3-1-2007

Variational Analysis in Bilevel Programming

S Dempe
Technical University Bergakademie Freiberg, Germany, dempe@tu-freiberg.de

J Dutta
Indian Institute of Technology, Kanpur, India, jdutta@iitk.ac.in

Boris S. Mordukhovich
Wayne State University, boris@math.wayne.edu

VARIATIONAL ANALYSIS IN BILEVEL PROGRAMMING

S. DEMPE, J. DUTTA and B. S. MORDUKHOVICH

WAYNE STATE UNIVERSITY

Detroit, MI 48202

Department of Mathematics
Research Report

2007 Series
#3

This research was partly supported by the National Science Foundation and the Australian Research Council

Variational Analysis in Bilevel Programming

S. Dempe[1], J. Dutta[2] and B. S. Mordukhovich[3]

Dedicated to the memory of Professor S. R. Mohan

Abstract. The paper is devoted to applications of advanced tools of modern variational analysis and generalized differentiation to problems of optimistic bilevel programming. In this way, new necessary optimality conditions are derived for two major classes of bilevel programs: those with partially convex and with fully convex lower-level problems. We provide detailed discussions of the results obtained and their relationships with known results in this area.

1 Introduction

In this paper we intend to discuss the interplay of variational analysis and bilevel programming. The term *Variational Analysis* is of quiet recent origin, and most probably the monograph by Rockafellar and Wets [13] had led the popularization of the term. In modern optimization, set-valued maps play a major role. Their role shot into prominence with the advent of nonsmooth analysis and nonsmooth optimization, since the role of the derivative in modern optimization is taken over by set-valued maps known as subdifferentials. Apart from that set-valued maps appear, for example, in the form of solution set maps in parametric optimization and play a very fundamental role in bilevel programming; see, e.g., Dempe [3]. Further, an important role in optimization is now played by derivatives and coderivatives of set-valued maps. For more details see Rockafellar and Wets [13] and the very recent two-volume monograph by Mordukhovich [9], [10].

On the other hand, bilevel programming grew out of the now classical Stackelberg games (see [16]) where a leader and a follower interact so that both can achieve their targeted objectives. In the language of optimization this can be framed as a two-level optimization problem as follows:

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

where $F : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$, $X \subseteq \mathbb{R}^n$, and where $S : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ is the solution set mapping to the lower-level problem:

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

[1]Department of Mathematics and Computer Science, Technical University Bergakademie Freiberg, Freiberg, Germany; dempe@tu-freiberg.de

[2]Department of Mathematics, Indian Institute of Technology, Kanpur, India; jdutta@iitk.ac.in

[3]Department of Mathematics, Wayne State University, Detroit, USA; boris@math.wayne.edu. Research of this author was partly supported by the US National Science Foundation under grants DMS-0304989 and DMS-0603846 and by the Australian Research Council under grant DP-0451168

where $f : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ and $K(x)$ is a closed set for each x. We denote the above optimization problem by (BP). So the idea is that the upper-level decision maker, or the leader, chooses a decision vector x and passes it onto the lower-level decision maker, or the follower, who then—based on the leader's choice x—minimizes his/her objective function and returns the solution y to the leader who then uses it to minimizes his objective function.

If for each x the lower-level problem has a unique solution, then the problem (BP) is well defined. However, if there are multiple solutions to the lower-level problem for a given x, then the upper-level objective becomes a set-valued map. In order to overcome this difficulty, two different solution concepts have been defined in the literature. These are namely the optimistic solution and the pessimistic solution.

For the optimistic case one first defines the function

$$\phi_0(x) := \inf_y \{F(x, y) : y \in S(x)\}.$$

Then the optimistic problem is:

$$\min \phi_0(x) \quad \text{subject to} \quad x \in X. \tag{1}$$

Thus a pair of points $(\bar{x}, \bar{y})$ is said to be an optimistic solution to the bilevel problem (BP) if $\phi_0(\bar{x}) = F(\bar{x}, \bar{y})$ and $\bar{x}$ is the optimal solution (local or global) to (1). On the other hand, in the pessimistic case we define the function

$$\phi_p(x) := \sup_y \{F(x, y) : y \in S(x)\}$$

and formulate the pessimistic problem as follows:

$$\min \phi_p(x) \quad \text{subject to} \quad x \in X.$$

In this paper we concentrate on the optimistic bilevel programming problem. An important situation where an optimistic bilevel formulation can be used is, e.g., that between a supplier and a store owner of some commodities. Since both want to do well in their businesses, the supplier will always give his/her best output to the store owner who in turn would like to do his/her best in the business. In some sense, both would like to minimize their loss or rather maximize their profit and thus act in the optimistic pattern. It is clear that in this example the store owner is the upper-level decision maker and the supplier is the lower-level decision maker. Thus in the study of supply chain management the optimistic bilevel problem can indeed play a fundamental role.

As it has been seen in Dutta and Dempe [5] and Dempe, Dutta and Mordukhovich [4], in studying the optimistic formulation of the bilevel programming problem it is useful to concentrate on the following problem (BPO):

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

If we consider global optimal solutions, then (BPO) is equivalent to the optimistic formulation of the bilevel problem (BP). This relationship is slightly more subtle when we consider local optimistic solutions. If the solution set map is uniformly bounded around the optimistic solution of the problem (BP), then the optimistic solution is a local minimum for problem (BPO). The converse however need not be true. Hence we will concentrate our efforts to analyze the local optimal points of problem (BPO).

A major bottleneck in developing necessary optimality conditions for bilevel programs is that most of the standard constraint qualifications (like, e.g., the Mangasarian-Fromovitz constraint qualification or the Abadie constraint qualification) are never satisfied for bilevel programs; see, e. g., [15]. This problem comes to light when the lower-level problem is replaced by its corresponding Karush-Kuhn-Tucker (KKT) conditions. This approach of replacing the lower-level problem by KKT conditions seems to be rather adequate if the lower-level problem is convex in the variable y and satisfies some regularity conditions; see Dutta and Dempe [5] for more detailed discussions. The presence of the complementarity slackness condition actually brings forth this violation of constraint qualifications; see, e.g., Dempe [3]. Thus various approaches have been used to develop necessary optimality conditions in bilevel programming. The reader may consult the book by Dempe [3] and the references therein for various necessary optimality conditions in bilevel programming. Let us mention that the approach in Dempe [3] requires an explicit representation of the feasible set of the lower-level problems via equality and inequality constraints. Dutta and Dempe [5] consider the case when the lower-level feasible sets are not explicitly expressed via functional constraints but are convex sets depending on the parameter x, and the lower-level objective function is convex in y for each x. In this setting, for smooth functions F and no constraint situation $X = \mathbb{R}^n$, necessary optimality conditions are expressed as

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

where $(\bar{x}, \bar{y})$ is a locally optimal solution of (BPO) and $N_{\mathrm{gph} S}(\bar{x}, \bar{y})$ is the basic/Mordukhovich normal cone to the graph of the solution set map S at the point $(\bar{x}, \bar{y})$; see Section 2. We can now shift our attention to variational analysis, since in order to develop necessary optimality conditions, we need to focus on calculating the basic normal cone in the above expression when the lower-level feasible set is explicitly defined, and also to see under what qualification conditions such a computation is possible. Thus the approach in Dutta and Dempe [5] brings forth the fundamental role that variational analysis plays in bilevel programming. Our aim here is to present the state-of-the-art on the role of variational analysis in bilevel programming.

This paper is planned is follows. In Section 2 we present some basic tools and facts from variational analysis, which are widely used in the sequel. In Section 3, which is one of the main sections of this paper, we aim to study bilevel programming problems with partially convex lower-level problems. The computation of the coderivative of the solution set map plays a major rôle in the analysis of the optimality conditions. This has been shown in [5], where results of coderivative computations from Levy and Mordukhovich [6] have been used.

We begin Section 3 with the explicit computation of the normal cone to the graph of a set-valued map defined as a solution set to a certain generalized variational inequality. Using this, we derive necessary optimality conditions for bilevel programs when the lower-level problem is partially convex, the feasible set does not depend on x, and $X = \mathbb{R}^n$. Then we move on to the case where X still equals $\mathbb{R}^n$ while the feasible set of the lower-level problem depends on x. At the end of this section we consider the general optimistic bilevel programming problem (BPO), where X is a proper subset of $\mathbb{R}^n$ and the lower-level feasible set depends on x. We provide examples where the qualification condition used hold and where they do not hold. It happens that the qualification conditions of Section 3 do not hold when the lower-level problem is linear. That leads us to consider the notion of partial calmness due to Ye and Zhu [18]. Then we move to Section 4, where we study the case in bilevel programming when the lower-level problem to be fully convex, which covers the case where the lower-level problem is linear. We derive necessary optimality conditions, which improve those in Section 3, at least for the fully convex lower-level problem.

2 Tools from Variational Analysis

In this section we briefly describe the basic tools of variational analysis needed in the sequel. We start with the variational geometry of constraint sets and describe various conic approximations associated with them.

Let us begin with the notion of the *regular normal cone* or the *Fréchet normal cone* at a point $\bar{x} \in C$, where C is a subset of $\mathbb{R}^n$. A vector $v \in \mathbb{R}^n$ is called a regular normal to C at $\bar{x}$ if

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

where $\lim_{x \to \bar{x}} \dfrac{o(\|x - \bar{x}\|)}{\|x - \bar{x}\|} = 0$. The collection of all regular normals to C at $\bar{x}$ is a cone denoted by $\hat{N}_C(\bar{x})$.

It is easy to show that if C is a convex set, the regular normal cone reduces to the standard normal cone of convex analysis (see, e.g., Rockafellar [14]). Though this definition of the regular normal cone might look as a natural generalization of the normal cone from the convex case to the nonconvex case, there are some serious pitfalls. One of the major drawbacks is that at points on the boundary of the set C the regular normal cone may just reduce to the trivial cone containing only the zero vector. To overcome this, a limiting procedure is employed, which leads us to the more robust notion of the *basic normal cone.*

A vector $\bar{v} \in \mathbb{R}^n$ is an element of the basic normal cone $N_C(\bar{x})$ to the set C at $\bar{x} \in C$ if there exist sequences $\{x_k\}$ with $x_k \in C$ and $x_k \to \bar{x}$ as well as $\{v_k\}$ with $v_k \to \bar{v}$ and $v_k \in \hat{N}_C(x_k)$. In a more compact form this is written as

$$N_C(\bar{x}) := \limsup_{x \to \bar{x}} \hat{N}_C(\bar{x})$$

in terms of the so-called Painlevé-Kuratowski upper/outer limit. It is important to note that the basic normal cone is closed but needs not be a convex set.

Further, when the set C is convex, it reduces to the classical normal cone of convex analysis.
Another concept, which is important for our study, is the notion of *normal regularity* of a set at a given point. The set C is said to be normally regular at $\bar{x} \in C$ if $\hat{N}_C(\bar{x}) = N_C(\bar{x})$.
Associated with the notion of the regular normal cone is the notion of the regular/Fréchet subdifferential of a function. Since in this study our functions are locally Lipschitz, we describe the regular subdifferential only for locally Lipschitz functions.

Let $f : \mathbb{R}^n \to \mathbb{R}$ be a locally Lipschitz function, and let $\bar{x} \in \mathbb{R}^n$ be given. The regular subdifferential $\hat{\partial} f(\bar{x})$ of the function f at $\bar{x}$ is given by

$$\hat{\partial} f(\bar{x}) := \{v \in \mathbb{R}^n : (v, -1) \in \hat{N}_{\text{epi}f}(\bar{x}, f(\bar{x}))\},$$

where epi f denotes the epigraph of f. The regular subdifferential also has a major drawback in the sense that there are points crucial, e.g., for optimization, where this subdifferential becomes empty. These are precisely the points where the regular normal cone to the epigraph of the function f reduces to the trivial cone containing only the zero element.

This trouble with the regular subdifferential is overcome by passing to the limit in order to obtain a more robust object called the *basic subdifferential*, which is given by

$$\partial f(\bar{x}) := \limsup_{x \to \bar{x}} \hat{\partial} f(x)$$

The above expression means that $v \in \partial f(\bar{x})$ if there exist sequences $\{v_k\}$ and $\{x_k\}$ with $x_k \in C$ such that $v_k \to v$ and $x_k \to \bar{x}$ with $v_k \in \hat{\partial} f(x_k)$. Knowing the fact that every basic normal can be realized as the limit of regular normals, we have the equivalent representation of the basic subdifferential:

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

The basic normal cone and the basic subdifferential were first introduced by Mordukhovich [8] in 1976. For more details see Rockafellar and Wets [13] or the recent monographs of Mordukhovich [9], [10].

Set-valued maps arise naturally in optimization, and it is very important to look at their differential properties. A significant concept in this direction is the notion of coderivative by Mordukhovich (see, e.g., his book [9]). Given a set-valued map $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ and a point $(\bar{x}, \bar{y}) \in$ gph F, the coderivative of F at $(\bar{x}, \bar{y})$ is a set-valued map $D^*F(\bar{x}, \bar{y}) : \mathbb{R}^m \rightrightarrows \mathbb{R}^n$ defined by

$$D^*F(\bar{x}, \bar{y})(y^*) := \{x^* \in \mathbb{R}^n : (x^*, -y^*) \in N_{\text{gph}F}(\bar{x}, \bar{y})\}.$$

We now consider the following optimization problem (P):

$$\min f_0(x) \quad \text{subject to} \quad F(x) \in U, \quad x \in X, \tag{2}$$

where $f_0 : \mathbb{R}^n \to \mathbb{R}$ and $F : \mathbb{R}^n \to \mathbb{R}^m$ are smooth functions, $U \subseteq \mathbb{R}^m$, and $X \subseteq \mathbb{R}^n$. The necessary optimality condition for (P) formulated in the next theorem can be found in Rockafellar and Wets [13] and Mordukhovich [9].

Theorem 2.1 *Consider problem (P) from (2), and let* $\bar{x}$ *be a local minimum to (P). Assume that the following qualification condition (Q) holds at* $\bar{x}$ *:*

$$y \in N_U(F(\bar{x})) \quad \text{with} \quad 0 \in \nabla F(\bar{x})^T y + N_X(\bar{x}) \text{ implies that } y = 0.$$

Then there exists $\bar{y} \in N_U(F(\bar{x}))$ *such that*

$$0 \in \nabla f_0(\bar{x}) + \nabla F(\bar{x})^T y + N_X(\bar{x}).$$

Using this result, we can compute the normal cone to the feasible set C, which is explicitly given in the above theorem by

$$C = \{x \in X : F(x) \in U\}. \tag{3}$$

However, the explicit computation of the normal cone can be done under certain qualification conditions, and we present the full result in the next theorem.

Theorem 2.2 *Consider the set* C *given by (3), where* $F : \mathbb{R}^n \to \mathbb{R}^m$ *is a smooth function and* X *is a closed set. Assume that the qualification condition (Q) of Theorem 2.1 holds at* $\bar{x}$*. Then one has*

$$N_C(\bar{x}) \subset \bigcup \left\{\nabla F(\bar{x})^T y + N_X(\bar{x}) : y \in N_U(F(\bar{x})\right\}.$$

Furthermore, if the set X *is normally regular at* $\bar{x}$ *and the set* U *is normally regular at* $F(\bar{x})$*, then equality holds in the above expression.*

The two theorems presented in this section play a fundamental role in the next section. We show there how to use these theorems to derive necessary optimality conditions for bilevel programs with partially convex lower-level problems.

3 Partially convex lower-level problems

In this section we consider partially convex lower-level problems in the bilevel programs (BPO) of our study. By a partially convex lower-level problem we mean that $y \mapsto f(x, y)$ is convex in y for each $x \in X$ and the set $K(x)$ is convex for each x. For simplicity of the presentation we assume the upper-level objective function to be smooth, i.e., with its data to be continuously differentiable. Furthermore, we assume that the lower-level objective function is twice continuously differentiable.
Our first step is to provide an explicit computation of the basic normal cone to the graph of a set-valued map defined as a solution set of a generalized variational inequality. This will play a fundamental role in the subsequent study, since—as we have discussed in Section 1—deriving necessary optimality condition for optimistic bilevel programming is based on computing the normal cone to the solution set of the lower-level problem. Let us begin with considering a set-valued map $S : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ given by

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

where $G : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^d$ is a smooth single-valued map and $M : \mathbb{R}^n \times \mathbb{R}^m \rightrightarrows \mathbb{R}^d$ is a set-valued map of closed graph. We first concern a more simpler version, where the set-valued map does not depend on x, i.e., $M(x,y) = M(y)$. Thus we concentrate on calculating the coderivative of the set-valued map $S : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ defined above. This is done through the following result.

Theorem 3.1 *Consider $S : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ given by*

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

where $G : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^d$ is smooth map and $M : \mathbb{R}^m \rightrightarrows \mathbb{R}^d$ is closed-graph. Taking $(\bar{x}, \bar{y}) \in gph\ S$, impose the qualification condition:

$$-\nabla_x G(\bar{x},\bar{y})^T z = 0 \quad \text{and} \quad w - \nabla_y G(\bar{x},\bar{y})^T z = 0$$

with $(w,z) \in N_{gph\ M}(\bar{y}, -G(\bar{x},\bar{y}))$ implies that $w = 0$ and $z = 0$. Then one has

$$N_{gph\ S}(\bar{x},\bar{y}) \subseteq \{(x^*,y^*) \in \mathbb{R}^n \times \mathbb{R}^m : x^* = -\nabla_x G(\bar{x},\bar{y})^T \bar{z},$$
$$y^* = \bar{w} - \nabla_y G(\bar{x},\bar{y})^T \bar{z}, (\bar{w},\bar{z}) \in N_{gph\ M}(\bar{y}, -G(\bar{x},\bar{y}))\}.$$

Equality holds in the above expression if any of the following two additional conditions are satisfied:

i) The graphical set $gph\ M$ is normally regular at $(\bar{y}, -G(\bar{x},\bar{y}))$.

ii) The matrix $\nabla_x G(\bar{x},\bar{y})$ is of full row rank.

Proof: To begin with, let us observe that the inclusion $y \in S(x)$ implies that $-G(x,y) \in M(y)$. Put then

$$F(x,y) := (y, -G(x,y))^T.$$

Thus we can equivalently rewrite $S(x)$ as

$$S(x) = \{y \in \mathbb{R}^m : F(x,y) \in \text{gph} M\},$$

which means that

$$\text{gph} S = \{(x,y) : F(x,y) \in \text{gph} M\}.$$

Observe that the qualification condition imposed in the theorem can be equivalently written as

$$\left[\nabla F(\bar{x},\bar{y})^T (w,z) = 0, \quad (w,z) \in N_{\text{gph}\ M}(F(\bar{x},\bar{y})\right] \Longrightarrow [w = 0, \quad z = 0],$$

where $\nabla F(\bar{x},\bar{y})$ stands for the Jacobian of F at $(\bar{x},\bar{y})$. It is easy to see that

$$\nabla F(\bar{x},\bar{y}) = \begin{pmatrix} 0 & I \\ -\nabla_x G(\bar{x},\bar{y}) & -\nabla_y G(\bar{x},\bar{y}) \end{pmatrix}.$$

Thus we have

$$\nabla F(\bar{x},\bar{y})^T = \begin{pmatrix} 0 & -\nabla_x G(\bar{x},\bar{y})^T \\ I & -\nabla_y G(\bar{x},\bar{y})^T \end{pmatrix}.$$

The above observation allows us to apply Theorem 2.2 and conclude that

$$N_{\text{gph }S}(\bar{x},\bar{y}) \subseteq \{(x^*,y^*) \in \mathbb{R}^m \times \mathbb{R}^n : (x^*,y^*) = \nabla F(\bar{x},\bar{y})^T(\bar{w},\bar{z}), \\ (\bar{w},\bar{z}) \in N_{\text{gph}M}(F(\bar{x},\bar{y}))\}.$$

This immediately gives

$$N_{\text{gph }S}(\bar{x},\bar{y}) \subseteq \{(x^*,y^*) \in \mathbb{R}^n \times \mathbb{R}^m : x^* = -\nabla_x G(\bar{x},\bar{y})^T\bar{z}, \\ y^* = \bar{w} - \nabla_y G(\bar{x},\bar{y})^T\bar{z}, (\bar{w},\bar{z}) \in N_{\text{gph }M}(\bar{y},-G(\bar{x},\bar{y}))\}.$$

If gph M is normally regular at $(\bar{y},-G(\bar{x},\bar{y}))$, we conclude from Theorem 2.2 that the equality holds. If furthermore $\nabla_x G(\bar{x},\bar{y})$ has full row rank, then the qualification condition is automatically satisfied, and the equality follows by application of Exercise 6.7 (page 202) from Rockafellar and Wets [13]. □

Remark 3.1 The above theorem estimates the normal cone to the graph of the set-valued map S defined as the set of solutions to a generalized variational inequality. This estimate naturally allows one to provide an estimation for the coderivative of S. It is not hard to see that

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

Of course, this estimate holds under the assumptions of Theorem 3.1, with equality holding under the same conditions as in Theorem 3.1.

Let us note that the conclusions of Theorem 3.1 can be easily extended to the case when the set-valued mapping M depends on both x and y. To proceed, we need to modify the qualification conditions in order to derive the corresponding estimate, which is slightly different from the previous one due to the change in the dependence pattern of M. We present the result below with no proof.

Theorem 3.2 *Consider the set-valued map $S : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ defined by*

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

where $G : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^d$ is smooth and $M : \mathbb{R}^n \to \mathbb{R}^m \rightrightarrows \mathbb{R}^d$ is closed-graph. Given $(\bar{x},\bar{y}) \in$ gph S, assume the qualification condition:

$$u - \nabla_x G(\bar{x},\bar{y})^T z = 0, \quad \text{and} \quad w - \nabla_y G(\bar{x},\bar{y})^T z = 0$$

with $(u,w,z) \in N_{gph\ M}(\bar{x},\bar{y},-G(\bar{x},\bar{y}))$ implies that $u = 0$, $w = 0$ and $z = 0$. Then one has the inclusion

$$N_{gph\ S}(\bar{x},\bar{y}) \subseteq \{(x^*,y^*) \in \mathbb{R}^n \times \mathbb{R}^m : x^* = \bar{u} - \nabla_x G(\bar{x},\bar{y})^T\bar{z}, \\ y^* = \bar{w} - \nabla_y G(\bar{x},\bar{y})^T\bar{z}, (\bar{u},\bar{w},\bar{z}) \in N_{gph\ M}(\bar{x},\bar{y},-G(\bar{x},\bar{y}))\}.$$

Equality holds in the above expression if gphM is normally regular at $(\bar{x},\bar{y},-G(\bar{x},\bar{y})$.

Theorem 3.1 allows us to derive necessary optimality conditions for bilevel programs (BPO) with partially convex lower-level problems. It is important to observe that, since the lower-level problem is partially convex, we can equivalently represent S as

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

It is convenient in what follows to define $N_K(y)$ for all $y \in \mathbb{R}^m$ extending it to $y \notin K$ by $N_K(y) = \emptyset$.

Theorem 3.3 *Consider problem (BPO) with $X = \mathbb{R}^n$ and $K(x) = K$ for all x. Let $(\bar{x}, \bar{y}) \in gphS$ be a local optimal solution to (BPO), and let the following qualification condition hold:*

$$-(\nabla^2_{xy} f(\bar{x},\bar{y}))^T z = 0, \quad w - (\nabla^2_{yy} f(\bar{x},\bar{y}))^T z = 0$$

with $(w,z) \in N_{gphN_K}(\bar{y}, -\nabla_y f(\bar{x},\bar{y}))$ implies that $w = 0$ and $z = 0$.
Then there exists $(\bar{w}, \bar{z}) \in N_{gphN_K}(\bar{y}, -\nabla_y f(\bar{x},\bar{y}))$ 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: Since $(\bar{x},\bar{y})$ is local optimal to (BP) with $X = \mathbb{R}^n$, we have

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

which implies that

$$-(\nabla_x F(\bar{x},\bar{y}), \nabla_y F(\bar{x},\bar{y})) \in N_{\text{gph}S}(\bar{x},\bar{y}). \tag{4}$$

Now setting $G(x,y) = \nabla_y f(x,y)$ and $N_K = M$, we see that the qualification condition in the theorem is the same as in Theorem 3.1. Thus applying Theorem 3.1, we get the inclusion

$$N_{\text{gph}\,S}(\bar{x},\bar{y}) \subseteq \{(x^*,y^*) \in \mathbb{R}^n \times \mathbb{R}^m : x^* = -\nabla^2_{xy} f(\bar{x},\bar{y})^T \bar{z},$$
$$y^* = \bar{w} - \nabla^2_{yy} f(\bar{x},\bar{y})^T \bar{z}, (\bar{w},\bar{z}) \in N_{\text{gph}\,N_K}(\bar{y}, -\nabla_y f(\bar{x},\bar{y}))\}.$$

Combining the above estimate with (4), we arrive at the desired result. □

Remark 3.2 Note that the above theorem is also derived in [5] by using Theorem 3.1 from Outrata [12]. Here we give a direct proof of this result, focusing more on structural and computational issues. Observe further that the qualification condition in the above theorem holds true if we assume that $\nabla^2_{xy} f(\bar{x},\bar{y})$ is of full row rank. To illustrate this, consider the function $f(x,y) := \langle y, Ax \rangle$, where $(x,y) \in \mathbb{R}^n \times \mathbb{R}^m$ and A is a $m \times n$ matrix of full row rank m. We have $\nabla^2_{xy} f(x,y) = A$, and hence the qualification condition of the above theorem is clearly satisfied.

Next we turn to the case where the feasible set of the lower-level problem needs not to remain constant for each x, assuming nevertheless that $X = \mathbb{R}^n$. In this case, the solution set to (BPO) is given by

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

Setting $N_K(x,y) := N_{K(x)}(y)$ if $y \in K(x)$ and $N_K(x,y) := \emptyset$ otherwise, we rewrite $S(x)$ as

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

Theorem 3.4 *Consider problem (BPO), where $X = \mathbb{R}^n$ and the feasible set to the lower-level problem varies with each x. Let $(\bar{x}, \bar{y}) \in gphS$ be a local optimal solution to (BPO), and let the following qualification condition hold:*

$$u - \nabla^2_{xy} f(\bar{x}, \bar{y})^T z = 0, \quad w - \nabla^2_{yy} f(\bar{x}, \bar{y})^T z = 0$$

with $(u, w, z) \in N_{gphN_K}(\bar{x}, \bar{y}, -\nabla_y f(\bar{x}, \bar{y}))$ implies that $u = 0, w = 0, z = 0$.
Then there exists $(\bar{u}, \bar{w}, \bar{z}) \in N_{gphN_K}(\bar{x}, \bar{y}, -\nabla_y f(\bar{x}, \bar{y}))$ such that

i) $-\nabla_x F(\bar{x}, \bar{y}) = \bar{u} - \nabla^2_{xy} f(\bar{x}, \bar{y})^T \bar{z}$,

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

Proof: Since $(\bar{x}, \bar{y})$ is a local optimal solution to (BPO), we have

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

Now the result follows from Theorem 3.2 by setting $G(x,y) := \nabla_y f(x,y)$ and $N_K := M$. □.

Remark 3.3 Note that the problem of Theorem 3.4 was studied in [5, Theorem 4.1], while our approach here is different.

The next most relevant question is about necessary optimality conditions for the constrained case $x \in X$ in (BPO). Here is the result in this case.

Theorem 3.5 *Let $(\bar{x}, \bar{y}) \in gphS$ be a local optimal solution to (BPO), and let the following qualification condition be satisfied:*

$$u - \nabla^2_{xy} f(\bar{x}, \bar{y})^T z + \gamma = 0, \quad w - \nabla^2_{yy} f(\bar{x}, \bar{y})^T z = 0$$

with $(u, w, z) \in N_{gphN_K}(\bar{x}, \bar{y}, -\nabla_y f(\bar{x}, \bar{y}))$ and $\gamma \in N_X(\bar{x})$ implies the equalities $u = 0, w = 0, z = 0$.
Then there are $(\bar{u}, \bar{w}, \bar{z}) \in N_{gphN_K}(\bar{x}, \bar{y}, -\nabla_y f(\bar{x}, \bar{y})$ and $\bar{\gamma} \in N_X(\bar{x})$ such that

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

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

Proof: Observe that problem (BPO) can be equivalently rewritten as

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

where the set C is given by

$$C = \{(x,y) \in X \times \mathbb{R}^m : H(x,y) \in \text{gph}N_K\} \text{ with } H(x,y) = (x, y, -\nabla_y f(x,y))^T.$$

It is well known that $N_{X\times\mathbb{R}^m}(\bar{x},\bar{y}) = N_X(\bar{x}) \times N_{\mathbb{R}^m}(\bar{y})$, and thus

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

Therefore, the qualification condition of the theorem is equivalent to

$$\Big[0 \in \nabla H(\bar{x},\bar{y})^T q + N_{X\times\mathbb{R}^m}(\bar{x},\bar{y}),\ q = (u,w,z) \in N_{\text{gph}N_K}(\bar{x},\bar{y},-\nabla_y f(\bar{x},\bar{y}))\Big]$$
$$\Longrightarrow \big[u = 0, w = 0, z = 0\big].$$

Observe further that

$$\nabla H(\bar{x},\bar{y})^T = \begin{pmatrix} I & 0 & -\nabla^2_{xy} f(\bar{x},\bar{y})^T \\ 0 & I & -\nabla^2_{yy} f(\bar{x},\bar{y})^T \end{pmatrix}.$$

Thus the qualification condition of this theorem reduces to the qualification condition of Theorem 2.1, and the result follows. □

Remark 3.4 We would like to note that in Dutta and Dempe [5] the optimistic bilevel programming problem with partially convex lower-level problems was not considered in its full generality as it is done in the above Theorem 3.5.

It is time to present an illustrative example for the reader convenience.

Example 3.1 Consider the optimistic bilevel programming problem in a two-dimensional setting:

$$\min_{x,y} (x-1)^2 + y^2 \quad \text{subject to} \quad x > 0, y \in S(x),$$

where S denotes the solution set mapping to the following lower-level problem:

$$\min_{y} x^2 y \quad \text{subject to} \quad y \geq 0.$$

Observe that $S(x) = \{0\}$ for all $x > 0$, and that the only solution to the above optimistic bilevel programming problem is $(1,0)$. It is clear that $\nabla^2_{xy} f(1,0) = 2$. Let us check that the qualification conditions of Theorem 3.5 is satisfied. To proceed, observe that the lower-level feasible set is $[0,+\infty)$, which is thus a convex set independent of x. Note that the vector u actually does not appear in the qualification condition of Theorem 3.5. Hence we may just set $u = 0$ throughout in this particular case. Since $X = \mathbb{R}_+$, we get $N_X(1) = \{0\}$, which easily yields $z = 0$ and $w = 0$. It is easy to check furthermore that the necessary condition of the theorem holds with $\bar{w} = 0$ and $\bar{z} = 0$. Observe finally that $\bar{y} = 0$, since $N_X(1) = \{0\}$.

One of our primary goals of this section is to highlight the fact that necessary optimality conditions for optimistic bilevel programs with partially convex lower-level problems can be basically deduced from Theorem 2.1 and Theorem 2.2, which are indeed fundamental results in optimization theory. An interesting fact that emerges here is that the second-order partial derivatives of the lower-level objective function naturally appear in the first-order optimality conditions for this class of bilevel programs. Another observation that emerges here is that the qualification conditions in the above results do not work if the lower-level problem is a linear optimization problem. This issue is addressed in the next section, where we discuss the property of partial calmness that automatically holds when the lower-level problem is linear. Further, in the next section we approach necessary optimality conditions in bilevel programs by using the idea of optimal value functions.

4 Fully convex lower-level problems

Now we investigate problem (BPO) under the assumption that both the lower-level and the upper-level objective functions are convex with respect to both x and y, and that gph K is also a convex set. Denote the optimal value function of the lower-level problem by

$$\varphi(x) := \min_y \{f(x,y) : y \in K(x)\}.$$

Then problem (BPO) is equivalent to the following problem (VPO):

$$\min_{x,y} F(x,y) \text{ subject to } f(x,y) \leq \varphi(x),\ y \in K(x),\ x \in X.$$

Usual constraint qualifications as, e.g., the Mangasarian-Fromowitz one (in its nondifferentiable version) are not satisfied at each feasible point of (VPO); see Ye and Zhu [18].

Following Ye and Zhu [18], we say that problem (VPO) is *partially calm* at a given point $(\overline{x},\overline{y})$ if there is a constant $M > 0$ and an open neighborhood D of the triple $(\overline{x},\overline{y},0)$ such that for each feasible point $(x,y,u) \in D$ of the problem

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

we have the relation

$$F(x,y) - F(\overline{x},\overline{y}) + M|u| \geq 0.$$

By [18], partial calmness is satisfied for problem (VPO) if, in particular, all optimal solutions to the lower-level problem are weak sharp minima in the sense of Burke and Ferris [2]: for fixed $\overline{x}$ there exists $\alpha > 0$ such that

$$f(\overline{x},y) \geq f(\overline{x},\overline{y}) + \alpha \ \mathrm{dist}(y, S(\overline{x})),$$

whenever $y \in K(\overline{x})$, where $\operatorname{dist}(y, S(\overline{x}))$ denotes the Euclidean distance of a point y to the set $S(\overline{x})$ and where $\overline{y} \in S(\overline{x})$. It has been shown in [2] that optimal solutions to linear programming problems are weak sharp minima (cf. also Mangasarian and Meyer [7]) whenever the problem has an optimal solution. Also, optimal solutions to quadratic programming problems are weak sharp minima provided that a certain relatively weak assumption is satisfied; see Burke and Ferris [2], Ye and Zhu [18]. Note that the assumption of partial calmness can be replaced by other assumptions; see Ye [17] for more discussions.

The main feature of partial calmness is the validity of an exact penalty function approach to problem (VPO):

Theorem 4.1 *([18, Proposition 3.2]) Let $(\overline{x}, \overline{y})$ be a local optimal solution to (VPO). Then, problem (VPO) is partially calm at $(\overline{x}, \overline{y})$ if and only if there exists $\lambda > 0$ such that $(\overline{x}, \overline{y})$ is a local optimal solution to the problem*

$$\min_{x,y} F(x,y) + \lambda(f(x,y) - \varphi(x)) \text{ subject to } y \in K(x),\ x \in X. \tag{5}$$

This is a significant tool in the proof of the next theorem, where the symbols ∂, ∂_x, ∂_y denote, respectively, the subdifferential, the partial subdifferential with respect to x and to y of convex functions in the sense of convex analysis.

Theorem 4.2 *Consider problem (VPO) under the assumptions that:*
i) $K(x) = \{y : g(x,y) \le 0\}$, $X = \mathbb{R}^n$, $g : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^p$;

ii) all functions F, f, g_i are convex on $\mathbb{R}^n \times \mathbb{R}^m$, $i = 1, \dots, p$;

iii) the point $(\overline{x}, \overline{y})$ is a local optimal solution, problem (VPO) is partially calm at $(\overline{x}, \overline{y})$, there exists a compact set C such that $\{(x,y) : g(x,y) \le 0\} \subseteq C$, and there is a point $(\widehat{x}, \widehat{y})$ with $g_i(\widehat{x}, \widehat{y}) < 0$, $i = 1, \dots, p$.

Then there exist $\lambda > 0$, λ_i, μ_i, and a point $\widetilde{y} \in S(\overline{x})$ such that the following conditions are satisfied:

$$\begin{aligned}
0 &\in \partial_x F(\overline{x}, \overline{y}) + \lambda(\partial_x f(\overline{x}, \overline{y}) - \partial_x f(\overline{x}, \widetilde{y})) + \sum_{i=1}^{p} (\mu_i \partial_x g_i(\overline{x}, \overline{y}) - \lambda \lambda_i \partial_x g_i(\overline{x}, \widetilde{y})), \\
0 &\in \partial_y F(\overline{x}, \overline{y}) + \lambda \partial_y f(\overline{x}, \overline{y}) + \sum_{i=1}^{p} \mu_i \partial_y g_i(\overline{x}, \overline{y}), \\
0 &\in \partial_y f(\overline{x}, \widetilde{y}) + \sum_{i=1}^{p} \lambda_i \partial_y g_i(\overline{x}, \widetilde{y}), \\
&\quad \lambda_i \ge 0,\ \lambda_i g_i(\overline{x}, \widetilde{y}) = 0,\ i = 1, \dots, p, \\
&\quad \mu_i \ge 0,\ \mu_i g_i(\overline{x}, \overline{y}) = 0,\ i = 1, \dots, p.
\end{aligned}$$

Proof: By our assumptions on the lower level problem, the optimal value function $\varphi(\cdot)$ is convex, and hence it is locally Lipschitzian; see [14]. Thus (VPO) is a problem of Lipschitzian programming. By partial calmness, the local optimal solution $(\overline{x}, \overline{y})$ is also a local optimal solution to the Lipschitz optimization problem (5) for some $\lambda > 0$. Applying to this problem the generalized multiplier rule from [10, Theorem 3.21 (iii)] together with the calculus rules for the basic subdifferential from [9, Theorem 2.23 (c)], we obtain the existence of multipliers $(\lambda_0, \mu_1, \ldots, \mu_p)$ such that $\lambda_0 \geq 0$ and

$$\mu_i \geq 0, \;\; \mu_i g_i(\overline{x}, \overline{y}) = 0, \;\; i = 1, \ldots, p, \tag{6}$$

$$0 \in \lambda_0 \partial F(\overline{x}, \overline{y}) + \lambda_0 \lambda (\partial f(\overline{x}, \overline{y}) - \partial_x \varphi(\overline{x}) \times \{0\}) + \sum_{i=1}^{p} \mu_i \partial g_i(\overline{x}, \overline{y}).$$

Observe that we have in fact $\lambda_0 > 0$, i.e., we can set $\lambda_0 = 1$ by the Slater-type qualification conditions assumed in the theorem. Using the important relationship between partial and full subdifferentials in convex analysis

$$\partial \theta(x, y) \subseteq \partial_x \theta(x, y) \times \partial_y \theta(x, y),$$

we obtain the inclusions

$$0 \in \partial F_x(\overline{x}, \overline{y}) + \lambda(\partial_x f(\overline{x}, \overline{y}) - \partial_x \varphi(\overline{x})) + \sum_{i=1}^{p} \mu_i \partial_x g_i(\overline{x}, \overline{y}), \tag{7}$$

$$0 \in \partial_y F(\overline{x}, \overline{y}) + \lambda \partial_y f(\overline{x}, \overline{y}) + \sum_{i=1}^{p} \mu_i \partial_y g_i(\overline{x}, \overline{y}). \tag{8}$$

By the symmetry property

$$\partial(-\varphi)(\overline{x}) \subseteq -\partial \varphi(\overline{x})$$

and the estimate

$$\partial \varphi(\overline{x}) \subseteq \bigcup_{y \in S(\overline{x})} \bigcup_{\lambda \in \Lambda(\overline{x}, y)} \{\partial_x f(\overline{x}, y) + \sum_{i=1}^{p} \lambda_i \partial_x g_i(\overline{x}, y)\}$$

given, e.g., in [11] with

$$\begin{aligned} \Lambda(\overline{x}, y) = \{\lambda_i \geq 0 : \;\; & \lambda_i g_i(\overline{x}, y) = 0, \; i = 1, \ldots, p, \\ & 0 \in \partial_y f(\overline{x}, y) + \sum_{i=1}^{p} \partial_y g_i(\overline{x}, y)\}, \end{aligned} \tag{9}$$

we transform (7) to

$$\begin{aligned} 0 \;\; & \in \partial F_x(\overline{x}, \overline{y}) + \lambda(\partial_x f(\overline{x}, \overline{y}) - (\partial_x f(\overline{x}, \widetilde{y}) \\ & + \sum_{i=1}^{p} \lambda_i \partial_x g_i(\overline{x}, \widetilde{y}))) + \sum_{i=1}^{p} \mu_i \partial_x g_i(\overline{x}, \overline{y}) \end{aligned} \tag{10}$$

for some $\widetilde{y} \in S(\overline{x})$ and $(\lambda_1, \ldots, \lambda_p) \in \Lambda(\overline{x}, \widetilde{y})$. Conditions (10), (8), (9), (6) together with $(\lambda_1, \ldots, \lambda_p) \in \Lambda(\overline{x}, \widetilde{y})$ are the desired necessary conditions, which thus completes the proof the theorem. □

Corollary 4.1 *If the compactness assumption of the theorem is replaced by the inner semicontinuity assumption that for each point $(\widehat{x}, \widehat{y}) \in$ gph S and any sequence $\{x^k\}$ with $S(x^k) \neq \emptyset$ converging to $\widehat{x}$ there is a sequence $\{y^k\}$ with $y^k \in S(x^k)$ converging to $\widehat{y}$, then we obtain by [11, Corollary 4] the inclusion*

$$\partial\varphi(\overline{x}) \subseteq \bigcup_{\lambda \in \Lambda(\overline{x}, y)} \{\partial_x f(\overline{x}, y) + \sum_{i=1}^{p} \lambda_i \partial_x g_i(\overline{x}, y)\}. \tag{11}$$

Replacing the formula for the subdifferential of φ at $\overline{x}$ in the above proof, we can take $\widetilde{y} = \overline{y}$ in the assertion of the theorem. If, moreover, the functions f, g_i, $i = 1, \ldots, p$, are continuously differentiable, the following necessary optimality conditions result from Theorem 4.2:

There exists $\lambda > 0, \lambda_i,\ \mu_i,\ i = 1, \ldots, p$, satisfying

$$\begin{aligned}
0 &\in \partial_x F(\overline{x}, \overline{y}) + \sum_{i=1}^{p} (\mu_i - \lambda\lambda_i) \nabla_x g_i(\overline{x}, \overline{y}), \\
0 &\in \partial_y F(\overline{x}, \overline{y}) + \lambda \nabla_y f(\overline{x}, \overline{y}) + \sum_{i=1}^{p} \mu_i \nabla_y g_i(\overline{x}, \overline{y}), \\
0 &\in \nabla_y f(\overline{x}, \overline{y}) + \sum_{i=1}^{p} \lambda_i \nabla_y g_i(\overline{x}, \overline{y}), \\
&\quad \lambda_i \geq 0,\ \lambda_i g_i(\overline{x}, \overline{y}) = 0,\ i = 1, \ldots, p, \\
&\quad \mu_i \geq 0,\ \mu_i g_i(\overline{x}, \overline{y}) = 0,\ i = 1, \ldots, p.
\end{aligned}$$

For a related result, obtained using different assumptions and a different method, we refer to [17, Theorem 4.1].

Optimal solutions to linear programming problems are weak sharp as shown by Burke and Ferris [2]. Moreover, the solution set map to linear programming problems of the type

$$\min c^\top y, \text{ subject to } Ay \leq x$$

with right-hand side perturbations x is lower semicontinuous by [1, Theorem 4.3.5] and hence also inner semicontinuous. This allows us to deduce the following simple necessary optimality conditions.

Corollary 4.2 *Consider the bilevel linear programming problem (VOP) with*

$$\varphi(x) = \min_y \{c^\top y : Ay \leq x\}$$

and $X = \mathbb{R}^n$. Assume for simplicity that F is continuously differentiable. If $(\overline{x}, \overline{y})$ is a local optimal solution of this problem, then there exist multipliers

$\lambda > 0, \mu \geq 0, \beta \geq 0$ *such that*

$$\begin{aligned} \nabla_x F(\overline{x},\overline{y}) - (\mu - \lambda\beta) &= 0, \\ \nabla_y F(\overline{x},\overline{y}) + \lambda c + \mu^{\mathsf{T}} A &= 0, \\ c + \beta^{\mathsf{T}} A &= 0, \\ \beta \geq 0,\ \beta^{\mathsf{T}}(A\overline{y} - \overline{x}) &= 0, \\ \mu \geq 0,\ \mu^{\mathsf{T}}(A\overline{y} - \overline{x}) &= 0. \end{aligned}$$

References

[1] B. Bank, J. Guddat, D. Klatte, B. Kummer and K. Tammer. *Non-Linear Parametric Optimization.* Akademie-Verlag, Berlin, 1982.

[2] J. V. Burke and M. C. Ferris. Weak sharp minima in mathematical programming. 31:1340–1359, 1993.

[3] S. Dempe. *Foundations of Bilevel Programming.* Kluwer Academie Publishers, Dordrecht, 2002.

[4] S. Dempe, J. Dutta and B. S. Mordukhovich. New necessary optimality conditions in optimistic bilevel programming. *Optimization*, to appear.

[5] J. Dutta and S. Dempe. Bilevel programming with convex lower level problems. In S. Dempe and V. Kalashnikov, editors, *Optimization with Multivalued Mappings: Theory, Applications and Algorithms.* Springer Science+Business Media, LLC, 2006.

[6] A. B. Levy and B. S. Mordukhovich. Coderivatives in parametric optimization. 99:311–327, 2004.

[7] O. L. Mangasarian and R. R. Meyer. Nonlinear perturbations of linear programs. 17:745–752, 1979.

[8] B. S. Mordukhovich. Maximum principle in problems of time optimal control with nonsmooth constraints. *J. Appl. Math. Mech.*, 40:960–969, 1976.

[9] B. S. Mordukhovich. *Variational Analysis and Generalized Differentiation, Vol. 1: Basic Theory.* Springer Verlag, Berlin et al., 2006.

[10] B. S. Mordukhovich. *Variational Analysis and Generalized Differentiation, Vol. 2: Applications.* Springer Verlag, Berlin et al., 2006.

[11] B. S. Mordukhovich, N. M. Nam and N. D. Yen. Subgradients of marginal functions in parametric mathematical programming. *Mathematical Programming*, to appear.

[12] J. V. Outrata. A generalized mathematical program with equilibrium constraints. *SIAM Journal on Control and Optimization*, 38:1623–1638, 2000.

[13] R. T. Rockafellar and R. J.-B. Wets. *Variational Analysis*. Springer Verlag, Berlin, 1998.

[14] R. T. Rockafellar. *Convex Analysis*. Princeton University Press, Princeton, 1970.

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

[16] H. von Stackelberg. *Marktform und Gleichgewicht*. Springer-Verlag, Berlin, 1934. Engl. transl.: The Theory of the Market Economy, Oxford University Press, 1952.

[17] J. J. Ye. Constraint qualifications and KKT conditions for bilevel programming problems. *Mathematics of Operations Research*, 31:811–824, 2007.

[18] J. J. Ye and D. L. Zhu. Optimality conditions for bilevel programming problems. *Optimization*, 33:9–27, 1995; with correction in *Optimization* 39:361-366, 1997.