cup \theta$, the family $\{a_i^g \mid i \in \nu \cup \theta\}$ is linearly independent and there exists a vector $d := (d^x, d^y, d^u)$ with $d^{xy} := (d^x, d^y)$, such that*

$$\begin{aligned} & d_i^u = 0 \text{ for all } i \in \eta \cup \theta, \\ & \langle d^{xy}, a_i^g \rangle = 0 \text{ for all } i \in \nu \cup \theta, \\ & \langle d^x, a_j^G \rangle < 0 \text{ for all } j \in J. \end{aligned}$$

Note that the prefix ‘‘M’’ is used here to label this CQ, in order to differentiate it with a similar one for the C-type approach to be introduced in the next section.

THEOREM 5.4 (M-stationarity under the M-MPEC-MFCQ). *Let $(\bar{x}, \bar{y}, \bar{u})$ be a local optimal solution of problem (4.1) and assume that the set-valued mappings $\partial_y f$ and $\partial_y g_i$ for $i = 1, \dots, p$, are closed and uniformly bounded around (\bar{x}, \bar{y}) . Furthermore, suppose that the SOLLCQ and the M-MPEC-MFCQ are satisfied and for all $t := (t^0, t^1, \dots, t^p)$ satisfying (5.2), the QC (3.16) holds. Then, there exist $t := (t^0, t^1, \dots, t^p)$ and $(\alpha, \beta, \gamma, \lambda)$ with $\lambda \in \mathbb{R}_+$ such that conditions (1.8) and (5.2)–(5.4) are satisfied together with the following one:*

$$\begin{aligned} 0 \in \partial F(\bar{x}, \bar{y}) + \sum_{j=1}^k \alpha_j (\partial G_j(\bar{x}), 0_m) + \sum_{i=1}^p \beta_i \bar{\partial} g_i(\bar{x}, \bar{y}) \\ + D^*(\partial_y f)((\bar{x}, \bar{y})|t^0)(\gamma) + \sum_{i=1}^p D^*(\partial_y g_i)((\bar{x}, \bar{y})|t^i)(\bar{u}_i \gamma). \end{aligned} \quad (5.14)$$

Proof. Let $(\bar{x}, \bar{y}, \bar{u})$ be a local optimal solution of problem (4.1), then applying [3, Proposition 2.4.3], it follows that there exists a number $\lambda > 0$ such that $(\bar{x}, \bar{y}, \bar{u})$ is also a local optimal solution of the problem

$$\begin{aligned} \min_{x, y, u} \{F(x, y) + \lambda d_{\psi^{-1}(\text{gph } \mathcal{L})}(x, y, u) \mid G_j(x) \leq 0, j = 1, \dots, k, \\ u_i \geq 0, g_i(x, y) \leq 0, u_i g_i(x, y) = 0, i = 1, \dots, p\}, \end{aligned} \quad (5.15)$$

where $\psi(x, y, u) := (x, y, u, 0)$. Now set $\mathcal{F}_\lambda(x, y, u) := F(x, y) + \lambda d_{\psi^{-1}(\text{gph } \mathcal{L})}(x, y, u)$, and observe that we can find a vector \bar{v} such that $(\bar{x}, \bar{y}, \bar{u}, \bar{v})$ locally solves the following problem

$$\begin{aligned} \min_{x, y, u, v} \{\mathcal{F}_\lambda(x, y, u) \mid G_j(x) \leq 0, j = 1, \dots, k, \\ g_i(x, y) + v_i = 0, i = 1, \dots, p, (u, v) \in \Pi\}, \end{aligned}$$

where the set Π is defined as in (5.6). Noting that this problem is Lipschitz continuous and applying the Fritz-John-type Lagrange multipliers rule of Mordukhovich [22] (also see [32, Corollary 4.2]), it holds that

$$\begin{aligned} 0 \in \kappa (\partial \mathcal{F}_\lambda(\bar{x}, \bar{y}, \bar{u}), 0_p) + \sum_{j=1}^k \alpha_j (\partial G_j(\bar{x}), 0_{m+2p}) + (0_{n+m+p}, \beta) \\ + \sum_{i=1}^p \beta_i (\bar{\partial} g_i(\bar{x}, \bar{y}), 0_{2p}) + (0_{n+m}) \times N_\Pi(\bar{u}, -g(\bar{x}, \bar{y})), \end{aligned} \quad (5.16)$$

for some vector $(\kappa, \alpha, \beta) \neq 0$ such that $\kappa \in \mathbb{R}_+$ and (1.8) are satisfied, while taking into account that $\partial g_i(\bar{x}, \bar{y}) \subseteq \bar{\partial} g_i(\bar{x}, \bar{y})$ for $i = 1, \dots, p$. If we assume that $\kappa = 0$ in (5.16), then there exist some $a_j^G \in \partial G_j(\bar{x})$ for $j = 1, \dots, k$, $a_i^g \in \bar{\partial} g_i(\bar{x}, \bar{y})$ for $i = 1, \dots, p$, and $\zeta \in \mathbb{R}^p$ such that

$$\sum_{j \in J} \alpha_j \langle a_j^G, 0_{m+p} \rangle + \sum_{i \in \nu \cup \theta} \beta_i \langle a_i^g, 0_p \rangle + \sum_{i \in \eta \cup \theta} \zeta_i \langle 0_{n+m}, e^i \rangle = 0, \quad (5.17)$$

$$\zeta_\nu = 0, \beta_\eta = 0, \forall i \in \theta : (\zeta_i < 0 \wedge \beta_i > 0) \vee \zeta_i \beta_i = 0, \quad (5.18)$$

where the second line is due to the expression of $N_\Pi(\bar{u}, -g(\bar{x}, \bar{y}))$ extracted from (5.9). Observe that the summations in (5.17) are restricted to J , $\nu \cup \theta$ and $\eta \cup \theta$ following (5.18) and the definition of J . Multiplying (5.17) with a vector d defined as in M-MPEC-MFCQ, we obtain

$$\sum_{j \in J} \alpha_j \langle a_j^G, d^x \rangle + \sum_{i \in \nu \cup \theta} \beta_i \langle a_i^g, d^{xy} \rangle + \sum_{i \in \eta \cup \theta} \zeta_i d_i^u = 0.$$

By further considering the definition of M-MPEC-MFCQ, $\sum_{j \in J} \alpha_j \langle a_j^G, d^x \rangle = 0$. Since $\langle a_j^G, d^x \rangle < 0$ and $\alpha_j \geq 0$ for $j \in J$, it holds that $\alpha = 0$. Inserting this value in (5.17), it follows that

$$\sum_{i \in \nu \cup \theta} \beta_i(a_i^g, 0_p) + \sum_{i \in \eta \cup \theta} \zeta_i(0_{n+m}, e^i) = 0. \quad (5.19)$$

Now, observe that we have $\sum_{i \in \eta \cup \theta} \zeta_i(0_{n+m}, e^i) = \sum_{i=1}^p \zeta_i(0_{n+m}, e^i) = (0_{n+m}, \zeta)$ given that $\zeta_i = 0$ for $i \in \nu$. Thus, we have from (5.19) that $\zeta = 0$. Moreover, taking into account that the family $\{a_i^g \mid i \in \nu \cup \theta\}$ is linearly independent (cf. M-MPEC-MFCQ), we also get from (5.19) that $\beta = 0$. We have now shown that if $\kappa = 0$, then all the other components of the vector (κ, α, β) also vanish. This contradicts the Fritz-John-type Lagrange multiplier rule. Thus $\kappa \neq 0$. Hence, by scaling, it follows from (5.16) that we can find $\zeta \in \mathbb{R}^p$ such that (5.18) holds together with

$$\begin{aligned} 0 \in \partial \mathcal{F}_\lambda(\bar{x}, \bar{y}, \bar{u}) + \sum_{j=1}^k \alpha_j (\partial G_j(\bar{x}), 0_{m+p}) \\ + \sum_{i=1}^p \beta_i (\bar{\partial} g_i(\bar{x}, \bar{y}), 0_p) + (0_{n+m}, \zeta). \end{aligned} \quad (5.20)$$

By the sum rule, an upper bound for the subdifferential of \mathcal{F}_λ can be obtained as

$$\partial \mathcal{F}_\lambda(\bar{x}, \bar{y}, \bar{u}) \subseteq (\partial F(\bar{x}, \bar{y}), 0_p) + \partial d_{\psi^{-1}(\text{gph } \mathcal{L})}(\bar{x}, \bar{y}, \bar{u}). \quad (5.21)$$

Furthermore, the following calculations give an upper estimate for the subdifferential of the involved distance function:

$$\begin{aligned} \partial d_{\psi^{-1}(\text{gph } \mathcal{L})}(\bar{x}, \bar{y}, \bar{u}) &= N_{\psi^{-1}(\text{gph } \mathcal{L})}(\bar{x}, \bar{y}, \bar{u}) \cap \mathbb{B}_{n+m+p} \\ &\subseteq \{ \nabla \psi(\bar{x}, \bar{y}, \bar{u})^\top v \mid v \in N_{\text{gph } \mathcal{L}}(\psi(\bar{x}, \bar{y}, \bar{u})) \} \cap \mathbb{B}_{n+m+p} \\ &= \bigcup_{\gamma \in \mathbb{R}^m} D^* \mathcal{L}((\bar{x}, \bar{y}, \bar{u})|0)(\gamma) \cap \mathbb{B}_{n+m+p} \end{aligned}$$

where the first line is due to (2.4), the third one to definition (2.5) and the inclusion in the second line is derived from (2.2), under the QC (2.3) with $\Xi := \text{gph } \mathcal{L}$ and $\bar{x} := (\bar{x}, \bar{y}, \bar{u})$. One can then easily check that this implication is equivalent to

$$0_{n+m+p} \in D^* \mathcal{L}((\bar{x}, \bar{y}, \bar{u})|0)(v) \implies v = 0.$$

Considering the estimate of $D^* \mathcal{L}$ from Theorem 3.3, it follows that the SOLLCQ (5.13) is a sufficient condition for the latter implication to hold. Furthermore, combining the estimate of $\partial d_{\psi^{-1}(\text{gph } \mathcal{L})}$ above with (5.18), (5.20), (5.21) and Theorem 3.3, we arrive at the desired result. \square

Observe that we have $\sum_{i=1}^p \beta_i \bar{\partial} g_i(\bar{x}, \bar{y})$ in the M-type optimality conditions of Theorem 5.4, instead of the term $\partial \langle \beta, g \rangle(\bar{x}, \bar{y})$ in Theorem 5.2. This is a purely technical consideration, which simplifies the implementation of the M-MPEC-MFCQ above. Further note that in the smooth case, the SOLLCQ reduces to the nonsingularity of the matrix $\nabla \mathcal{L}(\bar{x}, \bar{y}, \bar{u})^\top$. This is automatically the case if there exists some $i \in \{1, \dots, p\}$ such that the family of gradients $\{\nabla_{y_l} g_i(\bar{x}, \bar{y}) \mid l = 1, \dots, m\}$ is linearly independent.

COROLLARY 5.5 (justification of SP-M- and P-M-stationarity for nonsmooth bilevel programs). *Let (\bar{x}, \bar{y}) be a local optimal solution of (P). Suppose that the assumptions of Theorem 5.2 are satisfied for all $\bar{u} \in \Lambda(\bar{x}, \bar{y})$, then (\bar{x}, \bar{y}) is SP-M-stationary. If they hold just for at least one $\bar{u} \in \Lambda(\bar{x}, \bar{y})$, then (\bar{x}, \bar{y}) is P-M-stationary.*

Proof. Observe from Theorem 4.1 that if (\bar{x}, \bar{y}) is a local optimal solution of (P) , then, for all $\bar{u} \in \Lambda(\bar{x}, \bar{y})$, the point $(\bar{x}, \bar{y}, \bar{u})$ is a local optimal solution of problem (4.1). Combining this fact with Theorem 5.2, we get the first implication. The second one follows similarly, while noting that it is enough that the assumptions of Theorem 5.2 hold at just one lower-level multiplier \bar{u} . \square

An analogous result can be stated for the M-stationarity conditions derived in Theorem 5.4. Also observe that using inclusion (3.24), the above optimality conditions and the subsequent ones can be formulated in terms of the second subdifferentials of the functions involved in the lower-level problem.

6. C-stationarity in the nonsmooth case. Following the pattern of the stationarity concepts in Definition 1.2 valid for the smooth case, we are tempted to consider SP-C- and P-C-stationarity conditions for the nonsmooth framework, in a way similar to Definition 5.1 while replacing condition (5.4) by

$$\forall i \in \theta : \beta_i \sum_{l=1}^m t_l^i \gamma_l \geq 0. \quad (6.1)$$

This extension is rather *artificial* as it will be shown below. Nevertheless, we replace the ‘‘C’’ above by a ‘‘C^o’’ to designate the underlined stationarity concepts. These conditions can be deduced from Theorem 5.2 as follows.

COROLLARY 6.1 (artificial extension of C-stationarity in nonsmooth bilevel programming). *Let (\bar{x}, \bar{y}) be a local optimal solution of (P) . Suppose that the assumptions of Theorem 5.2 are satisfied for all $\bar{u} \in \Lambda(\bar{x}, \bar{y})$, then (\bar{x}, \bar{y}) is SP-C^o-stationary. If they hold just for at least one $\bar{u} \in \Lambda(\bar{x}, \bar{y})$, then (\bar{x}, \bar{y}) is P-C^o-stationary.*

Proof. Simply observe that if condition (5.4) holds, then (6.1) also holds. \square

In a general framework of a smooth MPCC, with the complementarity constraint $H_i(x) \geq 0$, $G_i(x) \geq 0$, $H_i(x)G_i(x) = 0$, $i = 1, \dots, d$, the C-type stationarity conditions are obtained while considering $\text{co}\{\nabla H_i(\bar{x}), \nabla G_i(\bar{x})\}$ for all i such that $H_i(\bar{x}) = G_i(\bar{x}) = 0$, where ‘‘co’’ stands for the convex hull. Based on this original idea, we now provide a *natural* extension of the C-stationarity conditions to the nonsmooth case, and that we label as such. To proceed, we set $q := n + m + p$ and consider the following sequence of equations in order to simplify the presentation:

$$\forall i \in \theta : r_i \in \{0, 1\}, \forall i \in \nu : a^i \in \bar{\partial}g_i(\bar{x}, \bar{y}), \quad (6.2)$$

$$\forall i \in \theta, s \in \{2, \dots, q+1\}, s' \in \{1, \dots, q+1\} : b^{is}, c^{is'} \in \bar{\partial}g_i(\bar{x}, \bar{y}), \quad (6.3)$$

$$\forall i \in \theta, s, s' \in \{1, \dots, q+1\} : v_{is}, w_{is'} \in \mathbb{R}_+, \sum_{s=1}^{q+1} v_{is} = \sum_{s'=1}^{q+1} w_{is'} = 1, \quad (6.4)$$

$$\forall i \in \nu : \sum_{l=1}^m t_l^i \gamma_l = 0, \forall i \in \eta : \mu_i - \sum_{l=1}^m t_l^i \gamma_l = 0, \quad (6.5)$$

$$\forall i \in \theta : r_i \mu_i v_{i1} - \sum_{l=1}^m t_l^i \gamma_l = 0. \quad (6.6)$$

Note the presence of the discrete variable $r_i \in \{0, 1\}$ for $i \in \theta$ that we introduce in order to be able to provide a detailed form of the stationarity conditions in the following theorem, which is the counterpart of Theorem 5.2. For the convenience of the reader, we recall that $\bar{\partial}g_i$ denotes the convexified/Clarke subdifferential of g_i .

THEOREM 6.2 (natural extension of C-stationarity conditions). *Let $(\bar{x}, \bar{y}, \bar{u})$ be a local optimal solution of (4.1) and assume that the set-valued mappings $\partial_y f$ and $\partial_y g_i$ for $i = 1, \dots, p$, are closed and uniformly bounded around (\bar{x}, \bar{y}) . Suppose that for all $t := (t^0, t^1, \dots, t^p)$ satisfying (5.2), the QC (3.16) holds. Furthermore, let the CQ*

$$0 \in \left. \begin{aligned} & \sum_{j=1}^k \alpha_j (\partial G_j(\bar{x}), 0_m) + D^*(\partial_y f)((\bar{x}, \bar{y})|t^0)(\gamma) \\ & + \sum_{i=1}^p D^*(\partial_y g_i)((\bar{x}, \bar{y})|t^i)(\bar{u}_i \gamma) + \sum_{i \in \nu} \mu_i a^i \\ & + \sum_{i \in \theta} \sum_{s=2}^{q+1} r_i \mu_i v_{is} b^{is} + \sum_{i \in \theta} \sum_{s'=1}^{q+1} \mu_i (1 - r_i) w_{is'} c^{is'} \\ & \text{with (1.8), (5.2), (6.2)–(6.4) and } \|\gamma\| \leq \lambda, \lambda \geq 0 \end{aligned} \right\} \implies \begin{cases} \alpha = 0, \\ \mu = 0, \\ \lambda = 0, \end{cases} \quad (6.7)$$

be satisfied. Then, there exist $(\alpha, \mu, \gamma, \lambda)$ with $\lambda \in \mathbb{R}_+$, $\|\gamma\| \leq \lambda$ and $\|(\alpha, \mu, \lambda)\| \leq r$ (for some $r > 0$), $t := (t^0, t^1, \dots, t^p)$, a^i with $i \in \nu$, $r_i \in \{0, 1\}$ with $i \in \theta$, v_{is} with $i \in \theta$ and $s = 1, \dots, q+1$, b^{is} with $i \in \theta$ and $s = 2, \dots, q+1$, $w_{is'}$ and $c^{is'}$ with $i \in \theta$ and $s' = 1, \dots, q+1$, such that (1.8), (5.2) and (6.2)–(6.6) hold together with

$$\begin{aligned} 0 \in & \partial F(\bar{x}, \bar{y}) + \sum_{j=1}^k \alpha_j (\partial G_j(\bar{x}), 0_m) \\ & + D^*(\partial_y f)((\bar{x}, \bar{y})|t^0)(\gamma) + \sum_{i=1}^p D^*(\partial_y g_i)((\bar{x}, \bar{y})|t^i)(\bar{u}_i \gamma) \\ & + \sum_{i \in \nu} \mu_i a^i + \sum_{i \in \theta} \sum_{s=2}^{q+1} r_i \mu_i v_{is} b^{is} + \sum_{i \in \theta} \sum_{s'=1}^{q+1} \mu_i (1 - r_i) w_{is'} c^{is'}. \end{aligned} \quad (6.8)$$

Proof. The proof technique here is similar to the one of Theorem 5.2 as we also start by considering the operator constraint reformulation of problem (4.1), but

$$\text{with } \begin{cases} \psi(x, y, u) := [G(x), h(x, y, u), V(x, y, u)], \\ \Xi := \mathbb{R}^k \times \mathbb{R}_- \times \{0_p\}, \\ V_i(x, y, u) := \min\{u_i, -g_i(x, y)\}, i = 1, \dots, p. \end{cases} \quad (6.9)$$

In this case it is elementary that we have $(\alpha, \lambda, \mu) \in N_{\Xi}(\psi(\bar{x}, \bar{y}, \bar{u}))$ if and only if $\lambda \geq 0$, $\alpha \geq 0$, $\alpha^\top G(\bar{x}) = 0$, while for the scalarization of ψ we get

$$\partial \langle v, \psi \rangle(\bar{x}, \bar{y}, \bar{u}) \subseteq \sum_{j=1}^k \alpha_j \partial G_j(\bar{x}) \times \{(0_m, 0_p)\} + \lambda \partial h(\bar{x}, \bar{y}, \bar{u}) + \sum_{i=1}^p \mu_i \bar{\partial} V_i(\bar{x}, \bar{y}, \bar{u}), \quad (6.10)$$

where $\bar{\partial} V_i$ denotes the convexified/Clarke subdifferential of V_i . Applying [3, Proposition 2.3.12] to $\bar{\partial} V_i$ we have

$$\bar{\partial} V_i(\bar{x}, \bar{y}, \bar{u}) \subseteq \begin{cases} \{(0_{n+m}, e^i)\} & \text{if } i \in \eta, \\ -\bar{\partial} g_i(\bar{x}, \bar{y}) \times \{0_p\} & \text{if } i \in \nu, \\ \text{co} \{ \{(0_{n+m}, e^i)\} \cup [-\bar{\partial} g_i(\bar{x}, \bar{y}) \times \{0_p\}] \} & \text{if } i \in \theta, \end{cases} \quad (6.11)$$

where $e^i := (0, \dots, 0, 1, 0, \dots, 0)^\top$ is a p -dimensional vector with 1 at position i with $i \in \eta \subseteq \{1, \dots, p\}$. Taking into account that we have set $q := n + m + p$, let us now show the following inclusion for any $\mu \in \mathbb{R}^p$:

$$\begin{aligned} \sum_{i=1}^p \mu_i \bar{\partial} V_i(\bar{x}, \bar{y}, \bar{u}) & \subseteq \{(v, w) \mid w = \sum_{i \in \eta} \mu_i e^i + \sum_{i \in \theta} r_i \mu_i v_{i1} e^i, \\ v & = -\sum_{i \in \nu} \mu_i a^i - \sum_{i \in \theta} \sum_{s=2}^{q+1} r_i \mu_i v_{is} b^{is} - \sum_{i \in \theta} \sum_{s'=1}^{q+1} (1 - r_i) \mu_i w_{is'} c^{is'}, \end{aligned} \quad (6.12)$$

with the conditions (6.2)–(6.4) satisfied.

To proceed, first observe that we have

$$\sum_{i=1}^p \mu_i \bar{\partial} V_i(\bar{x}, \bar{y}, \bar{u}) = \sum_{i \in \eta} \mu_i \bar{\partial} V_i(\bar{x}, \bar{y}, \bar{u}) + \sum_{i \in \nu} \mu_i \bar{\partial} V_i(\bar{x}, \bar{y}, \bar{u}) + \sum_{i \in \theta} \mu_i \bar{\partial} V_i(\bar{x}, \bar{y}, \bar{u}). \quad (6.13)$$

Now pick any $u^\eta \in \sum_{i \in \eta} \mu_i \bar{\partial} V_i(\bar{x}, \bar{y}, \bar{u})$ and $u^\nu \in \sum_{i \in \nu} \mu_i \bar{\partial} V_i(\bar{x}, \bar{y}, \bar{u})$, then from (6.11),

$$u^\eta + u^\nu = \begin{bmatrix} -\sum_{i \in \nu} \mu_i a^i \\ \sum_{i \in \eta} \mu_i e^i \end{bmatrix} \text{ for some } a^i \in \bar{\partial} g_i(\bar{x}, \bar{y}), \quad i \in \nu. \quad (6.14)$$

Furthermore, let us consider a vector $u^i \in \bar{\partial} V_i(\bar{x}, \bar{y}, \bar{u})$ for $i \in \theta$, then we also have from (6.11) that $u^i \in \text{co} \left\{ \{(0_{n+m}, e^i)\} \cup -\bar{\partial} g_i(\bar{x}, \bar{y}) \times \{0_p\} \right\}$. Hence, by the well-known theorem of Carathéodory, it holds that

$$\text{either } \begin{cases} u^i = v_{i1}(0_{n+m}, e^i) - \sum_{s=2}^{q+1} v_{is}(b^{is}, 0_p), \\ \forall s = 1, \dots, q+1, v_{is} \geq 0, \sum_{s=1}^{q+1} v_{is} = 1, \\ \forall s = 2, \dots, q+1, b^{is} \in \bar{\partial} g_i(\bar{x}, \bar{y}), \end{cases} \quad (6.15)$$

$$\text{or } \begin{cases} u^i = -\sum_{s'=1}^{q+1} w_{is'}(c^{is'}, 0_p), \\ \forall s' = 1, \dots, q+1, w_{is'} \geq 0, \sum_{s'=1}^{q+1} w_{is'} = 1, \\ \forall s' = 1, \dots, q+1, c^{is'} \in \bar{\partial} g_i(\bar{x}, \bar{y}). \end{cases} \quad (6.16)$$

Thus we have the following representation of u^i whenever $i \in \theta$:

$$u^i = r_i \left[v_{i1}(0_{n+m}, e^i) - \sum_{s=2}^{q+1} v_{is}(b^{is}, 0_p) \right] - (1-r_i) \left[\sum_{s'=1}^{q+1} w_{is'}(c^{is'}, 0_p) \right] \quad (6.17)$$

with $r_i \in \{0, 1\}$ and the other components defined as in (6.15) and (6.16). Adding the terms $\mu_i u^i$ on θ (with u^i in (6.17)) componentwise, we arrive at

$$\sum_{i \in \theta} \mu_i u^i = \begin{bmatrix} -\sum_{i \in \theta} \sum_{s=2}^{q+1} r_i \mu_i v_{is} b^{is} - \sum_{i \in \theta} \sum_{s'=1}^{q+1} (1-r_i) \mu_i w_{is'} c^{is'} \\ \sum_{i \in \theta} r_i \mu_i v_{i1} e^i \end{bmatrix}.$$

Adding this part to (6.14) and inserting the outcome in (6.13), while considering the definitions of (6.15) and (6.16), we have the inclusion in (6.12).

By inserting (6.12) in (6.11), while considering Theorem 3.3 and inclusion (5.11), we get a desired C-counterpart of (5.12) by including conditions (6.2)–(6.4) and respectively replacing ζ and $\partial \langle -\beta, g \rangle(\bar{x}, \bar{y})$ by

$$\begin{aligned} & \sum_{i \in \eta} \mu_i e^i + \sum_{i \in \theta} r_i \mu_i v_{i1} e^i \text{ and} \\ & -\sum_{i \in \nu} \mu_i a^i - \sum_{i \in \theta} \sum_{s=2}^{q+1} r_i \mu_i v_{is} b^{is} - \sum_{i \in \theta} \sum_{s'=1}^{q+1} (1-r_i) \mu_i w_{is'} c^{is'}. \end{aligned}$$

Now consider the generalized equation $0 \in \partial \langle v, \psi \rangle(\bar{x}, \bar{y}, \bar{u})$ from (5.8) and the one in (5.7). Inserting the aforementioned C-counterpart of the upper bound of $\langle v, \psi \rangle(\bar{x}, \bar{y}, \bar{u})$ in these equations, their u -components both reduce to

$$\sum_{i \in \eta} \mu_i e^i + \sum_{i \in \theta} r_i \mu_i v_{i1} e^i + \left(\sum_{l=1}^m t_l^1 v_l, \dots, \sum_{l=1}^m t_l^p v_l \right)^\top = 0, \quad (6.18)$$

taking into account that F is independent of u . Considering the definition of e^i , we get (6.5)–(6.6) from equation (6.18).

Combining all the above, we can easily check that CQ (6.7) is a sufficient condition for the counterpart of (5.8) in the framework of (6.9) to hold. Similarly, we get the optimality conditions (1.8), (5.2), (6.2)–(6.6) and (6.8) via our C-counterpart of (5.7). It is important to note the replacement of μ_i by $-\mu_i$ all over, for $i = 1, \dots, p$. \square

Based on this result, we will say that a point (\bar{x}, \bar{y}) is SP-C-STATIONARY (resp. P-C-STATIONARY) if for every $\bar{u} \in \Lambda(\bar{x}, \bar{y})$ (resp. for some $\bar{u} \in \Lambda(\bar{x}, \bar{y})$), we can find $(\alpha, \mu, \gamma, \lambda)$ with $\lambda \in \mathbb{R}_+$, $t := (t^0, t^1, \dots, t^p)$, a^i with $i \in \nu$, $r_i \in \{0, 1\}$ with $i \in \theta$, v_{is} with $i \in \theta$ and $s = 1, \dots, q+1$, b^{is} with $i \in \theta$ and $s = 2, \dots, q+1$, $w_{is'}$ and $c^{is'}$ with $i \in \theta$ and $s' = 1, \dots, q+1$, such that (1.8), (5.2), (6.2)–(6.6) and (6.8) hold.

Similarly to Corollary 5.5, these stationarity conditions can respectively be derived from Theorem 6.2. To get a closer outlook between the C^o- and C-type stationarity concepts, observe that we can set

$$\begin{aligned} \beta_i &:= 0 \text{ for } i \in \eta, \beta_i := \mu_i \text{ for } i \in \nu, \\ \beta_{is} &:= r_i \mu_i v_s \text{ for } i \in \theta, s = 2, \dots, q+1, \\ \beta_{is'} &:= \mu_i w_{is'} (1 - r_i) \text{ for } i \in \theta, s' = 1, \dots, q+1. \end{aligned}$$

Further note that from (6.6), $\sum_{l=1}^m t_l^i \gamma_l = r_i \mu_i v_{i1}$ for $i \in \theta$. It is then easy to see that for all $i \in \theta$, $s = 2, \dots, q+1$, $s' = 1, \dots, q+1$, we have

$$\beta_{is} \left(\sum_{l=1}^m t_l^i \gamma_l \right) \geq 0, \beta_{is'} \left(\sum_{l=1}^m t_l^i \gamma_l \right) \geq 0.$$

This corresponds to the counterpart of (6.1) in the framework of the natural extension of the C-stationarity. Obviously, both the artificial and natural extensions of the C-stationarity conditions of problem (P) coincide in the smooth case.

It was observed in [38] that the smooth counterpart of CQ (6.7) is a quite strong assumption. Hence, we now attempt in the next result to extend a rather standard CQ in the MPCC theory, in order to derived the C-stationarity conditions defined above. Namely, we introduce a nonsmooth version of the MPEC-MFCQ tailored to the C-stationarity.

DEFINITION 6.3 (C-type MPEC-Mangasarian-Fromowitz CQ). *The C-MPEC-MFCQ holds at $(\bar{x}, \bar{y}, \bar{u})$ if for all $a_j^G \in \partial G_j(\bar{x})$ with $j \in J$; for all a^i with $i \in \nu$, $r_i \in \{0, 1\}$ with $i \in \theta$, v_{is} with $i \in \theta$ and $s = 1, \dots, q+1$, b^{is} with $i \in \theta$ and $s = 2, \dots, q+1$, $w_{is'}$ and $c^{is'}$ with $i \in \theta$ and $s' = 1, \dots, q+1$, verifying (6.2)–(6.4),*

$$\begin{aligned} \left[\sum_{i \in \nu} \mu_i a^i + \sum_{i \in \theta} \sum_{s=2}^{q+1} r_i \mu_i v_{is} b^{is} + \sum_{i \in \theta} \sum_{s'=1}^{q+1} (1 - r_i) \mu_i w_{is'} c^{is'} = 0, \right. \\ \left. \forall i \in \theta : r_i \mu_i v_{i1} = 0 \right] \implies \mu_i = 0, i \in \nu \cup \theta \end{aligned} \quad (6.19)$$

and there exists a vector $d := (d^x, d^y, d^u)$ with $d^{xy} := (d^x, d^y)$, such that:

$$\begin{aligned} \langle d^{xy}, a^i \rangle &= 0 \text{ for all } i \in \nu, \\ \langle d^{xy}, \sum_{s=2}^{q+1} r_i v_{is} b^{is} + \sum_{s'=1}^{q+1} (1 - r_i) w_{is'} c^{is'} \rangle &= 0 \text{ for all } i \in \theta, \\ \langle d^u, e^i \rangle &= 0 \text{ for all } i \in \eta, \\ \langle d^u, r_i v_{i1} e^i \rangle &= 0 \text{ for all } i \in \theta, \\ \langle d^x, a_j^G \rangle &< 0 \text{ for all } j \in J. \end{aligned} \quad (6.20)$$

Observe that in the smooth case (i.e., essentially if g is C^1), the C-MPEC-MFCQ thus not necessarily coincide with the M-MPEC-MFCQ, but the latter CQ will also lead to the following result.

THEOREM 6.4 (C-stationarity conditions under the C-MPEC-MFCQ). *Let $(\bar{x}, \bar{y}, \bar{u})$ be a local optimal solution of (4.1) and assume that the maps $\partial_y f$ and $\partial_y g_i$ for $i = 1, \dots, p$, are closed and uniformly bounded around (\bar{x}, \bar{y}) . Furthermore, let the C-MPEC-MFCQ be satisfied and (5.2) hold for all $t := (t^0, t^1, \dots, t^p)$. Then $(\bar{x}, \bar{y}, \bar{u})$ satisfies the optimality conditions in Theorem 6.2.*

Proof. Proceeding as in the proof of Theorem 5.4, it follows under the Fritz-John-type Lagrange multipliers rule of Mordukhovich that there exists $(\kappa, \alpha, \beta) \neq 0$ with $\kappa, \lambda \in \mathbb{R}_+$ and α satisfying (1.8), such that we have

$$0 \in \kappa \partial \mathcal{F}_\lambda(\bar{x}, \bar{y}, \bar{u}) + \sum_{j=1}^k \alpha_j (\partial G(\bar{x}), 0_{m+p}) + \sum_{i=1}^p \beta_i \bar{\partial} V_i(\bar{x}, \bar{y}, \bar{u}). \quad (6.21)$$

If we suppose that $\kappa = 0$, while considering the upper estimate of $\sum_{i=1}^p \beta_i \bar{\partial} V_i(\bar{x}, \bar{y}, \bar{u})$ from (6.12), it follows that we can find some a^i with $i \in \nu$, $r_i \in \{0, 1\}$ with $i \in \theta$, v_{is} with $i \in \theta$ and $s = 1, \dots, q+1$, b^{is} with $i \in \theta$ and $s = 2, \dots, q+1$, $w_{is'}$ and $c^{is'}$ with $i \in \theta$ and $s' = 1, \dots, q+1$, satisfying (6.2)–(6.4) such that

$$\begin{aligned} & \sum_{j \in J} \alpha_j (a_j^G, 0_{m+p}) \\ & + \left[\begin{array}{l} \sum_{i \in \nu} \mu_i a^i + \sum_{i \in \theta} \sum_{s=2}^{q+1} r_i \mu_i v_{is} b^{is} + \sum_{i \in \theta} \sum_{s'=1}^{q+1} \mu_i (1 - r_i) w_{is'} c^{is'} \\ \sum_{i \in \eta} \mu_i e^i + \sum_{i \in \theta} r_i \mu_i v_{i1} e^i \end{array} \right] = 0 \quad (6.22) \end{aligned}$$

for some $a_j^G \in \partial G_j(\bar{x})$ with $j \in J$. Now consider a vector $d := (d^x, d^y, d^u)$ with $d^{xy} := (d^x, d^y)$ that satisfies (6.20). Then multiplying (6.22) with d , we obtain

$$\begin{aligned} & \sum_{j \in J} \alpha_j \langle a_j^G, d^x \rangle + \sum_{i \in \nu} \mu_i \langle a^i, d^{xy} \rangle \\ & + \sum_{i \in \theta} \sum_{s=2}^{q+1} r_i \mu_i v_{is} \langle b^{is}, d^{xy} \rangle + \sum_{i \in \theta} \sum_{s'=1}^{q+1} \mu_i (1 - r_i) w_{is'} \langle c^{is'}, d^{xy} \rangle \\ & + \sum_{i \in \eta} \mu_i \langle e^i, d^u \rangle + \sum_{i \in \theta} r_i \mu_i v_{i1} \langle e^i, d^u \rangle = 0. \end{aligned}$$

Further proceeding as in proof of Theorem 5.4, we respectively have $\alpha = 0$ from (6.19) and $\mu = 0$ from (6.20). Thus contradicting the fact that $(\kappa, \alpha, \beta) \neq 0$. Hence, by similarly setting $\kappa = 1$ in (6.21), we have the result by considering the estimate of $\partial \mathcal{F}_\lambda(\bar{x}, \bar{y}, \bar{u})$ from (5.21) and that of $\sum_{i=1}^p \beta_i \bar{\partial} V_i(\bar{x}, \bar{y}, \bar{u})$ from (6.12). \square

7. S-stationarity in the nonsmooth case. To motivate the discussion in this section, we first recall the following result from [38], establishing the equivalence between the P-S-stationarity from Definition 1.2 (iii) and the so-called KKT necessary optimality conditions of the smooth KKT reformulation (1.3).

PROPOSITION 7.1 (characterization of the P-S-stationarity in the smooth case). *The point (\bar{x}, \bar{y}) is P-S-stationary if and only if there exists $(\alpha, \beta, \bar{u}, \gamma, \xi)$ with $\xi \in \mathbb{R}_+$*

such that (1.8) holds together with the following conditions:

$$\begin{aligned} & \nabla_x F(\bar{x}, \bar{y}) + \sum_{i=1}^p (\beta_i - \xi \bar{u}_i) \nabla_x g_i(\bar{x}, \bar{y}) \\ & + \sum_{j=1}^k \alpha_j \nabla G_j(\bar{x}) + \sum_{l=1}^m \gamma_l \nabla_x \mathcal{L}_l(\bar{x}, \bar{y}, \bar{u}) = 0, \end{aligned} \quad (7.1)$$

$$\nabla_y F(\bar{x}, \bar{y}) + \sum_{i=1}^p (\beta_i - \xi \bar{u}_i) \nabla_y g_i(\bar{x}, \bar{y}) + \sum_{l=1}^m \gamma_l \nabla_y \mathcal{L}_l(\bar{x}, \bar{y}, \bar{u}) = 0, \quad (7.2)$$

$$\nabla_y f(\bar{x}, \bar{y}) + \sum_{i=1}^p \bar{u}_i \nabla_y g_i(\bar{x}, \bar{y}) = 0, \quad (7.3)$$

$$\forall i = 1, \dots, p : \bar{u}_i \geq 0, \bar{u}_i g_i(\bar{x}, \bar{y}) = 0, \beta_i \geq 0, \beta_i g_i(\bar{x}, \bar{y}) = 0, \quad (7.4)$$

$$\forall i = 1, \dots, p : \sum_{l=1}^m \gamma_l \nabla_{y_l} g_i(\bar{x}, \bar{y}) - \xi g_i(\bar{x}, \bar{y}) \geq 0, \bar{u}_i \sum_{l=1}^m \gamma_l \nabla_{y_l} g_i(\bar{x}, \bar{y}) = 0. \quad (7.5)$$

The combination of (1.8) and (7.1)–(7.5) corresponds to the KKT necessary optimality conditions of (1.3), which can be obtained at least by two possible ways. The first one is the application of Guignard's CQ, which as one of the weakest CQ in optimization is satisfied for some classes of MPCCs/bilevel programs [12, 38]. The second approach is the application of the partial calmness to help move the function $(x, y, u) \rightarrow \sum_{i=1}^p u_i g_i(x, y)$ from the constraints to the objective function. Afterwards, any other usual CQ can then be used to derive the conditions [38]. In the perspective to extend these ideas to the nonsmooth case, we now derive, in the next result, the Fritz-John counterpart of the KKT type necessary optimality conditions of problem (4.1) in terms of the generalized differentiation tools defined in Section 2.

THEOREM 7.2 (Fritz-John's type optimality conditions for (4.1)). *Let $(\bar{x}, \bar{y}, \bar{u})$ be a local optimal solution of (4.1) and the set-valued maps $\partial_y f$ and $\partial_y g_i$ for $i = 1, \dots, p$, be closed and uniformly bounded around (\bar{x}, \bar{y}) . Moreover, if for all $t := (t^0, t^1, \dots, t^p)$ satisfying (5.2), the QC (3.16) holds, then there exist (t^0, t^1, \dots, t^p) , (α, β, γ) and $\kappa, \xi, \lambda \in \mathbb{R}_+$ with $\|\gamma\| \leq \lambda$ such that (1.8), (5.2) and (7.4) hold together with:*

$$\begin{aligned} & 0 \in \kappa \partial F(\bar{x}, \bar{y}) + \sum_{j=1}^k \alpha_j (\partial G_j(\bar{x}), 0_m) \\ & + D^*(\partial_y f)((\bar{x}, \bar{y})|t^0)(\gamma) + \sum_{i=1}^p D^*(\partial_y g_i)((\bar{x}, \bar{y})|t^i)(\bar{u}_i \gamma) \\ & + \sum_{i=1}^p \left\{ \beta_i \bar{\partial} g_i(\bar{x}, \bar{y}) - \xi \bar{u}_i \bar{\partial} g_i(\bar{x}, \bar{y}) \right\}, \end{aligned} \quad (7.6)$$

$$\forall i = 1, \dots, p : \sum_{l=1}^m t_l^i \gamma_l - \xi g_i(\bar{x}, \bar{y}) \geq 0, \quad u_i \sum_{l=1}^m t_l^i \gamma_l = 0. \quad (7.7)$$

Proof. Observe that (4.1) can fully be rewritten in terms of inequality constraints:

$$\begin{aligned} \min_{x, y, u} \{ & F(x, y) \mid h(x, y, u) \leq 0, G_j(x) \leq 0, j = 1, \dots, k, \\ & -u \leq 0, g(x, y) \leq 0, \chi(x, y, u) := \sum_{i=1}^p (-u_i) g_i(x, y) \leq 0 \}, \end{aligned}$$

with h defined as in (5.6). This is a Lipschitz optimization problem. Thus, by the Fritz-John type Lagrange multiplier rule in terms of the limiting subdifferential, we can find $(\kappa, \alpha, \lambda, \mu, \beta, \xi)$ with $\kappa, \lambda, \xi \in \mathbb{R}_+$, such that (1.8), (7.4) and the conditions

$$\begin{aligned} 0 \in & \kappa(\partial F(\bar{x}, \bar{y}), 0_p) + \sum_{j=1}^k \alpha_j(\partial G_j(\bar{x}), 0_{m+p}) + \lambda \partial h(\bar{x}, \bar{y}, \bar{u}) \\ & + (0_{n+m}, -\mu) + \sum_{i=1}^p \beta_i(\partial g_i(\bar{x}, \bar{y}), 0_p) + \xi \partial \chi(\bar{x}, \bar{y}, \bar{u}), \end{aligned} \quad (7.8)$$

$$\forall i = 1, \dots, p : \mu_i \geq 0, \mu_i \bar{u}_i = 0, \quad (7.9)$$

hold. Applying the sum and product rules of basic subdifferentials to the function χ , while taking into account that the plus/minus symmetry holds for the Clarke subdifferential, we arrive at the following inclusion

$$\partial \chi(\bar{x}, \bar{y}, \bar{u}) \subseteq (0_{n+m}, -g(\bar{x}, \bar{y})) - \sum_{i=1}^p \bar{u}_i(\bar{\partial} g_i(\bar{x}, \bar{y}), 0_p). \quad (7.10)$$

Now consider the estimate of $\partial h(\bar{x}, \bar{y}, \bar{u})$ from Theorem 3.3 and (5.11), then we automatically get (5.2). Moreover, the (x, y) -component of (7.8) generates (7.6), while taking (7.10) into account. Finally, (7.7) results from the combination of (7.7) and the u -components of (7.8), (7.10) and of the just mentioned estimate of $\partial h(\bar{x}, \bar{y}, \bar{u})$. \square

On the CQs to get $\kappa = 1$ in the above result, note that it is not yet clear whether the partial calmness CQ mentioned above would still work here, as the proof requires differentiability of the functions [37, 38]. As for the Guignard CQ, if we define the linearized tangent cone in terms of the Clarke directional derivative, a result closely related to the above one (with $\kappa = 1$) can be generated by directly extending [38, Theorem 3.3.8] to the current Lipschitz case, provided the upper-level objective function is \mathcal{C}^1 . These topics will be carefully addressed in a future research.

To close this section, we assume that we are in a position to derive the KKT necessary optimality conditions of (4.1) in Theorem 7.2 with $\kappa = 1$. So the question is whether the result in Proposition 7.1 can be extended to the nonsmooth case. Concretely, following the patterns in the previous sections, we would say that a point (\bar{x}, \bar{y}) is P-S-stationarity in the nonsmooth case, if for some $\bar{u} \in \Lambda(\bar{x}, \bar{y})$, there exist $t := (t^0, t^1, \dots, t^p)$ and $(\alpha, \beta, \gamma, \lambda)$ with $\lambda \in \mathbb{R}_+$ such that (1.8), (5.2)–(5.3) and (5.14) hold together with the following condition

$$\forall i \in \theta : \beta_i \geq 0, \sum_{l=1}^m t_l^i \gamma_l \geq 0.$$

If the lower-level constraint function g is \mathcal{C}^1 , the result can easily be extended to (4.1), cf. [38, Proof of Theorem 3.1.9]. Otherwise, it is not difficult to find examples where the inclusion

$$\beta_i \bar{\partial} g_i(\bar{x}, \bar{y}) - \xi \bar{u}_i \bar{\partial} g_i(\bar{x}, \bar{y}) \supset (\beta_i - \xi \bar{u}_i) \bar{\partial} g_i(\bar{x}, \bar{y}) \quad (7.11)$$

is strict, when the cardinality of the set $\partial g_i(\bar{x}, \bar{y})$ is more than one, cf. Subsection 8.3. This seems therefore to suggest that the well-known result in MPCC/MPEC theory that the KKT necessary optimality conditions are equivalent to the S-stationarity concept (see Proposition 7.1 for the bilevel programming-counterpart) is not valid in the nonsmooth case; at least in the framework of the bilevel program.

8. A numerical example. In the example that we consider here, the coderivative calculations involve the evaluation of normal cones to unions of finite numbers of sets. Thus, we use the following formulas to proceed, see, e.g., [12, 34] for details. Let Ω_1 and Ω_2 be two closed subsets of \mathbb{R}^n . If Ω_1 is nonempty and $\bar{x} \in \Omega_1 \setminus \Omega_2$, then

$$\widehat{N}_{\Omega_1 \cup \Omega_2}(\bar{x}) = \widehat{N}_{\Omega_1}(\bar{x}) \quad (8.1)$$

whereas, if $\Omega_1 \cap \Omega_2$ is nonempty and $\bar{x} \in \Omega_1 \cap \Omega_2$, it holds that

$$\widehat{N}_{\Omega_1 \cup \Omega_2}(\bar{x}) = \widehat{N}_{\Omega_1}(\bar{x}) \cap \widehat{N}_{\Omega_2}(\bar{x}). \quad (8.2)$$

Next we consider a case of the optimistic bilevel optimization problem (P) in \mathbb{R}^2 and with a single upper and lower-level constraint, where $F(x, y) := |x - y|$, $G(x) := -x$, $f(x, y) := \max\{x, y\}$, and $g(x, y) := |y| - x$. One can easily check that $S(x) = [-x, x]$ for all $x \geq 0$. Thus implying that we have

$$\underset{x, y}{\operatorname{Argmin}} \{F(x, y) \mid G(x) \leq 0, y \in S(x)\} = \{(x, y) \mid y = x \geq 0\}.$$

8.1. Computing the coderivatives. The functions f and g are convex w.r.t. y and we have

$$\partial_y f(x, y) = \begin{cases} 0 & \text{if } x > y, \\ 1 & \text{if } x < y, \\ [0, 1] & \text{if } x = y, \end{cases} \quad \text{and} \quad \partial_y g(x, y) = \begin{cases} 1 & \text{if } y > 0, \\ -1 & \text{if } y < 0, \\ [-1, 1] & \text{if } y = 0. \end{cases}$$

These maps are obviously uniformly bounded. The graph of $\partial_y f$ can take the form

$$\operatorname{gph}(\partial_y f) = \{(x, y, 0) \mid x > y\} \cup \{(x, y, 1) \mid x < y\} \cup \{(x, y, z) \mid x = y, z \in [0, 1]\}.$$

Obviously, this set can be rewritten as $\operatorname{gph}(\partial_y f) = A \cup B \cup C$, where

$$A := \{(x, y, 0) \mid x \geq y\}, \quad B := \{(x, y, 1) \mid x \leq y\}, \quad \text{and} \quad C := \{(x, y, z) \mid x = y, z \in [0, 1]\}.$$

Thus, as the union of three closed sets, $\operatorname{gph}(\partial_y f)$ is also a closed set.

Considering the definition of the coderivative and the interplay between the Mordukhovich and Fréchet normal cones in (2.1), we first provide the expressions of $\widehat{N}_{\operatorname{gph}(\partial_y f)}(\bar{x}, \bar{y}, \bar{z})$. Five cases are considered:

(i) If $\bar{x} > \bar{y}$ and $\bar{z} = 0$, then $(\bar{x}, \bar{y}, \bar{z}) \in A$ and $(\bar{x}, \bar{y}, \bar{z}) \notin B \cup C$. Moreover, since $B \cup C$ is closed, we have from (8.1) that $\widehat{N}_{\operatorname{gph}(\partial_y f)}(\bar{x}, \bar{y}, \bar{z}) = \widehat{N}_A(\bar{x}, \bar{y}, \bar{z})$. On the other hand, one can easily check that $\widehat{N}_A(\bar{x}, \bar{y}, \bar{z}) = \{(0, 0)\} \times \mathbb{R}$. Thus $\widehat{N}_{\operatorname{gph}(\partial_y f)}(\bar{x}, \bar{y}, \bar{z}) = \{(0, 0)\} \times \mathbb{R} := \Omega_1$.

(ii) If $\bar{x} < \bar{y}$ and $\bar{z} = 1$, then $(\bar{x}, \bar{y}, \bar{z}) \in B$ and $(\bar{x}, \bar{y}, \bar{z}) \notin A \cup C$. Since $A \cup C$ is closed, we get $\widehat{N}_{\operatorname{gph}(\partial_y f)}(\bar{x}, \bar{y}, \bar{z}) = \widehat{N}_B(\bar{x}, \bar{y}, \bar{z}) = \Omega_1$ while proceeding as in the previous case.

(iii) If $\bar{x} = \bar{y}$ and $0 < \bar{z} < 1$, then $(\bar{x}, \bar{y}, \bar{z}) \in C$ and $(\bar{x}, \bar{y}, \bar{z}) \notin A \cup B$. Similarly, $\widehat{N}_{\operatorname{gph}(\partial_y f)}(\bar{x}, \bar{y}, \bar{z}) = \widehat{N}_C(\bar{x}, \bar{y}, \bar{z}) = \{(x, -x) : x \in \mathbb{R}\} \times \{0\} := \Omega_2$, as $A \cup B$ is closed.

(iv) If $\bar{x} = \bar{y}$ and $\bar{z} = 0$, then $(\bar{x}, \bar{y}, \bar{z}) \in A \cap C$ and $(\bar{x}, \bar{y}, \bar{z}) \notin B$. Combining (8.1) and (8.2) it holds that $\widehat{N}_{\operatorname{gph}(\partial_y f)}(\bar{x}, \bar{y}, \bar{z}) = \widehat{N}_{A \cap C}(\bar{x}, \bar{y}, \bar{z})$ given that B and $A \cup C$ are closed sets. One can easily check that $\widehat{N}_A(\bar{x}, \bar{y}, \bar{z}) = \{(x, -x) : x \leq 0\} \times \mathbb{R}$. and $\widehat{N}_C(\bar{x}, \bar{y}, \bar{z}) = \{(x, -x) : x \in \mathbb{R}\} \times \mathbb{R}_-$. In conclusion for this case, we have

$$\widehat{N}_{\text{gph}(\partial_y f)}(\bar{x}, \bar{y}, \bar{z}) = \{(x, -x) : x \leq 0\} \times \mathbb{R}_- := \Omega_3.$$

(v) If $\bar{x} = \bar{y}$ and $\bar{z} = 1$, then $(\bar{x}, \bar{y}, \bar{z}) \in B \cap C$ and $(\bar{x}, \bar{y}, \bar{z}) \notin A$. Following the same path as in (iv), $\widehat{N}_{\text{gph}(\partial_y f)}(\bar{x}, \bar{y}, \bar{z}) = \widehat{N}_{B \cap C}(\bar{x}, \bar{y}, \bar{z})$. By simple calculations, $\widehat{N}_B(\bar{x}, \bar{y}, \bar{z}) = \{(x, -x) : x \geq 0\} \times \mathbb{R}$ and $\widehat{N}_C(\bar{x}, \bar{y}, \bar{z}) = \{(x, -x) : x \in \mathbb{R}\} \times \mathbb{R}_+$. Thus, $\widehat{N}_{\text{gph}(\partial_y f)}(\bar{x}, \bar{y}, \bar{z}) = \{(x, -x) : x \geq 0\} \times \mathbb{R}_+ := \Omega_4$.

In summary, we have

$$\widehat{N}_{\text{gph}(\partial_y f)}(\bar{x}, \bar{y}, \bar{z}) = \begin{cases} \Omega_1 & \text{if } (\bar{x} > \bar{y}, \bar{z} = 0) \vee (\bar{x} < \bar{y}, \bar{z} = 1), \\ \Omega_2 & \text{if } \bar{x} = \bar{y}, 0 < \bar{z} < 1, \\ \Omega_3 & \text{if } \bar{x} = \bar{y}, \bar{z} = 0, \\ \Omega_4 & \text{if } \bar{x} = \bar{y}, \bar{z} = 1. \end{cases}$$

The sequence $(\frac{1}{2^n}, \frac{1}{n}, 0)$ satisfies $\frac{1}{2^n} < \frac{1}{n}$ for all $n \geq 1$ and $\frac{1}{2^n} \rightarrow 0$ and $\frac{1}{n} \rightarrow 0$. A similar observation can be made for the sequence $(\frac{1}{n}, \frac{1}{2^n}, 1)$. Thus from the first line of the latter formula, we have $\widehat{N}_{\text{gph}(\partial_y f)}(\frac{1}{2^n}, \frac{1}{n}, 0) = \widehat{N}_{\text{gph}(\partial_y f)}(\frac{1}{n}, \frac{1}{2^n}, 1) = \Omega_1$. Let us also consider the sequences $(\frac{1}{n}, \frac{1}{n}, \frac{1}{2^n})$ and $(\frac{1}{n}, \frac{1}{n}, 1 - \frac{1}{2^n})$ with $0 < \frac{1}{2^n} < 1$ and $0 < 1 - \frac{1}{2^n} < 1$ for all $n \geq 1$. Obviously, $\frac{1}{2^n} \rightarrow 0$ and $1 - \frac{1}{2^n} \rightarrow 1$. Thus, by the second line of the above formula, $\widehat{N}_{\text{gph}(\partial_y f)}(\frac{1}{n}, \frac{1}{n}, \frac{1}{2^n}) = \widehat{N}_{\text{gph}(\partial_y f)}(\frac{1}{n}, \frac{1}{n}, 1 - \frac{1}{2^n}) = \Omega_2$. Considering the behavior of these sequences, we get the following expression for the Mordukhovich normal cone to $\text{gph}(\partial_y f)$, by applying (2.1):

$$N_{\text{gph}(\partial_y f)}(\bar{x}, \bar{y}, \bar{z}) = \begin{cases} \Omega_1 & \text{if } (\bar{x} > \bar{y}, \bar{z} = 0) \vee (\bar{x} < \bar{y}, \bar{z} = 1), \\ \Omega_2 & \text{if } \bar{x} = \bar{y}, 0 < \bar{z} < 1, \\ \Omega_3 \cup \Omega_1 \cup \Omega_2 & \text{if } \bar{x} = \bar{y}, \bar{z} = 0, \\ \Omega_4 \cup \Omega_1 \cup \Omega_2 & \text{if } \bar{x} = \bar{y}, \bar{z} = 1. \end{cases}$$

Taking the expressions of the sets Ω_i with $i = 1, 2, 3, 4$ into account, it follows from the definition of the coderivative in (2.5) that

$$D^*(\partial_y f)((\bar{x}, \bar{y})|\bar{z})(z^*) = \begin{cases} \{(0, 0)\} & \text{if } \bar{x} > \bar{y}, \bar{z} = 0, z^* \in \mathbb{R}, \\ \{(0, 0)\} & \text{if } \bar{x} < \bar{y}, \bar{z} = 1, z^* \in \mathbb{R}, \\ \{(x, -x) : x \in \mathbb{R}\} & \text{if } \bar{x} = \bar{y}, 0 < \bar{z} < 1, z^* = 0, \\ \emptyset & \text{if } \bar{x} = \bar{y}, 0 < \bar{z} < 1, z^* \neq 0, \\ \{(x, -x) : x \in \mathbb{R}\} & \text{if } \bar{x} = \bar{y}, \bar{z} = 0, z^* = 0, \\ \{(x, -x) : x < 0\} & \text{if } \bar{x} = \bar{y}, \bar{z} = 0, z^* > 0, \\ \{(0, 0)\} & \text{if } \bar{x} = \bar{y}, \bar{z} = 0, z^* < 0, \\ \{(x, -x) : x \in \mathbb{R}\} & \text{if } \bar{x} = \bar{y}, \bar{z} = 1, z^* = 0, \\ \{(x, -x) : x > 0\} & \text{if } \bar{x} = \bar{y}, \bar{z} = 1, z^* < 0, \\ \{(0, 0)\} & \text{if } \bar{x} = \bar{y}, \bar{z} = 1, z^* > 0. \end{cases}$$

Note that the graph of $\partial_y g$ is equally closed. Further proceeding as above, we get

$$D^*(\partial_y g)((\bar{x}, \bar{y})|\bar{z})(z^*) = \begin{cases} \{(0, 0)\} & \text{if } \bar{x} \in \mathbb{R}, \bar{y} > 0, \bar{z} = 1, z^* \in \mathbb{R}, \\ \{(0, 0)\} & \text{if } \bar{x} \in \mathbb{R}, \bar{y} < 0, \bar{z} = -1, z^* \in \mathbb{R}, \\ \{0\} \times \mathbb{R} & \text{if } \bar{x} \in \mathbb{R}, \bar{y} = 0, -1 < \bar{z} < 1, z^* = 0, \\ \emptyset & \text{if } \bar{x} \in \mathbb{R}, \bar{y} = 0, -1 < \bar{z} < 1, z^* \neq 0, \\ \{0\} \times \mathbb{R}_+ & \text{if } \bar{x} \in \mathbb{R}, \bar{y} = 0, \bar{z} = -1, z^* > 0, \\ \{(0, 0)\} & \text{if } \bar{x} \in \mathbb{R}, \bar{y} = 0, \bar{z} = -1, z^* < 0, \\ \{0\} \times \mathbb{R} & \text{if } \bar{x} \in \mathbb{R}, \bar{y} = 0, \bar{z} = -1, z^* = 0, \\ \{0\} \times \mathbb{R}_- & \text{if } \bar{x} \in \mathbb{R}, \bar{y} = 0, \bar{z} = 1, z^* < 0, \\ \{(0, 0)\} & \text{if } \bar{x} \in \mathbb{R}, \bar{y} = 0, \bar{z} = 1, z^* > 0, \\ \{0\} \times \mathbb{R} & \text{if } \bar{x} \in \mathbb{R}, \bar{y} = 0, \bar{z} = 1, z^* = 0. \end{cases}$$

8.2. Verifying the CQs, M and C-stationarity. First observe that the Slater CQ is satisfied at all $x > 0$. In this subsection, we focus our attention on the point $(\bar{x}, \bar{y}) := (1, 1)$, which is an element of the set of optimal solutions of the upper-level problem. Obviously, $\Lambda(1, 1) = \{0\}$ and $\partial_y g(1, 1) = \{1\}$. Also note that for all $\gamma \in \mathbb{R}$, $D^*(\partial_y g)((1, 1)|1)(\gamma) = \{(0, 0)\}$. Thus the condition (3.16) is automatically satisfied. Moreover, it follows that the only vector (t^0, t^1) verifying (5.2) is $(0, 1)$. This implies that the SOLLCCQ (5.13) holds. To guaranty the M-stationarity, it remains to show that the M-MPEC-MFCQ in Definition 5.3 is satisfied. This is obviously the case with $d = (1, 1, 0)^\top$ since $\theta = \{1\}$ and $\partial g(1, 1) = \{(-1, 1)^\top\}$. Now recall from the expression of the coderivative of $\partial_y f$ above that

$$D^*(\partial_y f)((1, 1)|0)(\gamma) = \begin{cases} \{(x, -x) : x \in \mathbb{R}\} & \text{if } \gamma = 0, \\ \{(x, -x) : x < 0\} & \text{if } \gamma > 0, \\ \{(0, 0)\} & \text{if } \gamma < 0. \end{cases}$$

The M-stationarity conditions hold for example with $(\alpha, \beta, \gamma) = (0, 1, -1)$ or $(0, 1, 0)$. Since the subdifferential of g is a singleton here, the C-stationarity conditions are also satisfied under the M-MPEC-MFCQ as mentioned in Section 6.

8.3. Analyzing the S-stationarity. For $(\bar{x}, \bar{y}) := (0, 0)$, the Slater CQ and MFCQ both fail for the lower-level problem. But as for all $x \geq 0$, the mapping

$$\Psi^x(v) := \{y \in \mathbb{R} : |y| - x \leq v\}$$

is calm at any point of its graph, the lower-level KKT conditions remain valid. Thus, applying Theorem 4.1, $(0, 0, u)$ is a local optimal solution of the KKT reformulation for any $u \in \Lambda(0, 0)$. Notice that $\Lambda(0, 0) := \{u \in \mathbb{R}_+ : 0 \in [0, 1] + u[-1, 1]\}$. From the formulas above, it holds that for all $t^0 \in [0, 1]$ and $t^1 \in [-1, 1]$,

$$D^*(\partial_y f)((0, 0)|t^0)(0) = \{(x, -x) : x \in \mathbb{R}\} \text{ and } D^*(\partial_y g)((0, 0)|t^1)(0) = \{0\} \times \mathbb{R}. \quad (8.3)$$

With these expressions, one can easily check that QC (3.16) holds for any point $(0, 0, u)$, where $u \in \Lambda(0, 0)$.

Choose $\bar{u} = 1$, then we have the following implications

$$\begin{aligned} (t^0, t^1) \text{ satisfies (5.2)} &\implies (t^0, t^1) \in \{(t^0, -t^0) : t^0 \in [0, 1]\}, \\ (7.7) &\implies t^1 = 0 \text{ or } \gamma = 0. \end{aligned}$$

Take $\gamma = 0$ and $t^0 = 1$ (that is, $t^1 = -1$), then the corresponding coderivatives are obtained as in (8.3).

Now note that the functions F and g are both convex w.r.t. (x, y) and we respectively have $\partial F(0, 0) = \text{co} \{(1, -1)^\top, (-1, 1)^\top\}$ and $\partial g(0, 0) = \text{co} \{(-1, 1)^\top, (-1, -1)^\top\}$. Thus, we get the following equation from (7.6) while setting $\kappa = 1$, $\xi = \beta$, and considering $(-1, \frac{1}{2})^\top \in \partial g(0, 0)$ and the origin point from the coderivatives in (8.3):

$$\begin{bmatrix} 1 \\ -1 \end{bmatrix} + \alpha \begin{bmatrix} -1 \\ 0 \end{bmatrix} + \beta \begin{bmatrix} -1 \\ 1 \end{bmatrix} - \beta \begin{bmatrix} -1 \\ \frac{1}{2} \end{bmatrix} = 0.$$

This leads to $\alpha = 1$ and $\beta = 2$. As a summary of the discussion, this means that a vector $(t^0, t^1, \alpha, \beta, \gamma, \kappa, \xi, \lambda)$ with $t^0 = 1, t^1 = -1, \alpha = 1, \beta = 2, \gamma = 0, \kappa = 1, \xi = 2$ and $\lambda \in \mathbb{R}_+$ satisfies the optimality conditions in Theorem 7.2 when $\bar{x} = \bar{y} = 0$ and $\bar{u} = 1$.

An important point that we would like to make here is to show that inclusion (7.11) is strict in this case. In fact, note that the right hand side is the singleton $\{(0, 0)\}$ under the values obtained above. On the other hand, observe that in addition to $(0, 0)$, the left hand side also contains the point $(0, 1)^\top = 2(-1, 1)^\top - 2(-1, \frac{1}{2})^\top$.

9. Conclusion and extensions. We have considered the KKT reformulation of a nonsmooth bilevel program and study the relationship between both problems. The extension of various stationarity concepts to this case are also discussed under dual type constraint qualifications (CQs). Note that the CQs in Theorem 5.2 and Theorem 6.2 can be replaced by the calmness property of adequately chosen set-valued maps, see the comments immediately after Theorem 5.2. Following the pattern of results in [7, 8], the ideas developed in this paper can readily be applied for the stability analysis of the two-level value functions

$$\varphi_o(x) := \min_y \{F(x, y) \mid y \in S(x)\} \text{ and } \varphi_p(x) := \max_y \{F(x, y) \mid y \in S(x)\}$$

(S is defined as in (1.1)), in the case where the functions involved are nonsmooth. Concretely, this would consist to consider the extended solutions of the lower-level problem described by the set-valued mapping

$$S^h(x) := \{(y, u) \mid 0 \in \mathcal{L}(x, y, u), u \geq 0, g(x, y) \leq 0, u^\top g(x, y) = 0\}.$$

M- and C-type upper estimates of the coderivative of this map can be obtained by the same techniques used in Theorems 5.2 and 6.2, respectively. This can then lead directly to necessary optimality conditions for *strong Stackelberg/original optimistic* and *weak Stackelberg/pessimistic* problems [7, 8]. The latter problems respectively consist to minimize φ_o and φ_p on the upper-level feasible set X . To ensure the Lipschitz continuity of φ_o and φ_p , it is enough to derive conditions for the Aubin property of S^h [8] by using the well-known coderivative criterion [21, 29].

Acknowledgements. The authors are grateful to René Henrion for his useful remarks and suggestions on an initial version of the paper. We are also indebted to two anonymous referees for their constructive remarks, which allowed us to improve the quality of the paper.

REFERENCES

- [1] G.B. ALLENDE AND G. STILL, *Solving bilevel programs with the KKT–approach*, Math. Program., 138 (2013), pp. 309–332.
- [2] T.Q. BAO, P. GUPTA, AND B.S. MORDUKHOVICH, *Necessary conditions in multiobjective optimization with equilibrium constraints*, J. Optim. Theory Appl., 135 (2007), pp. 179–203.
- [3] F.H. CLARKE, *Optimization and nonsmooth analysis*, Wiley, New York, 1983.
- [4] S. DEMPE AND J. DUTTA, *Is bilevel programming a special case of mathematical programming with equilibrium constraints?*, Math. Program., 131 (2012), pp. 37–48.
- [5] S. DEMPE, J. DUTTA, AND B.S. MORDUKHOVICH, *New necessary optimality conditions in optimistic bilevel programming*, Optimization, 56 (2007), pp. 577–604.
- [6] S. DEMPE, N. GADHI, AND A.B. ZEMKOHO, *New optimality conditions in semivectorial bilevel optimization*, J. Optim. Theory Appl., 157 (2013), pp. 54–74.
- [7] S. DEMPE, B.S. MORDUKHOVICH, AND A.B. ZEMKOHO, *Necessary optimality conditions in pessimistic bilevel programming*, Optimization, 63 (2014), pp. 505–533.
- [8] 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), pp. 1309–1343.
- [9] S. DEMPE AND A.B. ZEMKOHO, *On the Karush-Kuhn-Tucker reformulation of the bilevel optimization problem*, Nonlinear Anal., 75 (2012), pp. 1202–1218.
- [10] X. DENG, *Complexity issues in bilevel linear programming*, Multilevel Optimization: Algorithms and Applications (A. Migdalas, P.M. Pardalos, and P. Värbrand, eds.), Kluwer Academic Publishers, Dordrecht, 1998, pp. 149–164.
- [11] B.V. DINH, P.G. HUNG, AND L.D. MUU, *Bilevel optimization as a regularization approach to pseudomonotone equilibrium problems*, Num. Fun. Anal. Optim., 35 (2014), pp. 539–563.

- [12] M. FLEGEL, *Constraint qualification and stationarity concepts for mathematical programs with equilibrium constraints*, Ph.D. thesis, Institute of Applied Mathematics and Statistics, University of Würzburg, 2005.
- [13] M.L. FLEGEL, C. KANZOW, AND J.V. OUTRATA, *Optimality conditions for disjunctive programs with application to mathematical programs with equilibrium constraints*, *J. Set-Valued Anal.*, 15 (2007), pp. 139–162.
- [14] R. HENRION, A. JOURANI, AND J.V. OUTRATA, *On the calmness of a class of multifunctions*, *SIAM J. Optim.*, 13 (2002), 603–618.
- [15] R. HENRION AND T. SUROWIEC, *On calmness conditions in convex bilevel programming*, *Applicable Anal.*, 90 (2011), pp. 951–970.
- [16] J.-B. HIRIART-URRUTY, *Refinements of necessary optimality conditions in nondifferentiable programming. I*, *Appl. Math. Optim.*, 5 (1979), pp. 63–82.
- [17] H.TH. JONGEN, J.-J. RÜCKMANN, AND V. SHIKHMAN, *MPCC: critical point theory*, *SIAM J. Optim.*, 20 (2009), pp. 473–484.
- [18] K. KUNISCH AND T. POCK, *A bilevel optimization approach for parameter learning in variational models*, *SIAM J. Imaging Sci.*, 6 (2013), pp. 938–983.
- [19] D. LI, L.-P. PANG, F.-F. GUO, AND Z.-Q. XIA, *An alternating linearization method with inexact data for bilevel nonsmooth convex optimization*, *J. Ind. Manag. Optim.*, 10 (2014), pp. 859–869.
- [20] B.S. MORDUKHOVICH, *Generalized differential calculus for nonsmooth and set-valued mappings*, *J. Math. Anal. Appl.*, 183 (1994), pp. 250–288.
- [21] B.S. MORDUKHOVICH, *Variational analysis and generalized differentiation. I: Basic theory. II: Applications*, Springer, Berlin, 2006.
- [22] B.S. MORDUKHOVICH, *Metric approximations and necessary optimality conditions for general classes of nonsmooth extremal problems*, *Sov. Math., Dokl.*, 22 (1980), pp. 526–530.
- [23] B.S. MORDUKHOVICH, N.M. NAM, AND H.M. PHAN, *Variational analysis of marginal functions with applications to bilevel programming problems*, *J. Optim. Theory Appl.*, 152 (2012), pp. 557–586.
- [24] B.S. MORDUKHOVICH AND R.T. ROCKAFELLAR, *Second-order subdifferential calculus with applications to tilt stability in optimization*, *SIAM J. Optim.*, 22 (2012), pp. 953–986.
- [25] N. MOVAHEDIAN AND S. NOBAKHIAN, *Constraint qualifications for nonsmooth mathematical programs with equilibrium constraints*, *Set-Valued Anal.*, 17 (2009), pp. 63–95.
- [26] N. MOVAHEDIAN AND S. NOBAKHIAN, *Necessary and sufficient conditions for nonsmooth mathematical programs with equilibrium constraints*, *Nonlinear Anal.*, 72 (2010), pp. 2694–2705.
- [27] S. M. ROBINSON, *Generalized equations and their solutions. I: Basic theory*, *Math. Program. Study*, 10 (1979), pp. 128–141.
- [28] R.T. ROCKAFELLAR, *Convex analysis*, Princeton University Press, 1970.
- [29] R.T. ROCKAFELLAR AND R.J-B WETS, *Variational analysis*, Springer, Berlin, 1998.
- [30] H. SCHEEL AND S. SCHOLTES, *Mathematical programs with complementarity constraints: stationarity, optimality, and sensitivity*, *Math. Oper. Res.*, 25 (2000), pp. 1–22.
- [31] M.V. SOLODOV, *A bundle method for a class of bilevel nonsmooth convex minimization problems*, *SIAM J. Optim.*, 18 (2007), pp. 242–259.
- [32] J. S. TREIMAN, *Lagrange multipliers for nonconvex generalized gradients with equality, inequality, and set constraints*, *SIAM J. Cont. Optim.*, 37 (1999), pp. 1313–1329.
- [33] J.J. YE, *Necessary and sufficient optimality conditions for mathematical programs with equilibrium constraints*, *J. Math. Anal. Appl.*, 307 (2005), pp. 350–369.
- [34] J.J. YE, *Optimality conditions for optimization problems with complementarity constraints*, *SIAM J. Optim.*, 9 (1999), pp. 374–387.
- [35] J.J. YE AND X.Y. YE, *Necessary optimality conditions for optimization problems with variational inequality constraints*, *Math. Oper. Res.*, 22 (1997), pp. 977–997.
- [36] J.J. YE AND D.L. ZHU, *Optimality conditions for bilevel programming problems*, *Optimization*, 33 (1995), pp. 9–27 (with Erratum in *Optimization*, 39 (1997), pp. 361–366).
- [37] J.J. YE, D.L. ZHU, AND Q.J. ZHU, *Exact penalization and necessary optimality conditions for generalized bilevel programming problems*, *SIAM J. Optim.*, 7 (1997), pp. 481–507.
- [38] A.B. ZEMKOHO, *Bilevel programming: reformulations, regularity, and stationarity*, Ph.D. thesis, Fakultät für Mathematik und Informatik, TU Bergakademie Freiberg, Freiberg, 2012.