Optimization,
Vol. 55, Nos. 5–6, October–December 2006, 505–524

Optimality conditions for bilevel programming problems

S. DEMPE*†, J. DUTTA‡ and S. LOHSE†

†Department of Mathematics and Computer Science, Technical University Bergakademie Freiberg, Germany
‡Department of Mathematics, Indian Institute of Technology, Kanpur, India

(*Received 2 August 2005; in final form 7 February 2006*)

Focus in the article is on necessary optimality conditions for bilevel programming problems. We first identify approaches that seem to be promising. Then, two such approaches are investigated. Using the nonexistence of a descent direction for the objective function within the tangent cone to the feasible set as necessary optimality condition we need efficient ways to describe this tangent cone. We describe possibilities for this in three different special cases. We conclude this article with the second approach applying Mordukhovich's coderivative to an appropriately reformulated bilevel programming problem.

Keywords: Bilevel programming; Necessary and sufficient optimality conditions

Mathematics Subject Classifications 2000: 90C46; 90C30

1. Introduction

Bilevel programming problems are hierarchical ones consisting of two combined optimization problems. The variables of the first (or upper-level) problem are the parameters of the second (or lower-level) problem, and the optimal solution of the latter is needed to calculate the objective function value of the former. In other words, the upper-level decision maker (or leader) fixes his selection x first, the second one, the follower or lower-level decision maker determines his solution y later in full knowledge of the leader's choice. This means that the variables x play the role of parameters in the follower's problem. Then, the leader has to anticipate the follower's selection since his revenue depends not only on his own selection but also on the follower's reaction.

*Corresponding author. Email: dempe@math.tu-freiberg.de

Optimization
ISSN 0233-1934 print: ISSN 1029-4945 online
http://www.tandf.co.uk/journals
DOI: 10.1080/02331930600816189

To formulate the problem, let the follower make his decision by solving a parametric optimization problem

$$\Psi(x) := \underset{y}{\operatorname{argmin}}\{f(x,y) \colon g(x,y) \le 0\}, \tag{1}$$

where $f, g_i \colon \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$, $i = 1, \ldots, p$ are smooth functions. To avoid difficulties related to the existence of local optimal solutions in the lower-level problem assume throughout this article that the functions $f(x, \cdot)$ and $g_i(x, \cdot), i = 1, \ldots, p$, are convex.

Then, the leader's problem consists in minimizing the function $F : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ subject to the constraints $y \in \Psi(x)$ and $x \in X$, where $X \subseteq \mathbb{R}^n$ is a closed set. This problem has been discussed in the monographs Bard [3] and Dempe [4] and in the annotated bibliography Dempe [5].

If the optimal solution of problem (1) is uniquely determined for all parameter values, the leader's problem reduces to a nondifferentiable (under certain assumptions Lipschitz continuous) problem with an implicitly determined objective function and/or feasible set. Since the resulting problem has a special structure, it earns extensive interest in nondifferentiable programming over the last two decades. Challenging topics are the formulation both of optimality conditions and solution algorithms. Optimality conditions can be found, e.g., in [4,23]. Sources for solution algorithms are [3] as well as [6].

Since, strongly speaking, the leader has control over the variable x only, this problem is properly determined only in the case when the optimal solution of the lower-level problem (1) is uniquely determined for all parameter values $x \in X$. If this is not the case, the optimistic and pessimistic approaches have been considered in the literature, see, e.g. [12]. Both approaches rest on the introduction of a new second-level problem.

Since focus is on the optimistic approach only in this article, we will not formulate the pessimistic one here. The optimistic approach can be applied if the leader assumes that the follower will in any case take an optimal solution which is a best one from the leader's point of view. This leads to the problem

$$\min\{\varphi_o(x) \colon x \in X\}, \tag{2}$$

where

$$\varphi_o(x) := \min_y\{F(x,y) \colon y \in \Psi(x)\}. \tag{3}$$

The use of problem (2) leads to the following notion of an optimal solution:

Definition 1.1 A point $(\bar{x}, \bar{y})$ is called a local optimistic optimal solution of the bilevel programming problem if

$$\bar{y} \in \Psi(\bar{x}), \bar{x} \in X, F(\bar{x}, \bar{y}) = \varphi_o(\bar{x})$$

and there is a number $\varepsilon > 0$ such that

$$\varphi_o(x) \ge \varphi_o(\bar{x}) \quad \forall x \in X, \ \|x - \bar{x}\| < \varepsilon.$$

Problem (2) is obviously equivalent to

$$\min_{x,y}\{F(x,y)\colon x\in X,\ \ y\in\Psi(x)\} \tag{4}$$

provided that the latter problem has an optimal solution. But note that this equivalence is true only for global minima [4].

Remark 1.1 Let $\bar{x}\in X$ be a local optimal solution of problem (2) and let $\bar{y}\in\Psi(\bar{x})$ be given with $F(\bar{x},\bar{y})=\varphi_o(\bar{x})$. Then $(\bar{x},\bar{y})$ is also a local optimal solution of problem (4) provided that Slater's condition is satisfied for problem (1) at $(\bar{x},\ \bar{y})$ and $\Psi(\cdot)$ is locally compact at $\bar{x}$.

Proof The assumptions of the remark imply that the point-to-set mapping $\Psi(\cdot)$ is upper-semicontinuous at $\bar{x}$ [2]. Thus, the proof follows from [9].

The opposite implication to Remark 1.1 is in general not true [4].

Now, $(\bar{x},\bar{y})\in M:=\{(x,y)\colon x\in X, y\in\Psi(x)\}$ is a local optimal solution of problem (4) provided there is a number $\varepsilon>0$ such that $F(x,y)\geq F(\bar{x},\bar{y})$ for all $(x,y)\in M$ with $\|x-\bar{x}\|<\varepsilon$.

For existence of an optimal solution of problem (4) we need closedness of the set M and either boundedness of M or coercivity of F over M. Closedness of M is implied by upper semicontinuity of the point-to-set mapping Ψ, which in turn follows from local boundedness of $\Psi(x)$ and validity of a regularity condition for the lower-level problem (1) [2].

Bilevel programming problems are closely related to other optimization problems and have many applications, see [3–5] and the references therein. To attack them, they need to be transformed into one-level problems. In the following, we present different reformulations and describe our opinion about their usefulness.

The most used reformulation is by replacing the lower-level problem of (4) with its Karush–Kuhn–Tucker (KKT) conditions. This leads to the problem

$$\min_{x,y,\lambda}\{F(x,y)\colon x\in X,\ \nabla_y L(x,y,\ \lambda)=0,\ g(x,y)\leq 0,\ \lambda\geq 0,\ \lambda^\top g(x,y)=0\}. \tag{5}$$

This problem is a special kind of a so-called mathematical program with equilibrium constraints (MPEC), which has widely been investigated in the literature, see e.g. [13, 17]. Optimality conditions for this problem can be found, e.g., in [19,21,24,25]. The main difficulty in using this reformulation for the construction of optimality conditions for the bilevel programming problem stems from the introduction of new variables, namely, the Lagrange multipliers for the lower-level problem. This implies that a local optimal solution $(\bar{x},\ \bar{y})$ of the bilevel programming problem (4) is a local optimal solution $(\bar{x},\bar{y},\lambda)$ of problem (5) for all Lagrange multipliers λ corresponding to the optimal solution $\bar{y}$ of the lower-level problem (1) with $x=\bar{x}$. On the other hand, a local optimal solution $(\bar{x},\bar{y},\bar{\lambda})$ of problem (5) need not to represent a local optimal solution of problem (4). To overcome this difficulty we have either to develop an optimality condition using all the Lagrange multipliers of the lower-level problem (and have hence to enumerate them) or to find a condition that is independent of the Lagrange multiplier. In the following, we will describe one such approach. It should be noted that this approach is not possible for the pessimistic solution.

A second approach uses the optimal value function of the lower-level problem (1)

$$\varphi(x) = \min_y \{f(x, y): g(x, y) \le 0\}$$

and replaces (4) by

$$\min_{x, y} \{F(x, y): x \in X,\ f(x, y) \le \varphi(y),\ g(x, y) \le 0\}. \tag{6}$$

This approach has been used in the articles Babahadda and Gadhi [1] and Ye and Zhu [26] to describe optimality conditions. We will not apply this approach here.

A third approach uses problem (4) directly and develops optimality conditions by means of the contingent or normal cone of the graph of the point-to-set mapping Ψ. This will be our first method in the following text.

A fourth approach replaces the lower-level problem by necessary and sufficient optimality conditions avoiding Lagrange multipliers. To simplify the notations define the feasible set mapping as

$$K(x) := \{y: g(x, y) \le 0\}.$$

Using this notation and assuming validity of a regularity condition, the optimistic bilevel programming problem (6) can equivalently be replaced by

$$\min_{x, y} \{F(x, y): x \in X,\ 0 \in \nabla_y f(x, y) + N_{K(x)}(y),\ g(x, y) \le 0\}, \tag{7}$$

where $N_{K(x)}(y)$ denotes some normal cone to $K(x)$ at the point y. We will show subsequently that this reformulation can fruitfully be used to derive optimality conditions using, e.g., the coderivative by Mordukhovich.

2. Optimality conditions using the contingent cone

2.1. *The general approach*

The contingent (or Bouligand) cone of the set $M = \{(x, y): x \in X, y \in \Psi(x)\}$ is defined via

$$\begin{aligned} C_M(x, y) := \Big\{ &(u, v)^\top : \exists \{t_k\}_{k=1}^\infty \subset \mathbb{R}_+,\ \exists \big\{(u^k, v^k)^\top\big\}_{k=1}^\infty \subset \mathbb{R}^n \times \mathbb{R}^m, \\ &\text{with } (x, y)^\top + t_k \big(u^k, v^k\big)^\top \in \operatorname{gph} \Psi\ \forall k,\ x + t_k u^k \in X, \\ &\lim_{k\to\infty} t_k = 0,\ \lim_{k\to\infty} \big(u^k, v^k\big)^\top = (u, v)^\top \Big\} \end{aligned}$$

where $\operatorname{gph} \Psi = \{(x, y)^\top : y \in \Psi(x)\}$ denotes the graph of the point-to-set mapping Ψ. In short,

$$C_M(x, y) = \limsup_{t\to 0+} t^{-1}(M - (x, y)^\top),$$

where the term lim sup denotes the outer limit or the Kuratowski–Painleve upper limit (see, for example, [20] for more details).

THEOREM 2.1 *If the point* $(\bar{x},\bar{y})^\top \in M$ *is a local optimal solution of the bilevel programming problem* (4), *then*

$$\nabla F(\bar{x},\bar{y})(d,r)^\top \geq 0$$

for all

$$(d,r)^\top \in C_M(\bar{x},\,\bar{y}).$$

On the other hand, if $(\bar{x},\bar{y})^\top \in M$ *and*

$$\nabla F(\bar{x},\bar{y})(d,r)^\top > 0$$

for all

$$(d,r)^\top \in C_M(\bar{x},\bar{y}),$$

then the point $(\bar{x},\bar{y})^\top$ *is a local optimal solution of* (4).

Proof (i) Let $(\bar{x},\bar{y})^\top \in M$ be a local optimal solution of problem (4). Assume that the proposition of the first part of the theorem is not satisfied. Then, there exists a direction $(d,r)^\top$ with

$$(d,r)^\top \in C_M(\bar{x},\bar{y})$$

and

$$\nabla F(\bar{x},\bar{y})(d,r)^\top < 0. \tag{8}$$

Hence, by definition there are two sequences $\{t_k\}_{k=1}^{\infty} \subset \mathbb{R}_+$ and $\{(u^k,v^k)^\top\}_{k=1}^{\infty} \subset \mathbb{R}^n \times \mathbb{R}^m$ with $(\bar{x},\bar{y})^\top + t_k(u^k,v^k)^\top \in \operatorname{gph}\Psi$, $\bar{x}+t_k u^k \in X$, $\forall k \in \mathbb{N}$, $\lim\limits_{k\to\infty} t_k = 0$, $\lim\limits_{k\to\infty}(u^k,v^k)^\top = (d,r)^\top$. Thus, using the definition of the derivative, we get

$$F(\bar{x}+t_k u^k, \bar{y}+t_k v^k) = F(\bar{x},\bar{y}) + t_k \nabla F(\bar{x},\bar{y})(u^k,v^k)^\top + o(t_k)$$

for each sufficiently large k, where $\lim_{k\to\infty} o(t_k)/t_k = 0$. Since

$$\lim_{k\to\infty}\left\{\nabla F(\bar{x},\bar{y})(u^k,v^k)^\top + \frac{o(t_k)}{t_k}\right\} = \nabla F(\bar{x},\bar{y})(d,r)^\top < 0$$

by the assumption this implies

$$\nabla F(\bar{x},\bar{y})(u^k,v^k)^\top + \frac{o(t_k)}{t_k} < 0$$

for all sufficiently large k and, hence,

$$F(\bar{x} + t_k u^k, \bar{y} + t_k v^k) < F(\bar{x}, \bar{y})$$

for large k. This leads to a contradiction to local optimality.

(ii) Now, let $\nabla F(\bar{x}, \bar{y})(d, r)^\top > 0$ for all $(d, r)^\top \in C_M(\bar{x}, \bar{y})$ and assume that there is a sequence $(x^k, y^k) \in M$ converging to $(\bar{x}, \bar{y})^\top$ with $F(x^k, y^k) < F(\bar{x}, \bar{y})$ for all k. Then,

$$\left(\frac{x^k - \bar{x}}{\|(x^k, y^k) - (\bar{x}, \bar{y})\|}, \frac{y^k - \bar{y}}{\|(x^k, y^k) - (\bar{x}, \bar{y})\|} \right)^\top$$

converges to some $(d, r)^\top \in C_M(\bar{x}, \bar{y})$. Using differential calculus, it is now easy to verify that

$$\nabla F(\bar{x}, \bar{y})(d, r)^\top \leq 0,$$

contradicting our assumption.

2.2. *The linear case*

If bilevel programming problems with linear lower-level problems are under consideration, a stronger formulation of the optimality condition is possible [7]. For this, consider a linear parametric optimization problem

$$\max_y \left\{ c^\top y \colon\ Ay = b,\ y \geq 0 \right\} \tag{9}$$

with a matrix A of appropriate dimension and parameters in the right-hand side as well as in the objective function. Let $\Psi_L(b, c)$ denote the set of optimal solutions of (9). A special optimistic bilevel programming problem reads as

$$\min_{y, b, c} \left\{ f(y) \colon Bb = \widetilde{b}, Cc = \widetilde{c}, y \in \Psi_L(b, c) \right\}. \tag{10}$$

Our aim is it now to apply Theorem 2.1 to this problem while finding an explicit description of the contingent cone to the feasible set. For this we start with the KKT reformulation of the lower-level problem (5):

$$\begin{aligned} f(y) &\longrightarrow \min_{y, b, c, u} \\ Ay &= b \\ y &\geq 0 \\ A^\top u &\geq c \\ y^\top (A^\top u - c) &= 0 \\ Bb &= \widetilde{b} \\ Cc &= \widetilde{c}. \end{aligned} \tag{11}$$

It should be noted that the objective function in the upper-level problem does not depend on the parameters of the lower-level one. This makes a more precise definition of a local optimal solution of problem (10) necessary.

Definition 2.1 A point $\bar{y}$, is a local optimal solution of problem (10) if there exists an open neighborhood U of $\bar{y}$ such that $f(\bar{y}) \leq f(y)$ for all y, b, c with $Bb = \widetilde{b}$, $Cc = \widetilde{c}$ and $y \in U \cap \Psi_L(b, c)$.

The main result of this definition is the possibility to drop the explicit dependence of the solution of problem (10) on c. This dependence rests on solvability of the dual problem and is guaranteed for index sets I in the set $\mathcal{I}(y)$.

Let the following index sets be determined at some point $\bar{y}$:

(1) $I(\bar{y}) = \{i\colon \bar{y}_i = 0\}$,
(2) $I(u, c) = \{i\colon (A^\top u - c)_i > 0\}$,
(3) $\mathcal{I}(\bar{y}) = \{I(u, c)\colon A^\top u \geq c,\ (A^\top u - c)_i = 0\ \forall i \notin I(\bar{y}),\ Cc = \widetilde{c}\}$, and
(4) $I^0(\bar{y}) = \bigcap_{I \in \mathcal{I}(\bar{y})} I$.

Using these definitions, problem (11) can be transformed into the following one by replacing the complementarity conditions:

$$
\begin{aligned}
f(y) &\longrightarrow \min_{y, b, I} \\
Ay &= b \\
y &\geq 0 \\
y_i &= 0 \quad \forall i \in I \\
Bb &= \tilde{b} \\
I &\in \mathcal{I}(y).
\end{aligned}
\tag{12}
$$

In this problem, the index set I enters the variables in the sense that problem (12) is to be solved for all fixed sets $I \in \mathcal{I}(y)$ and the best solution over all resulting problems is selected.

The contingent cone to the feasible set of the last problem (or rather its projection onto $\mathbb{R}^m$) is

$$C(\bar{y}) := \bigcup_{I \in \mathcal{I}(\bar{y})} C_I(\bar{y}),$$

where

$$C_I(\bar{y}) = \left\{d \mid \exists r\colon\ Ad = r,\ Br = 0,\ d_i \geq 0\ \forall i \in I(\bar{y}) \setminus I,\ d_i = 0\ \forall i \in I\right\}$$

for all $I \in \mathcal{I}(\bar{y})$.

Remark 2.1 If f is differentiable at $\bar{y}$, this point is a local optimal solution of (10) if and only if $\nabla f(\bar{y}) \cdot d \geq 0$ for all $d \in \operatorname{conv} C(\bar{y})$.

Consider the relaxed problem of (12)

$$
\begin{aligned}
f(y) &\longrightarrow \min_{y,b} \\
Ay &= b \\
y_i &\geq 0 \quad i \in I(\bar{y}) \backslash I^0(\bar{y}) \\
y_i &= 0 \quad i \in I^0(\bar{y}) \\
Bb &= \widetilde{b}
\end{aligned}
\tag{13}
$$

together with the contingent cone to its feasible set (or better, its projection onto $\mathbb{R}^m$)

$$C_R(\bar{y}) = \left\{ d \mid \exists r\colon Ad = r,\ Br = 0,\ d_i \geq 0, i \in I(\bar{y}) \backslash I^0(\bar{y}),\ d_i = 0, i \in I^0(\bar{y}) \right\}$$

at the point $\bar{y}$.

Remark 2.2 We have $j \in I(\bar{y}) \backslash I^0(\bar{y})$ if and only if the system

$$
\begin{aligned}
\left(A^\top u - c\right)_i &= 0 \quad \forall i \notin I(\bar{y}) \\
\left(A^\top u - c\right)_j &= 0 \\
\left(A^\top u - c\right)_i &\geq 0 \quad \forall i \in I(\bar{y}) \backslash \{j\} \\
Cc &= \tilde{c}
\end{aligned}
$$

has a solution. This implies that it is possible to compute the index set $I^0(\bar{y})$ in polynomial time. Hence, also the contingent cone $C_R(\bar{y})$ can be computed in polynomial time.

In the following theorem, we need an assumption: the point $\bar{y}$ is said to satisfy the *full-rank condition* (FRC) if

$$\operatorname{span}\left(\left\{A_i\colon\ i \notin I(\bar{y})\right\}\right) = \mathbb{R}^m, \tag{14}$$

where A_i denotes the ith column of the matrix A.

THEOREM 2.2 [7] *Let (FRC) be satisfied at the point $\bar{y}$. Then,*

$$\operatorname{conv} C(\bar{y}) = \operatorname{cone} C(\bar{y}) = C_R(\bar{y}). \tag{15}$$

We add the proof for completeness of the article.

Proof Let, for simplicity, $I(\bar{y}) = \{1, \ldots, k\}$ and $I^0(\bar{y}) = \{l+1, \ldots, k\}$. The inclusion

$$\operatorname{conv} C(\bar{y}) = \operatorname{cone} C(\bar{y}) \subseteq C_R(\bar{y}) \tag{16}$$

is obviously satisfied. Let $\bar{d}$ be an arbitrary element of $C_R(\bar{y})$, which means there is a $\bar{r}$ with $A\bar{d} = \bar{r}$, $B\bar{r} = 0$, $\bar{d}_i \geq 0 \quad i = 1, \ldots, l$, $\bar{d}_i = 0, \quad i = l+1, \ldots, k$. We consider the following linear systems

$$\begin{aligned} Ad &= \bar{r} \\ d_1 &= \bar{d}_1 \qquad\qquad (S_1) \\ d_i &= 0, \quad i = 2, \ldots, k \end{aligned}$$

and

$$\begin{aligned} Ad &= 0 \\ d_j &= \bar{d}_j \qquad\qquad (S_j) \\ d_i &= 0, \quad i = 1, \ldots, k,\ i \neq j \end{aligned}$$

for $j = 2, \ldots, l$. These systems are all feasible because of (FRC).

Furthermore let $d^1, \ldots, d^l$ be (arbitrary) solutions of the systems (S_j). We define now the direction $d = \sum_{j=1}^{l} d^j$ and get $d_i = \bar{d}_i$ for $i = 1, \ldots, k$ as well as $Ad = A\bar{d} = \bar{r}$. Because we chose arbitrary vectors $d^1, \ldots, d^l$ it is possible that $d \neq \bar{d}$. But we can achieve equality with a translation of the solution d^1 by a specific vector of $\mathcal{N}(A) = \{z\colon\ Az = 0\}$. Therefore, we define $\hat{d}^1 := d^1 + \bar{d} - d$, and because d^1 is feasible for (S_1) and $d_i = \bar{d}_i$ for $i = 1, \ldots, k$ as well as $Ad = A\bar{d} = \bar{r}$ we get $\hat{d}^1_i = 0$ for all $i = 2, \ldots, k$ and $A\hat{d}^1 = A(d^1 + \bar{d} - d) = \bar{r} + \bar{r} - \bar{r} = \bar{r}$. Hence $\hat{d}^1$ is also a solution of (S_1). Thus we have $\hat{d}^1 + \sum_{j=2}^{l} d^j = \bar{d} - d + \sum_{j=1}^{l} d^j = \bar{d}$. As a result of the definition of the set $I^0(\bar{y})$ there are index sets $I_j \in \mathcal{I}(\bar{y})$ with $j \notin I_j$ for all $j \in \{1, \ldots, l\} = I(\bar{y}) \backslash I^0(\bar{y})$. So $\hat{d}^1$ is an element of the contingent cone $C_{I_1}(\bar{y})$ and d^j are elements of the contingent cones $C_{I_j}(\bar{y})$ for $j = 2, \ldots, l$. Finally $\bar{d}$ is the sum of a finite number of elements of $C(\bar{y})$ and therefore $C_R(\bar{y}) \subseteq \operatorname{cone} C(\bar{y})$.

Remark 2.3 Using the Farkas lemma of the alternative it is now a simple task to formulate the conditions of Theorem 2.2 in dual space as KKT necessary optimality conditions. Moreover, due to Remark 2.2 verification of local optimality can be done in polynomial time.

2.3. *The nonlinear case*

We consider now the inverse optimization problem

$$\min_{x,y,c}\left\{f(y)\colon Cc = \widetilde{c}, Xx = \widetilde{x}, y \in \Psi_N(x,c)\right\} \tag{17}$$

with a nonlinear lower-level problem

$$\Psi_N(x,c) = \operatorname*{argmin}_y \left\{c^\top y\colon g(x,y) \leq 0\right\}, \tag{18}$$

where $g\colon \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^p$ and $f\colon \mathbb{R}^m \to \mathbb{R}$ are sufficiently smooth functions. Assume that $g_i(x,\cdot)$ are convex, $i = 1, \ldots, p$, and that Slater's condition is satisfied for problem (18) for all x. Assume without loss of generality that the matrix X has full row rank. The assumption that the lower-level problem has a linear objective function is of no loss of generality since this can be obtained by adding one variable bounding the objective function of a nonlinear programming problem from above and minimizing it.

Again we intend to use Theorem 2.1 to formulate optimality conditions. Once more we start with the KKT reformulation of the problem:

$$\begin{aligned} f(y) &\to \min_{x,y,u,c} \\ Xx = \widetilde{x},\ Cc &= \widetilde{c} \\ c + u^\top \nabla_y g(x,y) &= 0 \\ g(x,y) \leq 0, u \geq 0, u^\top g(x,y) &= 0. \end{aligned} \tag{19}$$

Denote by

$$M = \{(x,y)\colon \exists c \text{ with } Cc = \widetilde{c}, Xx = \widetilde{x}, y \in \Psi_N(x,c)\}$$

the feasible set of problem (17), where again, as in section 2.2, the lower-level objective function coefficients are considered as rather abdicable. This means that we use the following optimality notion:

Definition 2.2 A point $(\bar{x},\bar{y}) \in M$ is a local optimal solution of problem (17) if there exists an $\varepsilon > 0$ such that for all $(x,y) \in M$ with $\|(x,y)-(\bar{x},\bar{y})\| \leq \varepsilon$ we have $f(y) \geq f(\bar{y})$.

Let

$$\begin{aligned} &I(x,y) = \{i\colon g_i(x,y) = 0\}, \\ &\mathcal{I}(x,y) = \{I \subseteq I(x,y)\colon \exists c \text{ with } Cc = \widetilde{c},\ -c \in \operatorname{cone}\{\nabla_y g_i(x,y)\colon i \in I\}\} \end{aligned}$$

and

$$I^0(x,y) = \bigcap_{I \in \mathcal{I}(x,y)} I.$$

Similar to section 2.2, it is relatively easy to determine the set $I^0(x,y)$. To derive a verificable optimality condition for the problem (17) we will formulate two different results being based on different assumptions.

Problem (19) is related to the following problem with less variables locally around some feasible point $(\bar{x},\bar{y})$:

$$\begin{aligned} f(y) &\to \min_{x,y,I} \\ Xx &= \widetilde{x}, \\ g_i(x,y) &= 0 \quad i \in I \\ g_i(x,y) &\leq 0 \quad i \in I(\bar{x},\bar{y})\setminus I \\ I &\in \mathcal{I}(\bar{x},\bar{y}). \end{aligned} \tag{20}$$

The following example shows yet that both problems (19) and (20) are not equivalent.

Example 2.1 Let $C = \{(-2, 0, -1) + t(1, 0, 0)\colon\ t \in \mathbb{R}\}$ and let for simplification the lower-level problem be independent of the parameter x. It is easy to introduce this parameter and to relate this example to the bilevel programming problem (17). The example shows that the feasible set of problem (20) has in general a larger feasible set than problem (19). Let the feasible set of the lower-level problem (18) be given by

$$\begin{aligned} g_1(y) &= y_1^2 + y_2^2 - y_3 \\ g_2(y) &= y_1 - y_2^2 + y_3 \\ g_3(y) &= y_1. \end{aligned}$$

and let $\bar{y} = 0$. Then, Slater's condition is satisfied. Using $c = (-2, 0, -1)^\top$ we get

$$\begin{pmatrix} 2 \\ 0 \\ 1 \end{pmatrix} \in \text{cone}\left\{ \begin{pmatrix} 0 \\ 0 \\ -1 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} \right\}.$$

Take $I = \{1, 2, 3\}$ and $\widetilde{y} = (0, \epsilon, \epsilon^2)$. The point $\widetilde{y}$ is feasible for the problem (20) with I.

For feasibility of the point $\widetilde{y}$ for problem (19) we need solvability of the equation

$$\begin{pmatrix} 2 \\ 0 \\ 1 \end{pmatrix} + t\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} = \lambda_1 \begin{pmatrix} 0 \\ 2\epsilon \\ -1 \end{pmatrix} + \lambda_2 \begin{pmatrix} 1 \\ -2\epsilon \\ 1 \end{pmatrix} + \lambda_3 \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}.$$

But since this is not possible, $\widetilde{y}$ is not feasible for (19).

THEOREM 2.3 *The two problems* (19) *and* (20) *are locally equivalent if and only if there exists a sufficiently small neighborhood* U *of* $(\bar{x}, \bar{y})$ *such that for each point* $(x, y) \in U$ *there is an index set* $I \in \mathcal{I}(\bar{x}, \bar{y})$ *with* $I \subseteq I(x, y)$ *such that*

$$\left\{c\colon Cc = \widetilde{c},\ -c \in \text{cone}\left\{\nabla_y g_i(x, y)\colon i \in I\right\}\right\} \neq \emptyset.$$

COROLLARY 2.1 *The assumption in this theorem is obviously satisfied if*

$$\left\{c\colon Cc = \widetilde{c},\ -c \in \text{ int cone}\left\{\nabla_y g_i(x, y)\colon i \in I\right\}\right\} \neq \emptyset$$

is satisfied for all $I \in \mathcal{I}(\bar{x}, \bar{y})$.

The linearizing cone to the feasible set of (20) is

$$C_L(\bar{x}, \bar{y}) = \bigcup_{I \in \mathcal{I}(\bar{x}, \bar{y})} \left\{ (r, d)^\top\colon Xr = 0, \nabla g_i(\bar{x}, \bar{y})(r, d)^\top \begin{cases} = 0, & i \in I \\ \leq 0, & i \in I(\bar{x}, \bar{y}) \setminus I \end{cases} \right\}.$$

Under the assumption in Theorem 2.3, this (in general nonconvex) cone equals the contingent cone of M provided that the Mangasarian–Fromowitz constraint qualification is valid for each of the feasible sets of the problems in (20) for a fixed set $I \in \mathcal{I}(\bar{x}, \bar{y})$.

Now, consider the relaxed problem

$$\begin{aligned} f(y) &\to \min_{x,y} \\ Xx &= \widetilde{x}, \\ g_i(x,y) &= 0, \quad i \in I^0(\bar{x},\bar{y}) \\ g_i(x,y) &\leq 0, \quad i \in I(\bar{x},\bar{y}) \setminus I^0(\bar{x},\bar{y}). \end{aligned} \tag{21}$$

The linearizing cone to the feasible set of problem (21) is

$$C_N(\bar{x},\bar{y}) = \left\{ (r,d)^\top : Xr = 0, \nabla g_i(\bar{x},\bar{y})(r,d)^\top \begin{cases} = 0, & i \in I^0(\bar{x},\bar{y}) \\ \leq 0 & i \in I(\bar{x},\bar{y}) \setminus I^0(\bar{x},\bar{y}) \end{cases} \right\}$$

and we have $C_L(\bar{x},\bar{y}) \subseteq C_N(\bar{x},\bar{y})$. The cone $C_N(\bar{x},\bar{y})$ coincides with the contingent cone of problem (21) if the Mangasarian–Fromowitz constraint qualification is valid.

Theorem 2.4 *If the Mangasarian–Fromowitz constraint qualification is satisfied for problem* (21) *and*

$$\left\{ c \colon Cc = \widetilde{c}, \; -c \in \operatorname{cone}\left\{ \nabla_y g_i(x,y) \colon i \in I \right\} \right\} \neq \emptyset$$

for all sets $I \supset I^0(\bar{x},\bar{y})$ *with* $\{\nabla_y g_i(\bar{x},\bar{y}) \colon i \in I\}$ *are linearly independent, then the cones* $C_N(\bar{x},\bar{y})$ *and* $\operatorname{conv} C_L(\bar{x},\bar{y})$ *coincide.*

Proof Clearly, by definition we have $C_L(\bar{x},\bar{y}) \subseteq C_N(\bar{x},\bar{y})$ and the latter is a convex polyhedral cone. Hence, it is equal to the set of all nonnegative linear combinations of a finite set of vectors $(r^k, d^k), k = 1, \ldots, s$. As these vectors we can take solutions of the following systems of equations and inequalities:

$$\begin{aligned} &Xr = 0 \\ &\nabla g_i(\bar{x},\bar{y})(r,d)^\top \begin{cases} = 0, & \text{for } i \in I \\ \leq 0, & \text{for } i \in I(\bar{x},\bar{y}) \setminus I \end{cases} \end{aligned}$$

with $I^0(\bar{x},\bar{y}) \subseteq I \subseteq I(\bar{x},\bar{y})$ and I is an index set with maximal cardinality of linearly independent gradients $\nabla_y g_i(\bar{x},\bar{y})$. Hence, the rows in the matrix

$$\begin{pmatrix} X & 0 \\ \nabla_x g_I(\bar{x},\bar{y}) & \nabla_y g_I(\bar{x},\bar{y}) \end{pmatrix}$$

are linearly independent. Hence, $\{\nabla_y g_i(\bar{x},\bar{y}) \colon i \in I\}$ are linearly independent. The assumptions now show that $I \in \mathcal{I}(\bar{x},\bar{y})$, i.e., the vector (r^k, d^k) is an element of one of the convex subcones in $C_L(\bar{x},\bar{y})$.

Remark 2.4 The assumptions of the theorem are satisfied, e.g., if the system of equations $Cc = \widetilde{c}$ disappears.

COROLLARY 2.2 *As a simple implication of this theorem we again obtain necessary optimality conditions of KKT type. It is, moreover, possible to derive also sufficient conditions for local optimality.*

The following theorem shows that we can get equality of $C_N(\bar{x},\bar{y})$ and $\operatorname{conv} C_L(\bar{x},\bar{y})$ also under another assumption.

THEOREM 2.5 *Let $I \in \mathcal{I}(\bar{x},\bar{y})$ be an arbitrary index set for which all of the following systems with $j \in I\backslash I^0(\bar{x},\bar{y})$ have a feasible solution:*

$$\begin{aligned} Xd &= 0 \\ \nabla g_i(\bar{x},\bar{y})(d,r)^T &= 0 \quad \forall i \in I(\bar{x},\bar{y})\backslash\{j\} \\ \nabla g_j(\bar{x},\bar{y})(d,r)^T &= -1 \end{aligned} \tag{22}$$

Then, $\operatorname{cone} C_L(\bar{x},\bar{y}) = C_N(\bar{x},\bar{y})$.

Proof Obviously, $\operatorname{cone} C_L(\bar{x},\bar{y}) \subseteq C_N(\bar{x},\bar{y})$. Hence, let $(\bar{d},\bar{r}) \in C_N(\bar{x},\bar{y})$ be an arbitrary element. Then, due to $\nabla g_j(\bar{x},\bar{y})(\bar{d},\bar{r}) \leq 0$ and the assumption all the systems

$$\begin{aligned} Xd &= 0 \\ \nabla g_i(\bar{x},\bar{y})(d,r)^T &= 0 \quad \forall i \in I(\bar{x},\bar{y})\backslash\{j\} \\ \nabla g_j(\bar{x},\bar{y})(d,r)^T - \nabla g_j(\bar{x},\bar{y})(\bar{d},\bar{r})^T &= 0 \end{aligned} \tag{23}$$

have a solution for $j \in \left(I\backslash I^0(\bar{x},\bar{y})\right) \subseteq I(\bar{x},\bar{y})$. Let (d^j,r^j) denote an arbitrary such solution. Denote

$$(\tilde{d},\tilde{r}) = \sum_{j \in I\backslash I^0(\bar{x},\bar{y})} (d^j,r^j).$$

Then,

$$\begin{aligned} X(\bar{d}-\tilde{d}) &= 0 \\ \nabla g_i(\bar{x},\bar{y})(\bar{d}-\tilde{d},\bar{r}-\tilde{r}) &\leq 0 \quad \forall i \in I(\bar{x},\bar{y})\backslash I \\ \nabla g_i(\bar{x},\bar{y})(\bar{d}-\tilde{d},\bar{r}-\tilde{r}) &= 0 \quad \forall i \in I, \end{aligned} \tag{24}$$

implying that $(d^I,r^I) = (\bar{d}-\tilde{d},\bar{r}-\tilde{r}) \in C_L(\bar{x},\bar{y})$. Specifically, (d^I,r^I) belongs to the subcone of $C_L(\bar{x},\bar{y})$ for the fixed set I of the assumption of the theorem. Hence, the point

$$(\bar{d},\bar{r}) = (d^I,r^I) + \sum_{j \in I\backslash I^0(\bar{x},\bar{y})} (d^j,r^j)$$

equals the sum of elements of the cone $C_L(\bar{x},\bar{y})$ and hence is also an element of it.

COROLLARY 2.3 *If $I^0(\bar{x},\bar{y}) \in \mathcal{I}(\bar{x},\bar{y})$ then the assumption of the theorem is satisfied.*

2.4. *The case of a fixed constraint set in the lower level*

We will now consider another approach which still using the contingent cone. This approach will now be based on the graphical derivative of the solution set mapping of the lower-level problem. We will however consider the bilevel programming problem in a much simpler form. To begin with we will consider $X = \mathbb{R}^n$ and the lower-level problem as follows

$$\Psi(x) = \operatorname*{argmin}_y \big\{ f(x,y) \colon y \in K(x) \big\}, \tag{25}$$

i.e., we will use the abbreviation $K(x) = \{y \colon g(x,y) \leq 0\}$ in what follows and will not refer to the explicit description of the constraint set as solution set of inequalities. We intend to approach the optimality conditions for the bilevel programming problem by relating it to the following single-level optimization problem:

$$\min_{x,y} \{F(x,y) \colon (x,y) \in \operatorname{gph}\Psi\}. \tag{26}$$

For the relations between both problems we refer to Remark 1.1, which similarly applies to these problems.

Since the lower-level problem is convex one can express the solution set mapping as a solution set of a variational inequality problem as follows:

$$\Psi(x) = \big\{ y \in \mathbb{R}^m \colon 0 \in \nabla_y f(x,y) + N_{K(x)}(y), y \in K(x) \big\},$$

provided that Slater's condition is satisfied for the set $K(x)$. Since it is traditional in convex analysis to define $N_{K(x)}(y) = \emptyset$ if $y \notin K(x)$ one can equivalently write $\Psi(x)$ as

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

Recently there has been a considerable progress in the understanding of the solution set mapping of variational systems, see for example, [8,11]. Advances have been made in computing the set-valued derivative and the coderivative of the solution set mapping of a variational system and also sufficient conditions have been developed under which the solution set mapping is local Lipschitz around a given point. In the following, we will use the results of Dontchev and Rockafellar [8] to deduce an optimality condition for the bilevel programming problem under consideration. We begin by introducing the notion of the graphical derivative of a set-valued mapping. This is a generalization of the classical notion that the derivative is the slope of the tangent.

Let $\Psi \colon \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ be a set-valued mapping and $(\bar{x}, \bar{y}) \in \operatorname{gph}\Psi$. The graphical derivative of Ψ at $\bar{x}$ for any $\bar{y} \in \Psi(\bar{x})$ is the set-valued mapping $D\Psi(\bar{x}|\bar{y}) \colon \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ such that

$$z \in D\Psi(\bar{x}|\bar{y})(w) \Longleftrightarrow (w,z) \in C_{\operatorname{gph}\Psi}(\bar{x}, \bar{y}).$$

This means that

$$\operatorname{gph} D\Psi(\bar{x}|\bar{y}) = C_{\operatorname{gph}\Psi}(\bar{x}, \bar{y}).$$

Further the graphical derivative can also be realized as an outer limit (of sets) in the following manner

$$D\Psi(\bar{x}|\bar{y})(w) = \limsup_{\tau\downarrow 0,\, w'\to w} \frac{\Psi(\bar{x}+\tau w')-\bar{y}}{\tau}. \tag{27}$$

The notion that will play an important role in our study is the *protoderivative* of a set-valued mapping. A set-valued mapping Ψ is said to be protodifferentiable at $(\bar{x}, \bar{y})$ if the lim sup in (27) is actually realized as a limit, i.e.,

$$D\Psi(\bar{x}|\bar{y})(w) = \lim_{\tau\downarrow 0,\, w'\to w} \frac{\Psi(\bar{x}+\tau w')-\bar{y}}{\tau}.$$

We now have the following result. However, we will introduce the following notation. For a vector y by $y^{\perp}$ we mean the set $y^{\perp} = \{u\colon \langle y, u\rangle = 0\}$.

THEOREM 2.6 *Let $(\bar{x}, \bar{y})$ be a local optimistic solution of the bilevel programming problem* (26) *where $\Psi(x)$ is given by* (25). *Let $K(x) = K$ for all x in this setting and assume that K is a polyhedral set. Further assume that the solution set mapping Ψ is upper-semicontinuous. Moreover also assume that the following qualification condition also holds*:

$$\operatorname{rank}\Big(\nabla^2_{xy} f(x, y)\Big) = m \quad (\textit{full rank}). \tag{28}$$

Then one has

$$\Big\langle \nabla F(\bar{x}, \bar{y}), (u, v)\Big\rangle \geq 0$$

for all $(u, v) \in \mathbb{R}^n \times \mathbb{R}^m$ satisfying

$$0 \in \nabla^2_{xy} f(\bar{x}, \bar{y})u + \nabla^2_{xx} f(\bar{x}, \bar{y})v + N_{K_*}(v).$$

where $K_* = C_K(\bar{y}) \cap (\nabla_y f(\bar{x}, \bar{y}))^{\perp}$.

Proof Since $(\bar{x}, \bar{y})$ is a local optimistic solution and the solution set mapping Ψ is upper-semicontinuous and bounded then one can show that $(\bar{x}, \bar{y})$ also solves the following single-level optimization problem

$$\min_{x, y}\Big\{F(x, y)\colon (x, y) \in \operatorname{gph}\Psi\Big\}$$

locally [9]. Thus, one has

$$\langle F(\bar{x}, \bar{y}), (u, v)\rangle \geq 0 \quad \forall (u, v) \in C_{\operatorname{gph}\Psi}(\bar{x}, \bar{y}).$$

Also observe that Ψ can be described as

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

Since K is polyhedral and the qualification condition (28) hold from Theorem 7.1 in [8], we have that the solution set mapping Ψ is protodifferentiable and the protoderivative is given as follows:

$$D\Psi(\bar{x}|\bar{y})(u) = \left\{ v \in \mathbb{R}^m : 0 \in \nabla^2_{xy} f(\bar{x}, \bar{y})u + \nabla^2_{yy} f(\bar{x}, \bar{y})v + N_{K_*}(v) \right\},$$

where $K_* = C_K(\bar{y}) \cap (\nabla_y f(\bar{x}, \bar{y}))^{\perp}$. The result is established by noting that $\operatorname{gph} D\Psi(\bar{x}|\bar{y}) = C_{\operatorname{gph}\Psi}(\bar{x}, \bar{y})$.

Remark 2.5 The qualification condition considered in the above theorem is called the ample parametrization condition in [8]. Of course one can ask whether the polyhedrality of the set K is at all required. In the context of the above theorem it seems to be an essential requirement. In fact as seen from Theorem 6.1 in [8] we need that the normal cone mapping N_K is protodifferentiable at $(\bar{y}, \bar{z}) \in \operatorname{gph} N_K$, $\bar{z} = -\nabla_y f(\bar{x}, \bar{y})$. However if K is fully amenable (see [20] for details) then N_K is automatically protodifferentiable. Further every polyhedral set is fully amenable. Moreover, the polyhedrality of K is crucial for the appearance of the term $N_{K_*}(v)$ in the above theorem. In absence of polyhedrality this term would be replaced by the protoderivative of N_K at $(\bar{y}, \bar{z}) \in \operatorname{gph} N_K$, $\bar{z} = -\nabla_y f(\bar{x}, \bar{y})$. Polyhedrality reduces the graphical derivative of the normal cone mapping N_K at $(\bar{y}, \bar{z}) \in \operatorname{gph} N_K$, $\bar{z} = -\nabla_y f(\bar{x}, \bar{y})$ to $N_{K_*}(v)$.

It is interesting to observe that if the polyhedral set K is given as follows

$$K = \left\{ y \in \mathbb{R}^m : \langle a_i, y \rangle \leq b_i,\ i = 1, \ldots, p,\ y \geq 0 \right\}, \tag{29}$$

where $a_i \in \mathbb{R}^m$ for $i = 1, \ldots, p$ and $b_i \in \mathbb{R}$ for $i = 1, \ldots, p$, then the cone $N_{K_*}(v)$ is given as follows:

$$N_{K_*}(v) = \operatorname{cone}\left\{ a_i : i \in I(\bar{y}) \right\} + \operatorname{span}\left\{ \nabla_y f(\bar{x}, \bar{y}) \right\}.$$

In the preceding expression $I(\bar{y})$ denotes the set of active indices at $\bar{y}$.

Further consider K to be a non-empty and full-dimensional polyhedral set given by (29). Then it is clear that K has a non-empty interior. If we now assume that the solution set mapping Ψ is locally bounded then from Remark 1.1 it is clear that Ψ is upper-semicontinuous. Note that the fact that K has a nonempty interior is equivalent to the fact that Slater's condition holds. On the other hand, if we assume that K is a bounded polyhedral set, then we can just assume Ψ to be locally bounded in the statement of the above theorem. This is due to the fact that, if K is a bounded polyhedral set, then Ψ has a closed graph, and thus, combined with the fact that it is locally bounded, we conclude that Ψ is upper-semicontinuous.

3. Optimality conditions using the coderivative

However, it is clear that the result in the preceding theorem is not very compact in the sense that it is not possible to detect immediately the Lagrange multipliers associated with the optimality conditions. However, for an explicit representation in terms of Lagrange multipliers we need to take a different approach. To do so we need the following notions of set-valued coderivatives of a set-valued mapping. Levy and Mordukhovich [11] define the adjoint mapping to the graphical derivative of a set-valued mapping. Let $F\colon \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ be a set-valued mapping and let $\bar{y} \in F(\bar{x})$ and let $DF(\bar{x}|\bar{y})$ denote the graphical derivative of F at $(\bar{x}, \bar{y})$. Then the adjoint mapping at $(\bar{x}, \bar{y})$ is a set-valued mapping $\hat{D}^*F(\bar{x}|\bar{y}) : \mathbb{R}^m \rightrightarrows \mathbb{R}^n$ given as

$$\hat{D}^*F(\bar{x}|\bar{y})(y^*) = \left\{x^* \in \mathbb{R}^m \colon \langle x^*, u\rangle \leq \langle y^*, v\rangle,\ \forall (u, v) \in \operatorname{gph} DF(\bar{x}|\bar{y})\right\}.$$

The coderivative of F at $(\bar{x}, \bar{y})$ is a set-valued mapping $D^*F(\bar{x}|\bar{y})\colon \mathbb{R}^m \rightrightarrows \mathbb{R}^n$ given as

$$D^*F(\bar{x}|\bar{y})(y^*) = \limsup_{(x,y)\to(\bar{x},\bar{y}),\, y'\to y^*} \hat{D}^*F(x|y)(y'). \tag{30}$$

The notion of coderivative plays a very important role in variational analysis and optimization. This notion was first introduced by Mordukhovich [14] in 1976 and later studied in detail for example in [15,16,20]. The coderivative can also be described in terms of the *basic normal cone* (for details see, e.g. [15]). The coderivative will play a pivotal role in establishing the following result.

THEOREM 3.1 *Let $(\bar{x}, \bar{y})$ be a local optimistic solution of the bilevel programming problem* (26), *where $\Psi(x)$ is given as in* (25) *with $K(x) = K$ for all x. Assume that the solution set mapping Ψ is upper-semicontinuous. Let the qualification condition* (28) *be satisfied. 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) $0 \in \nabla_y F(\bar{x}, \bar{y}) + \nabla^2_{yy} f(\bar{x}, \bar{y})v^* + D^*N_K(\bar{y}| - \nabla_y f(\bar{x}, \bar{y}))(v^*)$.

Proof Since $(\bar{x}, \bar{y})$ is a local optimistic solution and the solution set mapping Ψ is upper-semicontinuous then $(\bar{x}, \bar{y})$ also solves the following single-level optimization problem

$$\min_{x,y}\left\{F(x, y)\colon (x, y) \in \operatorname{gph} \Psi\right\}.$$

locally [9]. Thus, one has

$$\left\langle F(\bar{x}, \bar{y}), (u, v)\right\rangle \geq 0 \quad \forall (u, v) \in C_{\operatorname{gph}\Psi}(\bar{x}, \bar{y}).$$

Thus, we have

$$-\left(\nabla_x F(\bar{x}, \bar{y}), \nabla_y F(\bar{x}, \bar{y})\right) \in \left(C_{\operatorname{gph}\Psi}(\bar{x}, \bar{y})\right)^\circ,$$

where $(C_{\operatorname{gph}\Psi}(\bar{x}, \bar{y}))^\circ$ denotes the polar cone to $C_{\operatorname{gph}\Psi}(\bar{x}, \bar{y})$. It is easy to observe that

$$(x^*, y^*) \in (C_{\operatorname{gph}\Psi}(\bar{x}, \bar{y}))^\circ \quad \text{if and only if } (-y^*, x^*) \in \operatorname{gph} \hat{D}^*\Psi(\bar{x}|\bar{y}).$$

This shows that

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

However, from (30) we see that $\hat{D}^*\Psi(\bar{x}|\bar{y})(y^*) \subseteq D^*\Psi(\bar{x}|\bar{y})(y^*)$. Thus, we have

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

Since the qualification condition in equation (28) by using Theorem 2.1 in [11], 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

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

Hence the result.

Remark 3.1 It is interesting to ask whether one can make a direct computation of $(C_{\mathrm{gph}\Psi}(\bar{x},\bar{y}))^\circ$ at least in the case when K is polyhedral. Suppose we need to find $(x^*,y^*) \in (C_{\mathrm{gph}\Psi}(\bar{x},\bar{y}))^\circ$. This will lead to the following set-valued optimization problem:

$$\min \phi(u,v) \quad \text{subject to } 0 \in G(u,v),$$

where

$$\phi(u,v) = -\langle (x^*,y^*),(u,v)\rangle.$$

and

$$G(u,v) = \nabla^2_{xy} f(\bar{x},\bar{y})u + \nabla^2_{yy} f(\bar{x},\bar{y})v + N_{K^*}(v).$$

It will be interesting to see if one can actually compute (x^*,y^*) by using the methods of set-valued optimization. For more details on set-valued optimization, see, for example [10] and the references therein.

An important question that arises is whether the optimality conditions for bilevel programming that are developed here are amenable to the development of numerical algorithms for solving bilevel programming problems. However, the optimality conditions that are developed here have an abstract term in their formulation in terms of the coderivative of the normal cone mapping to the set K. Except this term, the other terms that appear in the optimality conditions are computable. Thus, when K is expressed in terms of say convex inequality constraints it is important to know if it is possible to express the coderivative of the normal cone mapping in terms of the first and second derivatives of the constraint functions. This remains to be an open question. Let $F\colon \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ be a set-valued mapping. Then the coderivative of F at $(\bar{x},\bar{y}) \in \mathrm{gph}\, F$ is equivalently given as follows:

$$D^*F(\bar{x}|\bar{y})(y^*) = \left\{x^* \in \mathbb{R}^n : (-x^*,y^*) \in N^L_{\mathrm{gph}F}(\bar{x},\bar{y})\right\}.$$

In the preceding expression the term $N^L_{\text{gph}F}(\bar{x}, \bar{y})$ is the basic normal cone or the limiting normal cone of Mordukhovich. For more details, see [15,16,20]. Thus, in order to compute $D^*N_K(\bar{y}| - \nabla_y f(\bar{x}, \bar{y}))(v^*)$ one needs to compute $N^L_{\text{gph}N_K}(\bar{y}, \bar{z})$, $\bar{z} = -\nabla_y f(\bar{x}, \bar{y})$. The computation of this object appears to be very difficult. It is important to note that even though K is a convex set the set $\text{gph}N_K$ is not a convex set. It appears that it is quiet difficult to compute the limiting normal cone for a closed and convex set K. When $K = \mathbb{R}^m_+$ an explicit calculation of $N^L_{\text{gph}N_K}(\bar{y}, \bar{z})$ is given in Ye [22]. When $(x, y) \in \mathbb{R}^2$ and $K = [0, 1]$ an explicit calculation of $N^L_{\text{gph}N_K}(\bar{y}, \bar{z})$ is given in [9]. Thus, a more practical formulation of optimality conditions for bilevel programming remains to be a major issue of future research.

Moreover, for simplicity of presentation we have considered the case where the feasible set in the lower-level problem is independent of the parameter x. However, this need not be the case. For the case where K depends on x [9] for the optimistic case. Further, we have also made the upper-level variable free in our exposition. However, even if x is constrained to belong to a proper closed subset X of $\mathbb{R}^n$, one can still formulate a necessary optimality condition using the results in [18]. This has been demonstrated in [9]. Thus, it appears as for now that a lot of more research effort has to go into the study of optimality conditions for bilevel programming in order to develop conditions that are computationally tractable. However, this does not mean optimality conditions for bilevel programming developed by several researchers till now are of no use. Instead, these optimality conditions would act as a guide for future development since they allow us to pinpoint the exact issue that needs to looked into in greater detail in order to formulate more practical optimality conditions.

Acknowledgement

The authors would like to thank the anonymous referee whose constructive suggestions have vastly improved the presentation of the article.

References

[1] Babahadda, H. and Gadhi, N., 2005, Necessary optimality conditions for bilevel optimization problems using convexificators, Technical Report Department of Mathematics, Dhar El Mehrez, Sidi Mohamed Ben Abdellah University, Fes, Morocco.

[2] Bank, B., Guddat, J., Klatte, D., Kummer, B. and Tammer, K., 1982, *Non-linear Parametric Optimization* (Berlin: Akademie-Verlag).

[3] Bard, J.F., 1998, *Practical Bilevel Optimization: Algorithms and Applications* (Dordrecht: Kluwer Academic Publishers).

[4] Dempe, S., 2002, *Foundations of Bilevel Programming* (Dordrecht: Kluwer Academic Publishers).

[5] Dempe, S., 2003, Annotated bibliography on bilevel programming and mathematical programs with equilibrium constraints. *Optimization*, **52**, 333–359.

[6] Dempe, S. and Schmidt, H., 1996, On an algorithm solving two-level programming problems with non-unique lower level solutions. *Computational Optimization and Applications*, **6**, 227–249.

[7] Dempe, S. and Lohse, S., 2006, Inverse linear programming. In: A. Seeger (Ed.) *Recent Advances in Optimization*, Vol. 563, Lectures Notes in Economics and Mathematical Systems (Berlin: Springer), pp. 19–28.

[8] Dontchev, A. and Rockafellar, R.T., 2001, Ample parametrization of variational inclusions. *SIAM Journal on Optimization*, **12**, 170–187.

[9] Dutta, J. and Dempe, S., 2005, *Bilevel Programming with Convex Lower Level Problems*, Working Paper, TU Bergakademie Freiberg, Department of Mathematics and Computer Science.

[10] Jahn, J., 2004, *Vector Optimization* (Berlin: Springer).

[11] Levy, A.B. and Mordukhovich, B.S., 2004, Coderivatives in parametric optimization. *Mathematical Programming, Series A*, **99**, 311–327.

[12] Loridan, P. and Morgan, J., 1996, Weak via strong Stackelberg problem: new results. *Journal of Global Optimization*, **8**, 263–287.

[13] Luo, Z.-Q., Pang, J.-S. and Ralph, D., 1996, *Mathematical Programs with Equilibrium Constraints* (Cambridge: Cambridge University Press).

[14] Mordukhovich, B.S., 1976, Maximum principle in the problem of time optimal response with nonsmooth constraints. *Journal of Applied Mathematics and Mechanics*, **40**, 960–969.

[15] Mordukhovich, B.S., 1994, Generalized differential calculus for nonsmooth and set-valued mappings. *Journal of Mathematical Analysis and Applications*, **183**, 250–288.

[16] Mordukhovich, B.S., 2005, *Variational Analysis and Generalized Differentiaition, I: Basic Theory, II: Applications* (Berlin: Springer).

[17] Outrata, J., Kočvara, M. and Zowe, J., 1998, *Nonsmooth Approach to Optimization Problems with Equilibrium Constraints* (Dordrecht: Kluwer Academic Publishers).

[18] Outrata, J., 2000, A generalized mathematical program with equilibrium constraints. *SIAM Journal on Optimization*, **30**, 1623–1638.

[19] Pang, J.-S. and Fukushima, M., 1999, Complementarity constraint qualifications and simplified B-stationarity conditions for mathematical programs with equilibrium constraints. *Computational Optimization and Applications*, **13**, 111–136.

[20] Rockafellar, R.T. and Wets, R.J.B., 1998, *Variational Analysis* (Berlin: Springer).

[21] Scheel, H. and Scholtes, S., 2000, Mathematical programs with equilibrium constraints: stationarity, optimality, and sensitivity. *Mathematics of Operations Research*, **25**, 1–22.

[22] Ye, J.J., 2000, Constraint qualifications and necessary optimality conditions for optimization problems with variational inequality constraints. *SIAM Journal on Optimization*, **10**, 943–962.

[23] Ye, J.J., 2004, Nondifferentiable multiplier rules for optimization and bilevel optimization problems. *SIAM Journal on Optimization*, **15**, 252–274.

[24] Ye, J.J., 2005, Necessary and sufficient optimality conditions for mathematical programs with equilibrium constraints. *Journal of Mathematical Analysis and Applications*, **307**, 350–369.

[25] Ye, J.J. and Ye, X.Y., 1997, Necessary optimality conditions for optimization problems with variational inequality constraints. *Mathematics of Operations Research*, **22**, 977–997.

[26] Ye, J.J. and Zhu, D.L., 1995, Optimality conditions for bilevel programming problems. *Optimization*, **33**, 9–27.