Bilevel programming with convex lower level problems

Joydeep Dutta[1] and Stephan Dempe[2]

[1] Department of Mathematics, Indian Institute of Technology, Kanpur-208106, India jdutta@iitk.ac.in

[2] Department of Mathematics and Informatics, Technical University Bergakademie Freiberg, 09596 Freiberg, Germany dempe@tu-freiberg.de

Summary. In this article we develop certain necessary optimality condition for bilevel programming problems with convex lower-level problem. The results are abstract in nature and depend on an important construction in nonsmooth analysis called the coderivative of a set-valued map.

Key words: Optimistic bilevel programming problem, coderivative of set-valued map, necessary optimality conditions.

1 Introduction

Bilevel programming is among the frontier areas of modern optimization theory. Apart from its importance in application it is also a theoretically challenging field. The first such challenge comes when one wants to write Karush-Kuhn-Tucker type optimality conditions for bilevel programming problems. The major drawback is that most standard constraint qualifications are never satisfied for a bilevel programming problem. Thus it is interesting to devise methods in which one may be able to develop in a natural way constraint qualifications associated with bilevel problems and thus proceed towards obtaining Karush-Kuhn-Tucker type optimality conditions. The recent literature in optimization has seen quiet a few attempts to obtain optimality conditions for bilevel programming problems. See for example Ye and Zhu [27],[28],[29], Ye and Ye [26], Dempe [9],[10], Loridan and Morgan [15], Bard [3], [4],[5] and the references there in. In 1984 J. F. Bard [3] made an attempt to develop optimality conditions for bilevel programming problems though it was later observed to have some error. The recent monograph by Dempe [9] is one helpful source to study optimality conditions for bilevel programming problems.

In this article we consider the special type of bilevel programming which has a convex programming problem as its lower-level problem. Using the recent advances made in the understanding of the solutions sets of variational

S. Dempe and V. Kalashnikov (eds.), *Optimization with Multivalued Mappings*, pp. 51-71

systems (see for example Dontchev and Rockafellar [11],[12] and Levy and Mordukhovich [14]) we will develop some new necessary conditions for bilevel programming problems with convex lower-level problems. In section 2 we begin with the basic formulation of a bilevel programming problem and motivate the type of problems we intend to study in this article. Then we outline the variational tools that are required to represent the necessary optimality conditions. In section 4 we present the main results, i.e the necessary optimality conditions for each of the problem formulations that we have described earlier.

2 Motivation

Consider the following bilevel programming problem (P)

$$\min_x F(x,y) \quad \text{subject to} \quad y \in S(x), x \in X,$$

where $F : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$, $X \subseteq \mathbb{R}^n$ and $S(x)$ is the solution set of the following problem (LLP)

$$\min_y f(x,y) \quad \text{subject to} \quad y \in K(x),$$

where $f : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ and $K(x) \subseteq \mathbb{R}^m$ is a set depending on x. The problem (LLP) is called the lower-level problem and the problem (P) is called the upper-level problem. For simplicity we consider only the case where $X = \mathbb{R}^n$ and where for each x the function $f(x,\cdot)$ is convex and the set $K(x)$ is convex. Thus in our setting the problem (P) will be as follows

$$\min_x F(x,y) \quad \text{subject to} \quad y \in S(x)$$

where $S(x)$ as before denotes the solution set of the lower-level problem (LLP)

$$\min_y f(x,y) \quad \text{subject to} \quad y \in K(x).$$

From now on we will assume that the problem (LLP) always has a solution. However the term min in the upper level problem is slightly ambiguous since one is not sure whether the lower-level problem has an unique solution or not. If $f(x,\cdot)$ is strictly convex in y for each x then $S(x)$ is a singleton for each x or rather S is a single-valued function. If however $f(x,\cdot)$ is only assumed to be convex one cannot always guarantee the single-valuedness of $S(x)$. It is important to observe that the main complication in bilevel programming arises when the lower-level problem does not have a unique solution, i.e. $S(x)$ is not a singleton for some x. Thus for that particular case the objective function of the upper-level problem would look like

$$\bigcup_{y \in S(x)} F(x,y) = F(x, S(x)).$$

The bilevel programming problem (P) is then rather a set-valued optimization problem, see e.g. G.Y. Chen and J. Jahn [7]. To treat this situation within bilevel programming problems at least two approaches have been reported in the literature, namely the *optimistic solution* and the *pessimistic solution.* We will consider here the optimistic solution approach for reasons which will be clear as we progress further. For details on the pessimistic solution approach see for example Dempe [9]. One of the reasons are the less restrictive assumptions needed to guarantee the existence of an optimal solution in the optimistic case. To introduce the optimistic case consider the function

$$\varphi_0(x) = \inf_y \{F(x,y) : y \in S(x)\}.$$

We remark that $\varphi_0(x)$ denotes the infimal function value of the upper level objective over the solution set of the lower-level problem parameterized by x and that we do not demand that it is attained. For each $x \in \mathbb{R}^n$, this function gives the lowest bound for possible objective function values of the upper level objective function on the set of optimal solutions of (LLP). Then, the optimistic bilevel problem reads as

$$\min_x \varphi_0(x). \tag{1}$$

Definition 2.1 *A point $\overline{x}$ is called a (global) optimistic solution of the problem (P) if $\varphi_0(x) \geq \varphi_0(\overline{x})$ for all x.*

An optimal solution of this problem exists whenever the function $\varphi_0(x)$ is lower semicontinuous and some boundedness assumptions are satisfied.

Theorem 2.1 *Consider problem (P) with continuous functions F, f and a continuous point-to-set mapping $K : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$. Then, if gph K is bounded, problem (P) has a solution.*

Here gph $K = \{(x,y) : y \in S(x)\}$ denotes the graph of the mapping K. To guarantee continuity of the point-to-set mapping some regularity condition (as Slater's conditions for all x) is needed. The main reason for this result is that the assumptions imply upper semicontinuity of the point-to-set mapping $S : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ which in turn implies lower semicontinuity of the function $\varphi_0(\cdot)$, see Bank et al. [2].

Let us now consider the problem (P1) given as

$$\min_{x,y} F(x,y) \quad \text{subject to} \quad (x,y) \in \operatorname{gph} S.$$

If local optimal solutions are under consideration it is easy to find examples showing that local optimal solutions of problem (P1) need not to correspond to local optimal solutions of (P). But, for each local optimal solution $\overline{x}$ of (1), some point $(\overline{x}, \overline{y})$ with $\overline{y} \in S(\overline{x})$ is a local optimal solution of (P1).

Proposition 2.1 *Let $\bar{x}$ be a local optimistic solution to the bilevel programming problem (P) whose solution set mapping S is upper-semicontinuous as a set-valued map. Then $(\bar{x}, \bar{y})$ with $\bar{y} \in S(\bar{x})$ and $\varphi_0(\bar{x}) = F(\bar{x}, \bar{y})$ is also a solution of (P1).*

Remark 2.1 We note that we have used the implicit assumption in the proposition that the lower-level problem (LLP) has an optimal solution for $x = \bar{x}$. Let us recall that we have already made this assumption in the beginning of this section.

Proof. Let $\bar{x}$ be a local optimistic solution to (P) and assume that there exists $\bar{y}$ with the properties as formulated in the statement. Then we first have $\bar{y} \in S(\bar{x})$ and

$$F(\bar{x}, \bar{y}) \leq F(\bar{x}, y), \quad \forall y \in S(\bar{x}).$$

By assumption $\varphi_0(\bar{x}) = F(\bar{x}, \bar{y})$. Further we also have

$$\varphi_0(\bar{x}) \leq \varphi_0(x), \quad \forall x \in \mathbb{R}^n \tag{2}$$

sufficiently close to $\bar{x}$. By definition of $\varphi_0(x)$ one has $\varphi_0(x) \leq F(x, y)$ for all $y \in S(x)$. Using (2) we immediately have

$$F(\bar{x}, \bar{y}) = \varphi_0(\bar{x}) \leq \varphi_0(x) \leq F(x, y),\ \forall y \in S(x) \text{ and } x \text{ sufficiently close to } \bar{x}.$$

Let V be an open neighborhood of $S(\bar{x})$, i.e. $S(\bar{x}) \subset V$. Since S is upper-semicontinuous as a set-valued map we have that there exists an open neighborhood U of $\bar{x}$ such that for all $x \in U$ one has $S(x) \subset V$(For a definition of upper-semicontinuous set-valued map see for example Berge [6]). Thus we can find a $\delta > 0$ such that $B_\delta(\bar{x}) \subset V$ and for all $x \in B_\delta(\bar{x})$ we have $\varphi(x) \geq \varphi(\bar{x})$. Here $B_\delta(\bar{x})$ denotes a ball centered at $\bar{x}$ and of radius δ. Thus arguing in a similar manner as before one has

$$F(\bar{x}, \bar{y}) \leq F(x, y) \quad \forall y \in S(x) \quad \text{and} \quad x \in B_\delta(\bar{x}).$$

However for all $x \in B_\delta(\bar{x})$ we have $S(x) \subset V$. This shows that

$$F(\bar{x}, \bar{y}) \leq F(x, y) \quad \forall (x, y) \in (B_\delta(\bar{x}) \times V) \cap \operatorname{gph} S.$$

Thus $(\bar{x}, \bar{y})$ is a local optimal solution for (P1). □

Remark 2.2 It is important to note that assumption of upper-semicontinuity on the solution set-mapping is not a strong one since it can arise under natural assumptions. Assume that feasible set $K(x)$ of the lower-level problem (LLP) is described by convex inequality constraints, i.e.

$$K(x) = \{y \in \mathbb{R}^m : g_i(x, y) \leq 0, i = 1, \ldots, p\},$$

where for each x the function $g(x, \cdot)$ is convex in y. Assume now that the set $\{(x, y) \in \mathbb{R}^n \times \mathbb{R}^m : g_i(x, y) \leq 0, i = 1, \ldots, p\}$ is non-empty and compact and the lower-level problem (LLP) satisfies the Slater's constraint qualification then we can conclude that the solution set mapping S is upper-semicontinuous as a set-valued map, cf. Bank et al. [2].

Let us note that if we now consider a global optimistic solution then there is no necessity for any additional assumption on the solution set mapping. Thus we have the following proposition.

Proposition 2.2 *Let $\bar{x}$ be a global optimistic solution to the bilevel programming problem (P). Then $(\bar{x}, \bar{y})$ with $\overline{y} \in S(\overline{x})$ and $\varphi_0(\overline{x}) = F(\overline{x}, \overline{y})$ is also a global solution of (P1).*

Proof. Let $\overline{x}$ be a global optimistic solution to (P) and assume that there exists $\overline{y}$ with the properties as formulated in the statement. Then we first have $\bar{y} \in S(\bar{x})$ and

$$F(\bar{x}, \bar{y}) \leq F(\bar{x}, y), \quad \forall y \in S(\bar{x}).$$

By assumption $\varphi_0(\bar{x}) = F(\bar{x}, \bar{y})$. Further we also have

$$\varphi_0(\bar{x}) \leq \varphi_0(x), \quad \forall x \in \mathbb{R}^n. \tag{3}$$

By definition of $\varphi_0(x)$ one has $\varphi_0(x) \leq F(x, y)$ for all $y \in S(x)$. Using (3) we immediately have

$$F(\bar{x}, \bar{y}) = \varphi_0(\bar{x}) \leq \varphi_0(x) \leq F(x, y), \quad \forall y \in S(x).$$

Hence the result. □

The opposite implication is also valid for global optima.

Proposition 2.3 *Let $(\overline{x}, \overline{y})$ be a global optimal solution of problem (P1). Then, $\overline{x}$ is a global optimal solution of problem (P).*

Proof. Assume that $\overline{x}$ is not a global optimal solution of problem (1) then there is $\widetilde{x}$ with $\varphi_0(\widetilde{x}) < \varphi_0(\overline{x})$ and, by definition of the function $\varphi_0(\cdot)$ there is $\widetilde{y} \in S(\widetilde{x})$ with $\varphi_0(\widetilde{x}) \leq F(\widetilde{x}, \widetilde{y}) < \varphi_0(\overline{x})$. Now, $\overline{y} \in S(\overline{x})$ and, hence,

$$F(\overline{x}, \overline{y}) = \varphi_0(\overline{x}) > F(\widetilde{x}, \widetilde{y}).$$

Then $F(\widetilde{x}, \widetilde{y}) < F(\overline{x}, \overline{y})$ which contradicts optimality of $(\overline{x}, \overline{y})$. □

The last two propositions enable us to reformulate the bilevel problem in its optimistic version to the problem (P1). Note that this excludes the case when the function φ_0 is determined as the infimal objective function value of the lower-level problem (LLP) (which is then not assumed to have an optimal solution), implying that the function φ_0 may have a minimum even in the case when the problem (P1) has no solution.

We have already stated that we will restrict ourselves to the case where the lower-level problem is a convex minimization problem. Let for the moment the set $K(x)$ in (LLP) be expressed in terms of convex inequalities:

$$K(x) = \{y : g_i(x, y) \leq 0, i = 1, \ldots, p\},$$

where $g_i(x, \cdot)$ are convex in y for each x and are sufficiently smooth i.e of class C^2. Finding the minimum of a regular convex problem is equivalent to solving the Karush-Kuhn-Tucker type conditions associated with the problem. Thus a bilevel programming problem can be posed as single level problem with the lower-level problem being replaced with its Karush-Kuhn-Tucker system which now become additional constraints to the problem (P1). Thus the problem (P1) can be reformulated as

$$\begin{aligned} & \min_{x,y,\lambda} F(x, y) \\ \text{subject to } \; & \nabla_y f(x, y) + \sum_{i=1}^{p} \lambda_i \nabla_y g_i(x, y) = 0 \\ & \lambda_i g_i(x, y) = 0 \quad i = 1, \ldots, p. \\ & g_i(x, y) \leq 0, \;\; \lambda_i \geq 0 \quad i = 1, \ldots, p. \end{aligned} \tag{4}$$

This is the so-called Karush-Kuhn-Tucker (KKT) formulation of a bilevel programming problem with a convex lower-level problem. Problem (4) is a special kind of the so-called Mathematical Program with Equilibrium Constraints (MPEC). It is well-known that many standard constraint qualifications like the Mangasarian-Fromowitz constraint qualification and the Abadie constraint qualification fail due to the presence of the complementary slackness condition of the lower-level problem which is now posed as an equality constraint. The reader is referred to the paper Ye [23] where possible regularity conditions are identified. The challenge therefore is to devise natural qualification conditions which can lead to KKT type optimality conditions for a bilevel programming problem. Here we suggest one new approach through which this may be possible. Similar investigations have been done in Ye [23] under the assumptions that the problem functions are either Gâteaux differentiable or locally Lipschitz continuous using the Michel-Penot subdifferential.

It should be mentioned that the KKT reformulation is equivalent to the problem (P1) only in the case when the lower-level problem is a convex regular one and global optimal solutions of the upper level problem are investigated (cf. Propositions 2.2 and 2.3). Without convexity, problem (4) has a larger feasible set than (P1) and an optimal solution of (4) need not to correspond to a feasible solution of problem (P1). Even more, an optimal solution of problem (P1) need also not be an optimal solution of (4), see Mirrlees [16].

What concerns optimality conditions the main difficulty in using the reformulation (4) of the bilevel programming problem is the addition of new variables. If these Lagrange multipliers of the lower-level problem are not uniquely determined, the optimality conditions of the MPEC depend on the selection of the multiplier but the conditions for the bilevel problem must not. This can easily been seen e.g. in the case when the lower-level problem is a convex one for which the Mangasarian-Fromowitz constraint qualification together with the strong sufficient optimality condition of second order and the constant rank constraint qualification are satisfied at a point $(\overline{x}, \overline{y})$. Then, the optimal solution of the lower-level problem is strongly stable in the sense of Kojima [13], Lipschitz continuous and directionally differentiable, see Ralph and Dempe [20] and the bilevel programming problem can be reformulated as

$$\min\{F(x, y(x)) : x \in \mathbb{R}^n\}.$$

Necessary optimality conditions for this problem reduce to nonexistence of directions of descent for the function $x \mapsto F(x, y(x))$, cf. Dempe [10]. If this problem is reformulated as (4) and a Lagrange multiplier is fixed it is possible that there is no direction of decent in the problem (4). But what we have done is to compute the directional derivative of the function $x \mapsto F(x, y(x))$ only in directions which correspond to the selected Lagrange multiplier, i.e. directions for which a certain linear optimization problem has a solution, see Ralph and Dempe [20]. But there is no need that the directional derivative of the function $F(x, y(x))$ into other directions (corresponding to other Lagrange multipliers) does not give a descent.

With other words, if optimality conditions for an MPEC are investigated, a feasible solution of this problem is fixed and optimality conditions are derived as in Pang and Fukushima [19], Scheel and Scholtes [22]. Considering the optimality conditions in primal space (i.e. formulating them as nonexistence of descent directions in the contingent cone) we see some combinatorial structure since the contingent cone is not convex. This approach has been applied to the KKT reformulation of a bilevel programming problem in Ye and Ye [26]. But to obtain a more useful condition for selecting a locally optimal solution we have to investigate the resulting systems for all Lagrange multipliers of the lower-level problem, or at least for all the vertices of the set of Lagrange multipliers, if some condition as the constant rank constraint qualification in the differentiable case is satisfied. Hence, this approach needs to be complemented by e.g. a method for an efficient computation of all Lagrange multipliers.

Hence we believe that other approaches are more promising. These are on the one hand approaches using the normal cone (or the contingent cone) to the graph of the solution set mapping of the lower-level problem and on the other hand approaches using the reformulation of the bilevel programming problem using the optimal value function of the lower-level problem. The latter approach has been used e.g. in the papers Babahadda and Gadhi [1], Ye [23].

Here we investigate the possibility to derive necessary optimality conditions using the normal cone.

3 Basic tools

Let $(\bar{x}, \bar{y})$ be a local (or global) solution of (P1) and let us assume that F is smooth. Then one has

$$0 \in \nabla F(\bar{x}, \bar{y}) + N_{\mathrm{gph} S}(\bar{x}, \bar{y}).$$

In the above expression $N_{\mathrm{gph}\, S}(\bar{x}, \bar{y})$ denotes the *Mordukhovich normal cone* or the *basic normal cone* to the graph of the set-valued map S at $(\bar{x}, \bar{y})$. For more details on Mordukhovich normal cone and the derivation of the above necessary optimality condition see for example Mordukhovich [17] and Rockafellar and Wets [21]. It is moreover important that the Mordukhovich normal cone is in general a closed and non-convex object. The basic normal cone to a convex set coincides with the usual normal cone of convex analysis. In order to obtain a KKT type optimality condition our main task is now to compute the basic normal cone to the graph of the solution set mapping at the point $(\bar{x}, \bar{y})$. Thus the qualification conditions that are required to compute the normal cone are indeed the natural qualification conditions for the bilevel programming problem. However let us note that it is in fact a formidable task to compute the normal cone to the graph of the solution set mapping. This is mainly due to the fact that even if the lower-level problem is convex the graph $\mathrm{gph}\, S$ of the solution set mapping S need not be convex. The following simple example demonstrates this fact.

Example 3.1 Let the lower-level problem be given as

$$S(x) = \operatorname*{argmin}_{y} \{f(x, y) = -xy : 0 \leq y \leq 1\}.$$

Observe here that $K(x) = [0, 1]$ for all $x \in \mathbb{R}$. Observe that the problem is a convex problem in y. Also note that the solution set mapping S in this particular case is given as

$$S(x) = \begin{cases} \{0\} & : x < 0 \\ [0, 1] & : x = 0 \\ \{1\} & : x > 0. \end{cases}$$

It is now simple to observe that the $\mathrm{gph}\, S$ is a non-convex set. $\triangle$

Professor Rockafellar suggested that an interesting approach to bilevel programming may be obtained by having a minimax problem or rather a primal-dual problem in the lower-level instead of just a convex minimization problem. This can in fact be motivated from the KKT representation of a bilevel programming problem with convex lower-level problems. Observe that the KKT problem brings in an additional variable $\lambda \in \mathbb{R}^p_+$ which is actually the Lagrange multiplier associated with the problem as well as the dual variable associated with the Lagrangian dual of the convex lower-level problem. Hence both the primal variable y and the dual variable λ of the convex lower-level problem are present in the KKT formulation of the bilevel programming problem with convex lower-level problems. Thus one may as well define a lower-level problem which has both the primal and dual variable and that naturally suggests us to consider the lower-level problem as a minmax problem. Thus we can have a new formulation of the bilevel programming problem (P2) with a minimax lower-level problem as follows

$$\min_{(x,y,\lambda)} F(x,y,\lambda) \quad \text{subject to} \quad (y,\lambda) \in S(x),$$

where $F : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p \to \mathbb{R}$ and the set-valued map $S : \mathbb{R}^n \rightrightarrows \mathbb{R}^m \times \mathbb{R}^p$ is a solution set of the following problem (LLP2)

$$\text{minimaximize}\, L(x,y,\lambda) \quad \text{subject to} \quad (y,\lambda) \in Y \times W,$$

where $Y \subset \mathbb{R}^m$ and $W \subset \mathbb{R}^p$ are non-empty convex sets and $L(x,y,\lambda)$ is convex with respect to y for each $(x,\lambda) \in \mathbb{R}^n \times W$ and is concave in λ for each $(x,y) \in \mathbb{R}^m \times Y$. Thus we can write

$$S(x) = \{(y,\lambda) : (y,\lambda) \quad \text{solves} \quad (LLP2)\}.$$

By a solution $(y,\lambda) \in S(x)$ we mean

$$y \in \operatorname*{argmin}_{y \in Y} L(x,y,\lambda) \text{ and } \lambda \in \operatorname*{argmax}_{\lambda \in W} L(x,y,\lambda).$$

Let us end this section by defining the nonsmooth tools that would be required for the proofs of the optimality conditions. We first begin with the definition of the normal cone to a set C at a given point in C. Let C be a non-empty subset of $\mathbb{R}^n$ and let $\bar{x} \in C$. A vector v is called a *regular normal* to C at $\bar{x}$ if

$$\langle v, x - \bar{x} \rangle \leq o(\|x - \bar{x}\|),$$

where $\dfrac{o(\|x-\bar{x}\|)}{\|x-\bar{x}\|} \to 0$ as $\|x - \bar{x}\| \to 0$. The set of all regular normals form a convex cone denoted by $\hat{N}_C(\bar{x})$. This is also known as the *Fréchet normal cone* in the literature.

A vector $v \in \mathbb{R}^n$ is said to be a normal or a basic to C at $\bar{x}$ if there exist a sequence $\{v_k\}$, with $v_k \to v$ and a sequence $\{x_k\}$, $x_k \in C$ with $x_k \to \overline{x}$ and $v_k \in \hat{N}_C(x_k)$. The set of all normals forms a closed (but not necessarily convex) cone denoted as $N_C(\bar{x})$. The basic normal cone has also been referred

to as the Mordukhovich normal cone in the literature. For more details on the basic normal cone in the finite dimensional setting see for example Mordukhovich [17] or Rockafellar and Wets [21]. It is important to note that if the interior of C is nonempty and $\bar{x} \in \text{int}C$ then $N_C(\bar{x}) = \{0\}$.
Let $S : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ be a set-valued map and let $(x, y) \in \operatorname{gph} S$. Then the coderivative at $(\bar{x}, \bar{y})$ is a set-valued map $D^*S(\bar{x}|\bar{y}) : \mathbb{R}^m \rightrightarrows \mathbb{R}^n$ given as

$$D^*S(\bar{x}|\bar{y})(w) = \{v \in \mathbb{R}^n : (v, -w) \in N_{\operatorname{gph} S}(\bar{x}, \bar{y})\}.$$

For more details on the properties of the coderivative see for example Mordukhovich [17] and Rockafellar and Wets [21]. Further given a function $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ and a point $\bar{x}$ where f is finite the *subdifferential* or the *basic subdifferential* at $\bar{x}$ is given as

$$\partial f(\bar{x}) = \{\xi \in \mathbb{R}^m : (\xi, -1) \in N_{\operatorname{epi} f}(\bar{x}, f(\bar{x}))\},$$

where $\operatorname{epi} f$ denotes the epigraph of the function f. The asymptotic subdifferential of f at $\bar{x}$ is given as

$$\partial^\infty f(\bar{x}) = \{\xi \in \mathbb{R}^m : (\xi, 0) \in N_{\operatorname{epi} f}(\bar{x}, f(\bar{x}))\}.$$

We will now present Theorem 2.1 in Levy and Mordukhovich [14] in the form of two lemmas whose application would lead to the necessary optimality conditions.

Lemma 3.1 *Consider the set-valued map $S : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ given as follows*

$$S(x) = \{y \in \mathbb{R}^m : 0 \in G(x, y) + M(x, y)\}, \tag{5}$$

where $G : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^d$ is a smooth vector-valued function and $M : \mathbb{R}^n \times \mathbb{R}^m \rightrightarrows \mathbb{R}^d$ is a set-valued map with closed graph. Let $(\bar{x}, \bar{y}) \in$ gph S and let the following qualification condition hold

$$v \in \mathbb{R}^d \quad \textit{with} \quad 0 \in \nabla G(\bar{x}, \bar{y})^T v + D^*M((\bar{x}, \bar{y})| - G(\bar{x}, \bar{y}))(v) \Longrightarrow v = 0.$$

Then one has

$$\begin{aligned} D^*S(\bar{x}|\bar{y})(y^*) \subseteq \{x^* : \exists v^* \in \mathbb{R}^d, (x^*, -y^*) \in \nabla G(\bar{x}, \bar{y})^T v^* + \\ D^*M((\bar{x}, \bar{y})| - G(\bar{x}, \bar{y}))(v^*)\}. \end{aligned}$$

Lemma 3.2 *Consider the set-valued map $S : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ given in formula (5) where $G : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^d$ is a smooth vector-valued function and $M : \mathbb{R}^n \times \mathbb{R}^m \rightrightarrows \mathbb{R}^d$ is a set-valued map with closed graph. Further assume that M only depends on y i.e $M(x, y) = M(y)$. Assume that the matrix $\nabla_x G(\bar{x}, \bar{y})$ has full rank. Then one has*

$$\begin{aligned} D^*S(\bar{x}|\bar{y})(y^*) = \{x^* : \exists v^* \in \mathbb{R}^d, x^* = \nabla_x G(\bar{x}, \bar{y}) v^*, \\ -y^* = \nabla_y G(\bar{x}, \bar{y})^T v^* + D^*M(\bar{y}| - G(\bar{x}, \bar{y}))(v^*)\}. \end{aligned}$$

4 Main Results

In this section we shall present necessary optimality conditions for the two classes of bilevel programming problems which we have discussed in the previous sections. First we shall derive necessary optimality conditions for the problem format defined by (1) in which the lower-level problem is a convex minimization problem. Then we shall consider the case when $K(x) = K$ for all x and then move on to the case where the lower-level problem is given in the form a primal-dual problem. Then we will present a more refined optimality condition using the second-order subdifferential of the indicator function which would appear to be a very novel feature. Before we begin let us define the following set-valued map which can also be called as the normal cone map

$$N_K(x,y) = \begin{cases} N_{K(x)}(y) & : y \in K(x) \\ \emptyset & : y \notin K(x) \end{cases}$$

Theorem 4.1 *Consider the problem (P1) given as*

$$\min_{x,y} F(x,y) \quad \textit{subject to} \quad (x,y) \in gph\ S,$$

where $F : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ is a smooth function and $S : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ is a set-valued map denoting the solution set of the problem (LLP) i.e.

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

where $f(x,\cdot)$ is a smooth convex function in y for each x and $K(x)$ is a closed convex set for each x. Let $(\bar{x},\bar{y})$ be a local (or global) solution of (P1). Further assume that $\nabla_y f : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^m$ is continuously differentiable. Set $\bar{p} = \nabla_y f(\bar{x},\bar{y})$. Assume also that the following qualification condition holds at $(\bar{x},\bar{y})$:

$$v \in \mathbb{R}^m \quad \textit{with} \quad 0 \in \nabla(\nabla_y f(\bar{x},\bar{y}))^T v + D^* N_K((\bar{x},\bar{y})| - \bar{p})(v) \Longrightarrow v = 0.$$

Then there exists $v^* \in \mathbb{R}^m$ such that

$$0 \in \nabla F(\bar{x},\bar{y}) + \nabla(\nabla_y f(\bar{x},\bar{y}))^T v^* + D^* N_K((\bar{x},\bar{y})| - \bar{p})(v^*).$$

Proof. They key to the proof of this result is Lemma 3.1. To begin with note that since (LLP) is a convex minimization problem in y for each given x we can write $S(x)$ equivalently as

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

It is not much difficult to show that the normal cone map has a closed graph. Since $(\bar{x},\bar{y})$ is a local (or global) solution of the problem (P1) then we have

$$-\nabla F(\bar{x},\bar{y}) \in N_{\operatorname{gph} S}(\bar{x},\bar{y}).$$

Now by using the definition of the coderivative and then applying Lemma 3.1 we have that there exists $v^* \in \mathbb{R}^m$ such that

$$-\nabla F(\bar{x},\bar{y}) \in \nabla(\nabla_y f(\bar{x},\bar{y}))^T v^* + D^* N_K((\bar{x},\bar{y})| - \bar{p})(v^*).$$

This proves the result □

We will now apply the above result to bilevel programming

Corollary 4.1 *Let us consider the bilevel programming problem (P)*

$$\min_x F(x,y) \quad \textit{subject to} \quad y \in S(x),$$

where $F : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ *is a smooth function and* $S(x)$ *is the solution set of the following problem (LLP)*

$$\min_y f(x,y) \quad \textit{subject to} \quad y \in K(x),$$

where $f : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ *is a smooth strictly convex function in* y *for each* x *and* $K(x)$ *is a compact convex set for each* x*. Let* $(\bar{x},\bar{y})$ *be a local solution of (P). Further assume that* $\nabla_y f : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^m$ *is continuously differentiable. Set* $\bar{p} = \nabla_y f(\bar{x},\bar{y})$*. Assume further that the following qualification condition holds at* $(\bar{x},\bar{y})$ *:*

$$v \in \mathbb{R}^m \quad \textit{with} \quad 0 \in \nabla(\nabla_y f(\bar{x},\bar{y}))^T v + D^* N_K((\bar{x},\bar{y})| - \bar{p})(v) \Longrightarrow v = 0.$$

Then there exists $v^* \in \mathbb{R}^m$ *such that*

$$0 \in \nabla F(\bar{x},\bar{y}) + \nabla(\nabla_y f(\bar{x},\bar{y}))^T v^* + D^* N_K((\bar{x},\bar{y})| - \bar{p})(v^*).$$

Proof. By the hypothesis of the theorem for each x the problem (LLP) has a unique solution. Hence the solution of the problem (P) is also a solution of (P1). The rest of the proof follows as in Theorem 4.1. □

Corollary 4.2 *Let us consider the bilevel programming problem (P)*

$$\min_x F(x,y) \quad \textit{subject to} \quad y \in S(x),$$

where $F : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ *is a smooth function and* $S(x)$ *is the solution set of the following problem (LLP)*

$$\min_y f(x,y) \quad \textit{subject to} \quad y \in K(x),$$

where $f : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ *is a smooth convex function in* y *for each* x *and* $K(x)$ *is a convex set for each* x*. Further assume that the solution set mapping* S *is upper-semicontinuous as a set-valued map. Let* $\overline{x}$ *be a local optimistic solution of (P) and assume that* $\overline{y} \in S(\overline{x})$ *with* $F(\overline{x},\overline{y}) = \varphi_0(\overline{x})$ *exists. Further assume* *that* $\nabla_y f : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^m$ *is continuously differentiable. Set* $\bar{p} = \nabla_y f(\bar{x}, \bar{y})$. *Assume further that the following qualification condition hold at* $(\bar{x}, \bar{y})$ *:*

$$v \in \mathbb{R}^m \quad \textit{with} \quad 0 \in \nabla(\nabla_y f(\bar{x}, \bar{y}))^T v + D^* N_K((\bar{x}, \bar{y})| - \bar{p})(v) \Longrightarrow v = 0.$$

Then there exists $v^* \in \mathbb{R}^m$ *such that*

$$0 \in \nabla F(\bar{x}, \bar{y}) + \nabla(\nabla_y f(\bar{x}, \bar{y}))^T v^* + D^* N_K((\bar{x}, \bar{y})| - \bar{p})(v^*).$$

Proof. Our assumptions imply that $(\overline{x}, \overline{y})$ is a local solution of the problem (P1) due to Proposition 2.1. The rest of the proof follows as in Theorem 4.1. □

Remark 4.1 An interesting feature in the optimality conditions presented in the above theorem is the presence of second-order partial derivatives in the expression of first order optimality conditions. This is essentially due to presence of the matrix $\nabla(\nabla_y f(\bar{x}, \bar{y}))$. The presence of second-order partial derivatives in the first conditions is a hallmark of bilevel programming. Further note that one can have analogous results for global optimistic solution using Proposition 2.2 and without any additional assumption on the nature of the solution set mapping S.

We will now turn to the case when $K(x) = K$ for all $x \in \mathbb{R}^n$. In such a case we have a much simplified qualification condition which amounts to checking whether a matrix is of full rank.

Theorem 4.2 *Consider the problem (P1) given as*

$$\min_{x,y} F(x, y) \quad \textit{subject to} \quad (x, y) \in \operatorname{gph} S,$$

where $F : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ *is a smooth function and* $S : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ *is a set-valued map denoting the solution set of the problem (LLP) i.e.*

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

where $f(x, \cdot)$ *is a smooth convex function in* y *for each* x *and* $K(x) = K$ *for all* x *where* K *is a fixed closed and convex set. Let us also assume that the function* f *is twice continuously differentiable. Let* $(\bar{x}, \bar{y}) \in \operatorname{gph} S$ *be a solution of problem (P1). Set* $\bar{p} = \nabla_y f(\bar{x}, \bar{y})$. *Further assume that the matrix* $\nabla_x(\nabla_y f(\bar{x}, \bar{y})) = \nabla^2_{xy} f(\bar{x}, \bar{y})$ *has full rank, i.e.*

$$\textit{rank}\left(\nabla^2_{xy} f(\bar{x}, \bar{y})\right) = m.$$

Then there exists $v^* \in \mathbb{R}^m$ *such that the following conditions hold*

i) $0 = \nabla_x F(\bar{x}, \bar{y}) + \nabla^2_{xy} f(\bar{x}, \bar{y}) v^*$
ii) $0 \in \nabla_y F(\bar{x}, \bar{y}) + \nabla^2_{yy} f(\bar{x}, \bar{y}) v^* + D^* N_K(\bar{y}| - \bar{p})(v^*)$.

Proof. In this particular case when $K(x) = K$ then one can write $N_K(x, y) = N_K(y)$. Further the solution set mapping S can also be equivalently written as

$$S(x) = \{y \in \mathbb{R}^m : 0 \in \nabla_y f(x, y) + N_K(y)\}.$$

Since $(\bar{x}, \bar{y})$ solves (P1) we have

$$-\nabla F(\bar{x}, \bar{y}) \in N_{\mathrm{gph}\, S}(\bar{x}, \bar{y}).$$

This shows that

$$-(\nabla_x F(\bar{x}, \bar{y}), \nabla_y F(\bar{x}, \bar{y})) \in N_{\mathrm{gph}\, S}(\bar{x}, \bar{y}).$$

Hence by definition of the coderivative we have

$$-\nabla_x F(\bar{x}, \bar{y}) \in D^* S(\bar{x}|\bar{y})(\nabla_y F(\bar{x}, \bar{y})).$$

Now by using Lemma 3.2 we see that there exists $v^* \in \mathbb{R}^m$ such that

$$-\nabla_x F(\bar{x}, \bar{y}) = \nabla^2_{xy} f(\bar{x}, \bar{y}) v^*$$

and

$$-\nabla_y F(\bar{x}, \bar{y}) \in \nabla^2_{yy} f(\bar{x}, \bar{y})^T v^* + D^* N_K(\bar{y}| - \bar{p})(v^*).$$

Hence the result. □

Remark 4.2 The qualification condition that we have used in the above theorem is called the *ample parametrization condition* in Dontchev and Rockafellar [11]. However in Dontchev and Rockafellar [11] the proto-derivative of the solution set mapping S is computed. The proto-derivative is the tangent cone to the graph of S at $(\bar{x}, \bar{y})$. Thus the approach due to Dontchev and Rockafellar [11] can be used in the dual setting given in terms of the tangent cone. However as we have noted the approach through coderivatives is essential in surpassing the computation (a difficult one that too) that is required to compute the normal cone to the graph of S at $(\bar{x}, \bar{y})$. Thus the results in Levy and Mordukhovich [14] will play a very fundamental role in the study of mathematical programming with equilibrium constraints (MPEC) and also bilevel programming with convex lower-level problems.

It is now easy to observe that the above theorem can be used to deduce optimality conditions for a bilevel programming problem with a convex lower-level problem with $K(x) = K$ for all $x \in \mathbb{R}^n$ if the lower-level problem has a unique solution or we consider an optimistic solution of the bilevel programming problem. However we are not going to explicitly state the results here since this can be done as in the corollaries following Theorem 4.1.

One of the main drawback of the optimality conditions derived above for problem (P) and (P1) is the presence of the coderivative of the normal cone mapping. Thus the optimality conditions are more abstract in nature. The computation of the coderivative of the normal cone map seems to be very difficult. However by using an approach due to Outrata [18] by using some different qualification condition we can derive an optimality condition in which the explicit presence of the coderivative of the normal cone map is not there though as we will see that it will be implicity present. We now present the following result.

Theorem 4.3 *Consider the problem (P1) given as*

$$\min_{x,y} F(x,y) \quad \textit{subject to} \quad (x,y) \in \operatorname{gph} S,$$

where $F : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ *is a smooth function and* $S : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ *is a set-valued map denoting the solution set of the problem (LLP) i.e.*

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

where $f(x,\cdot)$ *is a smooth convex function in* y *for each* x *and* $K(x) = K$ *for all* x *where* K *is a fixed closed and convex set. Let us also assume that the function* f *is twice continuously differentiable. Let* $(\bar{x},\bar{y}) \in \operatorname{gph} S$ *be a local solution of problem (P1). Further assume that the following qualification condition holds at* $(\bar{x},\bar{y})$ *:*

$$(w,z) \in N_{\operatorname{gph} N_k}(\bar{y}, -\nabla_y f(\bar{x},\bar{y})) \quad \textit{with}$$
$$(\nabla^2_{xy} f(\bar{x},\bar{y}))^T z = 0, \quad w - (\nabla^2_{yy} f(\bar{x},\bar{y}))^T z = 0 \Longrightarrow w = 0, z = 0.$$

Then there exists a pair $(\bar{w},\bar{z}) \in N_{\operatorname{gph} N_k}(\bar{y}, -\nabla_y f(\bar{x},\bar{y}))$ *such that*

i) $\nabla_x F(\bar{x},\bar{y}) = (\nabla^2_{xy} f(\bar{x},\bar{y}))^T \bar{z}$.
ii) $-\nabla_y F(\bar{x},\bar{y}) = \bar{w} - (\nabla^2_{yy} f(\bar{x},\bar{y}))^T \bar{z}$.

Proof. Observe that according the hypothesis of the theorem the problem (P1) is equivalent to the following problem (P4)

$$\min_{x,y} F(x,y) \quad \text{subject to} \quad 0 \in \nabla_y f(x,y) + N_K(y).$$

Now by applying Theorem 3.1 in Outrata [18] we reach our desired conclusion. □

Observe that the qualification condition in Theorem 4.3 guarantees that $(\nabla^2_{xy} f(\bar{x},\bar{y}))^T$ has full rank which is similar to the qualification condition appearing in Theorem 4.2. However there is also an extra qualification condition since we now have two Lagrange multipliers instead of one. Though the coderivative does not appear explicitly in the representation of the optimality condition but the condition $(\bar{w}, \bar{z}) \in N_{\mathrm{gph}\, N_k}(\bar{y}, -\nabla_y f(\bar{x}, \bar{y}))$ tells us that

$$\bar{w} \in D^*(\bar{y}| - \nabla_y f(\bar{x}, \bar{y}))(-\bar{z}).$$

Thus the conditions obtained in Theorem 4.2 are same as that of Theorem 4.3. However the approach due to Outrata [18] seems to have an additional advantage. This apparent advantage is that we can use Outrata's approach even when in the problem x is lying in a proper closed set X of $\mathbb{R}^m$. In such a situation the problem (P1) gets slightly modified and looks as follows

$$\min_{x,y} F(x,y), \quad \text{subject to} \quad (x,y) \in \mathrm{gph}\, S \quad x \in X.$$

Also note that since $N_K(y) = \emptyset$, when $y \notin K$ it is clear that $y \in K$ is implied by $0 \in \nabla_y f(x,y) + N_K(y)$. Thus $S(x)$ can also be written as

$$\begin{aligned} S(x) &= \{y \in \mathbb{R}^m : 0 \in \nabla_y f(x,y) + N_K(y)\} \\ &= \{y \in K : 0 \in \nabla_y f(x,y) + N_K(y)\}. \end{aligned}$$

Hence $S(x) \subset K$. So when we write the expression for $S(x)$ there is no need to explicitly write that $y \in K$. Thus when $x \in X$ the modified version of the problem (P1) is equivalent to

$$\min_{x,y} F(x,y) \quad \text{subject to} \quad 0 \in \nabla_y f(x,y) + N_K(y), \quad (x,y) \in X \times \mathbb{R}^m$$

Hence, if $(\bar{x}, \bar{y})$ is a solution of the modified (P1) then it also solves the above problem. Thus in this scenario the qualification condition in Theorem 3.1 in Outrata reduces to the following. Consider $(w, z) \in N_{\mathrm{gph}\, N_k}(\bar{y}, -\nabla_y f(\bar{x}, \bar{y}))$. Then

$$((\nabla^2_{xy} f(\bar{x}, \bar{y}))^T z, w - (\nabla^2_{yy} f(\bar{x}, \bar{y}))^T z) \in N_{X \times \mathbb{R}^m}(\bar{x}, \bar{y}) \Longrightarrow w = 0, z = 0.$$

Then by applying Theorem 3.1 in Outrata [18] we arrive at the conclusion that there exists a pair $(\bar{w}, \bar{z}) \in N_{\mathrm{gph}\, N_k}(\bar{y}, -\nabla_y f(\bar{x}, \bar{y}))$ and $(\gamma, 0) \in N_{X \times \mathbb{R}^m}(\bar{x}, \bar{y})$ such that

i) $-\nabla_x F(\bar{x}, \bar{y}) = -(\nabla^2_{xy} f(\bar{x}, \bar{y}))^T \bar{z} + \gamma$.
ii) $-\nabla_y F(\bar{x}, \bar{y}) = \bar{w} - (\nabla^2_{yy} f(\bar{x}, \bar{y}))^T \bar{z}$.

Note that $N_{X \times \mathbb{R}^m}(\bar{x}, \bar{y}) = N_X(\bar{x}) \times N_{\mathbb{R}^m}(\bar{y})$ (see for example Rockafellar and Wets [21]). And since $N_{\mathbb{R}^m}(\bar{y}) = \{0\}$, it is clear that

$$N_{X \times \mathbb{R}^m}(\bar{x}, \bar{y}) = \{(\gamma, 0) : \gamma \in N_X(\bar{x})\}.$$

Further if $X = \mathbb{R}^n$ which is the case in Theorem 4.3 one has $N_{\mathbb{R}^n \times \mathbb{R}^m}(\bar{x}, \bar{y}) = \{(0,0)\}$.
Let us now turn our attention of how to calculate the normal cone to the

graph of the normal cone mapping associated with a given set K in the lower-level problem. However if K has some special form then one can have an explicit expression for $N_{\mathrm{gph}\, N_K}(\bar{y}, \bar{z})$. For example if $K = \mathbb{R}^m_+$ then such an explicit expression for $N_{\mathrm{gph}\, N_K}(\bar{y}, \bar{z})$ is given by Proposition 3.7 in Ye [24]. The result in Proposition 3.7 in Ye [24] depends on Proposition 2.7 in Ye [25]. In Proposition 2.7 of Ye [25] the normal cone to the graph of the normal cone mapping $N_{\mathbb{R}^m_+}$ is calculated. Let $C \subset \mathbb{R}^n$ be a closed set. Then $v \in \mathbb{R}^n$ is said to be proximal normal to C at $\bar{x} \in C$ if there exists $\sigma > 0$ such that

$$\langle v, x - \bar{x} \rangle \leq \sigma \|x - \bar{x}\|^2$$

The set of all proximal normals forms a cone called the proximal normal cone which is denoted by $N^P_C(\bar{x})$. It is also important to note that if C is a closed set then a normal vector can be realized as a limit of proximal normal vectors. More precisely if C is closed and $v \in N_C(\bar{x})$ then there exist sequences $v_k \to v$ and $x_k \to \bar{x}$ with $v_k \in N^P_C(\bar{x})$. It is clear from the definition of the proximal normal cone that

$$N^P_C(\bar{x}) \subseteq \hat{N}_C(\bar{x}) \subseteq N_C(\bar{x}).$$

For more details on the proximal normal cone see for example Clarke, Ledyaev, Stern and Wolenski [8].
We will now consider the simple case when $(x, y) \in \mathbb{R}^2$ and we shall consider the set $K(x) = K = [0, 1]$ as the feasible set of the lower-level problem. Our aim is to precisely calculate $N_{\mathrm{gph}\, N_K}(\bar{y}, \bar{z})$. Observe that

$$N_K(y) = \begin{cases} (-\infty, 0] & : y = 0 \\ \{0\} & : 0 < y < 1 \\ [0, +\infty) & : y = 1 \end{cases}$$

It is easy to sketch the graph of the normal cone map N_K where $K = [0, 1]$. The proximal normal cone to $\mathrm{gph}\, N_K$ is given as follows.

$$N^P_{\mathrm{gph}\, N_K}(\bar{y}, \bar{z}) = \begin{cases} (-\infty, 0] \times [0, +\infty) & : \bar{y} = 0, \bar{z} = 0 \\ \{0\} \times \mathbb{R} & : 0 < \bar{y} < 1, \bar{z} = 0 \\ \mathbb{R} \times \{0\} & : \bar{y} = 1, \bar{z} > 0 \\ \mathbb{R} \times \{0\} & : \bar{y} = 0, \bar{z} < 0 \\ [0, +\infty) \times (-\infty, 0] & : \bar{y} = 1, \bar{z} = 0 \end{cases}$$

Using the fact that the basic normal cone can be obtained as a limit of the proximal normal cone we obtain the following

$$\begin{aligned} N_{\mathrm{gph}\, N_K}(0, 0) = \{(w, v) \in \mathbb{R}^2 : w < 0, v > 0\} &\cup \{(w, v) \in \mathbb{R}^2 : v = 0\} \\ &\cup \{(w, v) \in \mathbb{R}^2 : w = 0\} \end{aligned}$$

and

$$N_{\mathrm{gph}\, N_K}(1,0) = \{(w,v) \in \mathbb{R}^2 : w > 0, v < 0\} \cup \{(w,v) \in \mathbb{R}^2 : v = 0\} \\ \cup \{(w,v) \in \mathbb{R}^2 : w = 0\}.$$

For all other points the basic normal cone coincides with the proximal normal cone.
We have shown earlier that using the approach of Outrata [18] we are able to develop optimality conditions for the problem (P1) when $x \in X$ and X is a closed subset of $\mathbb{R}^n$. We would like to remark that by using the conditions (a), (b) and (c) in Theorem 3.2 of Ye and Ye [26] we can arrive at the same conditions as we have obtained using Outrata's approach. However the most interesting condition in Theorem 3.2 of [26] is (b). In our case this corresponds to the assumption that for each fixed $x \in X$ the function $y \mapsto f(x,y)$ is strongly convex. This may actually appear in practical situations.

We will now turn our attention to study the bilevel programming problem (P2) whose lower-level problem (LLP2) is of a primal-dual nature i.e. a minimax problem.

Theorem 4.4 *Let us consider the problem (P3) given as follows*

$$\min_{x,y,\lambda} F(x,y,\lambda) \quad \textit{subject to} \quad (x,y,\lambda) \in gph\ S,$$

where $F : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p \to \mathbb{R}$ is a smooth function and the set-valued map $S : \mathbb{R}^n \rightrightarrows \mathbb{R}^m \times \mathbb{R}^p$ is a solution set of the following problem (LLP2)

$$minimaximize L(x,y,\lambda) \quad \textit{subject to} \quad (y,\lambda) \in Y \times W,$$

where $Y \subset \mathbb{R}^m$ and $W \in \mathbb{R}^p$ are non-empty convex sets and $L(x,y,\lambda)$ is convex with respect to y for each $(x,\lambda) \in \mathbb{R}^n \times W$ and is concave in λ for each $(x,y) \in \mathbb{R}^n \times Y$. Further assume that $L(x,y,\lambda)$ is a twice continuously differentiable function. Let $(\bar{x},\bar{y},\bar{\lambda})$ be a solution to (P3). Set

$$\bar{p} = (\nabla_y L(\bar{x},\bar{y},\bar{\lambda}), -\nabla_\lambda L(\bar{x},\bar{y},\bar{\lambda})).$$

Further assume that the following qualification condition holds :

$$rank \left[\nabla^2_{xy} L(\bar{x},\bar{y},\bar{\lambda}) | \nabla^2_{x\lambda} L(\bar{x},\bar{y},\bar{\lambda})\right] = m + p$$

Then there exists $v^* \in \mathbb{R}^{m+p}$ such that

i) $0 \in \nabla_x F(x,y,\lambda) + \left[\nabla^2_{xy} L(\bar{x},\bar{y},\bar{\lambda}) | \nabla^2_{x\lambda} L(\bar{x},\bar{y},\bar{\lambda})\right] v^*$
ii) $0 \in \nabla_{(y,\lambda)} F(\bar{x},\bar{y},\bar{\lambda}) + \left[\nabla^2_{yy} L(\bar{x},\bar{y},\bar{\lambda}) | \nabla^2_{y\lambda} L(\bar{x},\bar{y},\bar{\lambda})\right] v^*$
$+ D^* N_{Y \times W}((\bar{y},\bar{\lambda}) | -\bar{p})(v^*)$.

Proof. Since $(\bar{x},\bar{y},\bar{\lambda})$ is a solution of (P3) then $(\bar{x},\bar{y},\bar{\lambda}) \in \mathrm{gph}\, S$. Hence from Proposition 1.4 in Dontchev and Rockafellar [12] we have that

$$-\nabla_y L(\bar{x},\bar{y},\bar{\lambda}) \in N_Y(\bar{y}) \quad \text{and} \quad \nabla_\lambda L(\bar{x},\bar{y},\bar{\lambda}) \in N_W(\bar{\lambda}).$$

This is equivalent to the fact that $(\bar{y}, \bar{\lambda})$ is solving the following variational inequality over $Y \times W$ namely

$$0 \in G(\bar{x}, y, \lambda) + N_{Y \times W}(y, \lambda),$$

where

$$G(\bar{x}, y, \lambda) = (\nabla_y L(\bar{x}, y, \lambda), -\nabla_\lambda L(\bar{x}, y, \lambda)).$$

The result then follows by direct application of Lemma 3.2. □

It is important to note that all the above optimality conditions are expressed in terms of the coderivative of the normal cone map. We can however provide a slightly different reformulation of the optimality conditions by using the second-order subdifferential of the indicator function. To do this let us observe that if the lower-level problem (LLP) in (P) is convex then the solution set mapping S can be equivalently written as follows

$$S(x) = \{y \in \mathbb{R}^m : 0 \in \nabla_y f(x, y) + \partial_y \delta_K(x, y)\},$$

where $\delta_K(x, y) = \delta_{K(x)}(y)$ denotes the indicator function for the set $K(y)$. For any function $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ which is finite at $\bar{x}$ the second-order subdifferential of f at $(\bar{x}, \bar{y})$ is the coderivative of the subdifferential map i.e.

$$\partial^2 f(\bar{x}|\bar{y})(u) = D^*(\partial f)(\bar{x}|\bar{y})(u).$$

Theorem 4.5 *Consider the problem (P1) and let $(\bar{x}, \bar{y})$ be a local solution of the problem. Consider that f is a twice continuously differentiable function. Assume that the following qualification condition holds:*

$$(u, 0) \in \partial^\infty \delta_K(\bar{x}, \bar{y}) \Longrightarrow u = 0.$$

Additionally assume that the following qualification condition also holds

$$0 \in \nabla^2 f(\bar{x}, \bar{y})^T(0, v_2) + \bigcup_{w \in \partial\delta_K(\bar{x},\bar{y}), \operatorname{proj}_2 w = -\nabla_y f(\bar{x},\bar{y})} \partial^2 \delta_K((\bar{x}, \bar{y})|w)(0, v_2)$$
$$\Longrightarrow v_2 = 0,$$

where proj_2 *denotes the projection on $\mathbb{R}^m$. Then there exists $v_2^* \in \mathbb{R}^m$ and $\bar{w} \in \partial\delta_K(\bar{x}, \bar{y})$ with* $\operatorname{proj}_2 \bar{w} = -\nabla_y f(\bar{x}, \bar{y})$ *such that*

$$0 \in \nabla F(\bar{x}, \bar{y}) + \nabla^2 f(\bar{x}, \bar{y})^T(0, v_2^*) + \partial^2 \delta_K((\bar{x}, \bar{y})|\bar{w})(0, v_2^*).$$

Proof. The key to the proof of this result is the Corollary 2.2 in Levy and Mordukhovich [14]. As per the Corollary 2.2 in Levy and Mordukhovich [14] the function δ_K should satisfy the following properties. First $\operatorname{gph} \partial_y \delta_K$ is closed. This is true since $\operatorname{gph} \partial_y \delta_K = \operatorname{gph} N_K$ and we know that $\operatorname{gph} N_K$ is closed. Second δ_k should be subdifferentially continuous at $(\bar{x}, \bar{y})$ for any $v \in \operatorname{gph} N_K$. This is true since the indicator function is subdifferentially continuous (see Rockafellar and Wets [21] pages 610-612). Now the result follows by a direct application of Corollary 2.2 in [14] □

Acknowledgments
We are grateful to Jiri Outrata and an anonymous referee whose constructive suggestions have vastly improved the presentation of the paper.

References

1. H. Babahadda and N. Gadhi: *Necessary optimality conditions for bilevel optimization problems using convexificators*, Technical Report, Department of Mathematics, Dhar El Mehrez, Sidi Mohamed Ben Abdellah University, Fes, Morocco, 2005.
2. B. Bank and J. Guddat and D. Klatte and B. Kummer and K. Tammer, *Non-Linear Parametric Optimization*, Akademie-Verlag, Berlin, 1982.
3. J. F. Bard, Optimality conditions for bilevel programming, *Naval Research Logistics Quaterly*, Vol 31, 1984, pp 13-26.
4. J. F. Bard, Convex two level optimization, *Mathematical Programming*, Vol 40, 1988, 15-27.
5. J. F. Bard, On some properties of bilevel programming, *Journal of Optimization Theory and Applications*, Vol 68, 1991, pp 371-378.
6. C. Berge, *Espaces topologiques: Fonctions multivoques*, (French) Collection Universitaire de Mathématiques, Vol. III Dunod, Paris, 1959. (In English) *Topological spaces : Including a treatment of multi-valued functions*, vector spaces and convexity. Translated from the French original by E. M. Patterson. Reprint of the 1963 translation. Dover Publications, Inc., Mineola, NY, 1997.
7. , G. Y. Chen and J. Jahn (eds.), *Mathematical Methods of Operations Research* Vol. 48, 1998.
8. F. H. Clarke, Yu. Ledyaev, R. Stern and P. Wolenski, *Nonsmooth Analysis and Control Theory*, Springer, 1998.
9. S. Dempe, *Foundations of Bilevel Programming*, Kluwer Academic Publishers, 2002.
10. S. Dempe, A necessary and sufficient optimality condition for bilevel prgramming problems, *Optimization*, Vol 25, 1992, pp 341-354.
11. A. L. Dontchev and R. T. Rockafellar, Ample paramerization of variational inclusions, *SIAM Journal on Optimization*, Vol 12, 2001, pp 170-187.
12. A.L. Dontchev and R. T. Rockafellar, Primal-dual solution perturbations in convex optimization, *Set-Valued Analysis*, Vol 9, 2001, pp 49-65.
13. M. Kojima, Strongly stable stationary solutions in nonlinear programs, In: S.M. Robinson (ed.): Analysis and Computation of Fixed Points, Academic Press, New York, 1980,, pp. 93–138.
14. A. B. Levy and B.S.Mordukhovich, Coderivatives in parametric optimization, *Mathematical Programming*, Vol 99, 2004, pp 311-327.
15. P. Loridan and J. Morgan, Weak via strong Stackelberg problem : New Results. *Journal of Global Optimization*, Vol 8, 1996, pp 263-287.
16. J. A. Mirrlees, The theory of moral hazard and unobservable bevaviour: part I, *Review of Economic Studies*, Vol. 66, 1999, pp. 3-21.
17. B. S. Mordukhovich, Generalized differential calculus for nonsmooth and set-valued mappings, *Journal of Mathematical Analysis and Applications*, Vol 183, 1994, pp 250-288.

18. J. Outrata, A generalized mathematical program with equilibrium constraints, *SIAM Journal on Control and Optimization*, Vol 38, 2000, pp 1623-1638.
19. J.-S. Pang and M. Fukushima, Complementarity constraint qualifications and simplified B-stationarity conditions for mathematical programs with equilibrium constraints, *Computational Optimization and Applications*, Vol. 13, 1999, pp. 111-136.
20. D. Ralph and S. Dempe, Directional Derivatives of the Solution of a Parametric Nonlinear Program, *Mathematical Programming*, Vol. 70, 1995, pp. 159-172.
21. R. T. Rockafellar and R. J. B. Wets, *Variational Analysis*, Springer-Verlag, 1998 (Second corrected printing 2004).
22. H. Scheel and S. Scholtes, *Mathematical programs with equilibrium constraints: stationarity, optimality, and sensitivity*, Mathematics of Operations Research, Vol. 25, 2000, pp. 1-22.
23. J. J. Ye, Nondifferentiable multiplier rules for optimization and bilevel optimization problems, *SIAM Journal on Optimization*, Vol. 15, 2004, pp. 252-274.
24. J. J. Ye, Constraint qualifications and necessary optimality conditions for optimization problems with variational inequality constraints, *SIAM Journal on Optimization*, Vol 10, 2000, pp 943-962.
25. J. J. Ye, Optimality conditions for optimization problems with complementarity constraints, *SIAM Journal on Optimization* Vol9, 1999, pp 374-387.
26. J. J. Ye and X. Y. Ye, Necessary optimality conditions for optimization problems with variational inequality constraints, *Mathematics of Operations Research*, Vol 22, 1997, pp 977-997.
27. J. J. Ye and D. L. Zhu, Optimality conditions for bilevel programming problems, *Optimization* Vol 33, 1995, pp 9-27.
28. J. J. Ye and D.L. Zhu, A note on optimality conditions for bilevel programming,*Optimization 39*,1997,pp 361-366.
29. J. J. Ye, D. L. Zhu and Q. Z. Zhu, Exact penalization and necessary optimality conditions for generalized bilevel programming problem, *SIAM Journal on Optimization*,Vol 4, 1994, pp 481-507.