Optimality conditions for bilevel programming problems

Stephan Dempe[1], Vyatcheslav V. Kalashnikov[2] and Nataliya Kalashnykova[3]

[1] Department of Mathematics and Informatics, Technical University Bergakademie Freiberg, 09596 Freiberg, Germany dempe@tu-freiberg.de
[2] Departamento de Ingeniería Industrial y de Sistemas, Centro de Calidad ITESM, Campus Monterrey, Ave. Eugenio Garza Sada 2501 Sur Monterrey, N.L., México, 64849, on leave from the Central Economics and Mathematics Institute (CEMI) of the Russian Academy of Sciences, Nakhimovsky prospect 47, Moscow 117418, Russian Federation kalash@itesm.mx
[3] Facultad de Ciencias Físico-Matemáticas, Universidad Autónoma de Nuevo León, San Nicolás de los Garza, N.L., México, 66450 nkalash@fcfm.uanl.mx

Summary. Focus in the paper is on optimality conditions for bilevel programming problems. We start with a general condition using tangent cones of the feasible set of the bilevel programming problem to derive such conditions for the optimistic bilevel problem. More precise conditions are obtained if the tangent cone possesses an explicit description as it is possible in the case of linear lower level problems. If the optimal solution of the lower level problem is a PC^1-function, sufficient conditions for a global optimal solution of the optimistic bilevel problem can be formulated. In the second part of the paper relations of the bilevel programming problem to set-valued optimization problems and to mathematical programs with equilibrium constraints are given which can also be used to formulate optimality conditions for the original problem. Finally, a variational inequality approach is described which works well when the involved functions are monotone. It consists in a variational re-formulation of the optimality conditions and looking for a solution of the thus obtained variational inequality among the points satisfying the initial constraints. A penalty function technique is applied to get a sequence of approximate solutions converging to a solution of the original problem with monotone operators.

Key words: Bilevel Programming, Set-valued Optimization, Mathematical Programs with Equilibrium Constraints, Necessary and Sufficient Optimality Conditions, Variational Inequality, Penalty Function Techniques

1 The bilevel programming problem

Bilevel programming problems are hierarchical in the sense that two decision makers make their choices on different levels of hierarchy. While the first one,

S. Dempe and V. Kalashnikov (eds.), *Optimization with Multivalued Mappings*, pp. 3-28

the so-called upper level decision maker or leader fixes his selections 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. Hence, the variables x play the role of parameters in the follower's problem. On the other hand, 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.

To be more precise, let the follower make his decision by solving a parametric optimization problem

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

where $f, g_i : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R},\ i = 1, \dots, p$ are smooth (at least twice continuously differentiable) functions, convex with respect to y for each fixed x.

Then, the leader's problem consists in minimizing the continuously differentiable 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 [1] and [5] and in the annotated bibliography [6].

Since the leader controls only the variable x , this problem is well-defined 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. [32]. Both approaches rest on the introduction of a new lower level problem.

The optimistic approach can be applied if the leader assumes that the follower will always take an optimal solution which is the best one from the leader's point of view, which leads to the problem

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

where

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

Problem (2)-(3) is obviously equivalent to

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

provided that the latter problem has an optimal solution. But note that this equivalence is true only for global minima [5]. It is easy to see that each locally optimal solution of problem (2)-(3) is also a locally optimal solution of problem (4), but the opposite implication is in general not true.

Example 1.1 Consider the simple linear bilevel programming problem

$$\min\{x : y \in \Psi(x), -1 \le x \le 1\},$$

where

$$\Psi(x) = \operatorname*{argmin}_y \{xy : 0 \le y \le 1\}$$

at the point $(x^0, y^0) = (0,0)$. Then, this point is a locally optimal solution to problem (4), i.e. there exists an open neighborhood $W_\varepsilon(0,0) = (-\varepsilon,\varepsilon)\times(-\varepsilon,\varepsilon)$ with $0 < \varepsilon < 1$ such that $x \ge 0$ for all $(x,y) \in W_\varepsilon(0,0)$ with $y \in \Psi(x)$ and $-1 \le x \le 1$. The simple reason for this is that there is no $-\varepsilon < y < \varepsilon$ with $y \in \Psi(x)$ for $x < 0$ since $\Psi(x) = \{1\}$ for $x < 0$. But if we consider the definition of a locally optimistic optimal solution by solving problem (2) then the point $(0,0)$ is not a locally optimistic optimal solution since $x^0 = 0$ is not a local minimum of the function $\varphi_o(x) = x$. △

The basic assumption for this approach is cooperation between the follower and the leader. If the follower cannot be assumed to cooperate with the leader, the latter applies the pessimistic approach

$$\min\{\varphi_p(x) : x \in X\}, \tag{5}$$

where

$$\varphi_p(x) := \max_y\{F(x,y) : y \in \Psi(x)\}. \tag{6}$$

Then, the following notions of optimality can be used:

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

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

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

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

Definition 1.2 *A point $(\overline{x},\overline{y})$ is called a locally pessimistic optimal solution of the bilevel programming problem if*

$$\overline{y} \in \Psi(\overline{x}), \overline{x} \in X, F(\overline{x},\overline{y}) = \varphi_p(\overline{x})$$

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

$$\varphi_p(x) \ge \varphi_p(\overline{x})\ \forall\ x \in X,\ \|x - \overline{x}\| < \varepsilon.$$

Using these definitions it is possible to determine assumptions guaranteeing the existence of locally optimal solutions [5].

(C) The set

$$\{(x,y) : x \in X,\ g(x,y) \le 0\}$$

is nonempty and bounded.

(MFCQ) The Mangasarian-Fromowitz constraint qualification is satisfied at a point $(\overline{x},\overline{y})$ if there is a direction d such that

$$\nabla_y g_i(\overline{x},\overline{y})d < 0,\ \forall\ i \in \{j : g_j(\overline{x},\overline{y}) = 0\}.$$

A point-to-set mapping $\Gamma : \mathbb{R}^p \to 2^{\mathbb{R}^q}$ maps points $w \in \mathbb{R}^p$ to sets $\Gamma(w) \subseteq \mathbb{R}^q$.

Definition 1.3 *A point-to-set mapping $\Gamma : \mathbb{R}^p \to 2^{\mathbb{R}^q}$ is said to be upper semicontinuous at a point $\overline{w} \in \mathbb{R}^p$ if for each open set $A \supseteq \Gamma(\overline{w})$ there is an open set $V \ni \overline{w}$ such that $\Gamma(w) \subseteq A$ for all $w \in V$. The point-to-set mapping Γ is lower semicontinuous at $\overline{w} \in \mathbb{R}^p$ provided that for each open set $A \subseteq \mathbb{R}^q$ with $\Gamma(\overline{w}) \cap A \neq \emptyset$ there is an open set $V \ni \overline{w}$ with $\Gamma(w) \cap A \neq \emptyset$ for all $w \in V$.*

Theorem 1.1 ([21],[33]) *A locally optimal optimistic solution of the bilevel programming problem exists provided the point-to-set mapping $\Psi(\cdot)$ is upper semicontinuous at all points $x \in X$ and assumption (C) is satisfied. A locally optimal pessimistic solution exists if upper semicontinuity of the mapping $\Psi(\cdot)$ is replaced by lower semicontinuity.*

It should be mentioned that the point-to-set mapping $\Psi(\cdot)$ is upper semicontinuous at a point $\overline{x} \in X$ if (C) and (MFCQ) are satisfied at all points $(\overline{x}, \overline{y})$ with $\overline{y} \in \Psi(\overline{x})$. In most cases, to guarantee lower semicontinuity of the point-to-set mapping $\Psi(\cdot)$, uniqueness of an optimal solution of problem (1) is needed.

2 Optimality conditions

To derive optimality conditions for the optimistic bilevel programming problem we have two possibilities. Either we apply the contingent or some other cone to the feasible set of the bilevel programming problem

$$M := \operatorname{Gph} \Psi \cap (X \times \mathbb{R}^m),$$

where $\operatorname{Gph} \Psi := \{(x,y)^\top : y \in \Psi(x)\}$ denotes the graph of the point-to-set mapping $\Psi(\cdot)$, or we use one of the known reformulations of the bilevel programming problem to get a one-level optimization problem and formulate optimality conditions for the latter problem. Focus in this paper is on possible advantages and difficulties related with the one or the other of these approaches. We start with the first one.

Definition 2.1 *The cone*

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

is the contingent (or Bouligand) cone of M.

Theorem 2.1 *If the point $(\overline{x}, \overline{y})^\top \in \operatorname{Gph} \Psi$, $\overline{x} \in X$ is a locally optimal solution of the optimistic problem (4), then*

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

for all

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

On the other hand, if $(\overline{x},\overline{y})^\top \in \operatorname{Gph}\Psi$, $\overline{x} \in X$ *and*

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

for all

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

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

Proof. Let $(\overline{x},\overline{y})^\top \in \operatorname{Gph}\Psi$, $\overline{x} \in X$ be a locally optimal solution of problem (4). Assume that the proposition of the theorem is not satisfied. Then, there exists a direction $(d,r)^\top$ with

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

and

$$\nabla F(\overline{x},\overline{y})(d,r)^\top < 0. \tag{7}$$

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

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

for sufficiently large k, where $\lim\limits_{k\to\infty} \frac{o(t_k)}{t_k} = 0$. Since

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

by the assumption this implies

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

for all sufficiently large k and, hence,

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

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

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

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

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

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

contradicting our assumption. □

Applying this theorem the main difficulty is the computation of the contingent cone. This has been done e.g. in the paper [7].

2.1 The linear case

If bilevel programming problems with linear lower level problems are under consideration, an explicit description of this contingent cone is possible [8] under a certain regularity condition. For this, consider a linear parametric optimization problem

$$\max_y\{c^\top y:\ Ay=b,\ y\geq 0\} \tag{8}$$

with a (m,n)-matrix A 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 (8). A special optimistic bilevel programming problem reads as

$$\min_{y,b,c}\{f(y):Bb=\widetilde{b},Cc=\widetilde{c},y\in\Psi_L(b,c)\}. \tag{9}$$

Using linear programming duality problem (9) has a reformulation as

$$\begin{aligned} f(y) &\longrightarrow \min_{y,b,c,u} \\ Ay &= b \\ y &\geq 0 \\ A^\top u &\geq c \\ y^\top(A^\top u-c) &= 0 \\ Bb &= \widetilde{b} \\ Cc &= \widetilde{c}. \end{aligned} \tag{10}$$

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 locally optimal solution of problem (9) necessary:

Definition 2.2 *A point $\overline{y}$ is a locally optimal solution of problem (9) if there exists an open neighborhood U of $\overline{y}$ such that $f(\overline{y})\leq f(y)$ 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 the problem (10) on c. This dependence rests on solvability of the dual problem and is guaranteed for index sets I in the set $\mathcal{I}(y)$ below.

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

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

Using these definitions, problem (10) can be transformed into the following one by replacing the complementarity conditions:

$$\begin{aligned} f(y) &\longrightarrow \min_{y,b,I} \\ Ay &= b \\ y &\geq 0 \\ y_i &= 0 \quad \forall i \in I \\ Bb &= \tilde{b} \\ I &\in \mathcal{I}(y). \end{aligned} \tag{11}$$

The tangent cone to the feasible set of the last problem is

$$T(\overline{y}) := \bigcup_{I \in \mathcal{I}(\overline{y})} T_I(\overline{y}),$$

where

$$T_I(\overline{y}) = \{d |\ \exists r :\ Ad = r,\ Br = 0,\ d_i \geq 0,\ \forall i \in I(\overline{y}) \setminus I,\ d_i = 0,\ \forall i \in I\}$$

for all $I \in \mathcal{I}(\overline{y})$. Note that $T(\overline{y})$ is the tangent cone to the feasible set of problem (9) with respect to Definition 2.2.

Theorem 2.2 *[Optimality conditions, [8]] If f is differentiable at $\overline{y}$, this point is a local optimum of (9) if and only if $\nabla f(\overline{y}) \cdot d \geq 0$ for all $d \in \operatorname{conv} T(\overline{y})$.*

For an efficient verification of the condition in Theorem 2.2 a compact formula for the convex hull of the tangent cone of the feasible set is crucial. For that consider the relaxed problem of (10)

$$\begin{aligned} f(y) &\longrightarrow \min_{y,b} \\ Ay &= b \\ y_i &\geq 0 \quad i = 1, \dots, l \\ y_i &= 0 \quad i = l+1, \dots, k \\ Bb &= \widetilde{b} \end{aligned} \tag{12}$$

together with the tangent cone to its feasible set

$$T_R(\overline{y}) = \{d|\ \exists r : Ad = r,\ Br = 0,\ d_i \geq 0,\ i = 1, \ldots, l,\ d_i = 0,\ i = l{+}1, \ldots, k\}$$

(relative to y only) at the point $\overline{y}$. Here, it is assumed that $I(\overline{y}) = \{1, \ldots, k\}$ and $I^0(\overline{y}) = \{l + 1, \ldots, k\}$.

Remark 2.1 ([8]) *We have* $j \in I(\overline{y}) \setminus I^0(\overline{y})$ *if and only if the system*

$$\begin{array}{rl} (A^\top u - c)_i = 0 & \forall i \notin I(\overline{y}) \\ (A^\top u - c)_j = 0 & \\ (A^\top u - c)_i \geq 0 & \forall i \in I(\overline{y}) \setminus \{j\} \\ Cc = \tilde{c} & \end{array}$$

has a solution.

In the following theorem we need an assumption: The point $\overline{y}$ is said to satisfy the *full rank condition*, if

$$\mathrm{span}(\{A_i:\ i = k + 1, \ldots, n\}) = \mathbb{R}^m, \tag{FRC}$$

where A_i denotes the ith column of the matrix A.

Theorem 2.3 ([8]) *Let (FRC) be satisfied at the point* $\overline{y}$*. Then,*

$$\mathrm{conv}\, T(\overline{y}) = \mathrm{cone}\, T(\overline{y}) = T_R(\overline{y}). \tag{13}$$

This theorem together with Theorem 2.2 enables us to check local optimality for the problem (9) in polynomial time while, in general, this is an $\mathcal{NP}$-hard problem [31].

2.2 The regular case

It is clear that

$$C_M(\overline{x}, \overline{y}) \subseteq C_\Psi(\overline{x}, \overline{y}) \cap (C_X(\overline{x}) \times \mathbb{R}^m), \tag{14}$$

where $C_\Psi(\overline{x}, \overline{y})$ denotes the contingent cone to the graph of $\Psi(\cdot)$:

$$\begin{aligned} C_\Psi(x, y) := \{\, (u, v)^\top : \exists \{t_k\}_{k=1}^\infty \subset \mathbb{R}_+,\ \exists \{(u^k, v^k)^\top\}_{k=1}^\infty \subset \mathbb{R}^n \times \mathbb{R}^m \\ \text{with } (x, y)^\top + t_k (u^k, v^k)^\top \in \mathrm{Gph}\, \Psi\ \forall k, \\ \lim_{k\to\infty} t_k = 0,\ \lim_{k\to\infty} (u^k, v^k)^\top = (u, v)^\top \} \end{aligned}$$

and $C_X(\overline{x})$ is the contingent cone for the set X at $\overline{x}$. This implies that the sufficient conditions in Theorem 2.1 can be replaced by the assumption that $\nabla F(\overline{x}, \overline{y})(d, r)^\top > 0$ for all $(d, r)^\top \in C_\Psi(\overline{x}, \overline{y}) \cap (C_X(\overline{x}) \times \mathbb{R}^m)$. Conditions for the contingent cone of the solution set mapping of a parametric optimization problem can be found in the monograph [29] and in [30]. Moreover, $C_M(\overline{x}, \overline{y}) = C_\Psi(\overline{x}, \overline{y})$ if $X = \mathbb{R}^n$.

Theorem 2.4 *If $\Psi(x) = \{y(x)\}$ for some locally Lipschitz continuous, directionally differentiable function $y(\cdot)$, then $C_M(\overline{x}, \overline{y}) = C_\Psi(\overline{x}, \overline{y}) \cap (C_X(\overline{x}) \times \mathbb{R}^m)$.*

Here the directional derivative of a function $h : \mathbb{R}^q \to \mathbb{R}^s$ in direction $d \in \mathbb{R}^q$ at a point $w \in \mathbb{R}^q$ is given by

$$h(w; d) = \lim_{t \to 0+} t^{-1}[h(w + td) - h(w)].$$

Proof of Theorem 2.4: Obviously,

$$C_M(\overline{x}, \overline{y}) \subseteq C_\Psi(\overline{x}, \overline{y}) \cap (C_X(\overline{x}) \times \mathbb{R}^m).$$

Let $(d, r)^\top \in C_\Psi(\overline{x}, \overline{y}) \cap (C_X(\overline{x}) \times \mathbb{R}^m)$. Then, by the assumptions $(d, r)^\top \in C_\Psi(\overline{x}, \overline{y})$, i.e. $r = y'(\overline{x}; d)$ and the directional derivative of $y(\cdot)$ is also locally Lipschitz continuous with respect to perturbations of the direction d [10]. Now, take any sequences $\{u^k\}_{k=1}^\infty$ and $\{t_k\}_{k=1}^\infty$ converging to d respectively to zero from above with $\overline{x} + t_k u^k \in X$ for all k existing by definition of $T_X(\overline{x})$. Then, $(y(\overline{x} + t_k u^k) - y(\overline{x}))/t_k$ converges to $y'(\overline{x}; d)$, which completes the proof. □

To determine conditions guaranteeing the assumptions of the last theorem to be valid consider the lower level problem (1) under the assumptions (SSOC), (MFCQ), and (CRCQ):
(SSOC) The strong second-order sufficient optimality condition for problem (1) is satisfied at a point $(\overline{x}, \overline{y})$ with $g(\overline{x}, \overline{y}) \leq 0$ if:

1. The set

$$\Lambda(\overline{x}, \overline{y}) := \{\lambda : \lambda \geq 0, \lambda^\top g(\overline{x}, \overline{y}) = 0, \nabla_y L(\overline{x}, \overline{y}, \lambda) = 0\}$$

is not empty and
2. for all $\overline{\lambda} \in \Lambda(\overline{x}, \overline{y})$ and for all $d \neq 0$ with

$$\nabla_y g_i(\overline{x}, \overline{y}) d = 0 \ \forall \ i : \overline{\lambda}_i > 0$$

there is

$$d^\top \nabla^2_{yy} L(\overline{x}, \overline{y}, \overline{\lambda}) d > 0.$$

Here, $L(x, y, \lambda) = f(x, y) + \lambda^\top g(x, y)$ is the Lagrange function of problem (1).
(CRCQ) The constant rank constraint qualification is satisfied for the problem (1) at the point $(\overline{x}, \overline{y})$ with $g(\overline{x}, \overline{y}) \leq 0$ if there exists an open neighborhood V of $(\overline{x}, \overline{y})$ such that for each subset $J \subseteq \{i : g_i(\overline{x}, \overline{y}) = 0\}$ the set of gradients

$$\{\nabla_y g_i(\overline{x}, \overline{y}) : i \in J\}$$

has a constant rank on V.

Theorem 2.5 ([28], [38]) *Consider problem (1) at a point $(x, y) = (\overline{x}, \overline{y})$ with $\overline{y} \in \Psi(\overline{x})$ and let the assumptions (MFCQ) and (SSOC) be satisfied. Then* *there are an open neighborhood U of $\overline{x}$ and a uniquely determined function $y : U \to \mathbb{R}^m$ such that $y(x)$ is the unique (globally) optimal solution of problem (1) for all $x \in U$. Moreover, if the assumption (CRCQ) is also satisfied, then the function $y(\cdot)$ is locally Lipschitz continuous and directionally differentiable at $\overline{x}$.*

To compute the directional derivative of the solutions function $y(x)$ it is sufficient to compute the unique optimal solution of a quadratic optimization problem using an optimal solution of a linear programming problem as data [38].

Under the assumptions in Theorem 2.5, the bilevel programming problem (both in its optimistic (2) and pessimistic (5) formulations) is equivalent to the problem

$$\min\{G(x) := F(x, y(x)) : x \in X\}. \tag{15}$$

The necessary and sufficient optimality conditions resulting from Theorem 2.1 under the assumptions of Theorem 2.5 and convexity of the lower level problem can be found in [4]:

Theorem 2.6 ([4]) *Consider the bilevel programming problem and let the assumptions (SSOC), (MFCQ), (CRCQ) be valid at a point $(\overline{x}, \overline{y}) \in M$. Then,*

1. *if $(\overline{x}, \overline{y})$ is a locally optimal solution, we have*
$$G'(x; d) \geq 0 \ \forall \ d \in C_X(\overline{x}).$$
2. *if*
$$G'(x; d) > 0 \ \forall \ d \in C_X(\overline{x}),$$
the point $(\overline{x}, \overline{y})$ is a locally optimal solution.

2.3 Application of the protoderivative

Consider the bilevel programming problem in its optimistic formulation (4) and assume that the lower level problem is given in the simpler form

$$\Psi_K(x) := \operatorname*{argmin}_y \{f(x, y) : y \in K\}, \tag{16}$$

where $K \subseteq \mathbb{R}^m$ is a polyhedral set. Then,

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

where $N_K(y)$ denotes the normal cone of convex analysis to the set K at y which is empty if $y \notin K$. Hence, assuming that $X = \mathbb{R}^n$ the problem (4) reduces to

$$\min_{x,y}\{F(x, y) : 0 \in \nabla_y f(x, y) + N_K(y)\}.$$

Then, if we assume that the regularity condition

$$\text{rank}(\nabla^2_{xy} f(x, y)) = m \quad \text{(full rank)} \tag{17}$$

is satisfied, from Theorem 7.1 in [11] we obtain that the solution set mapping Ψ_K is protodifferentiable. Using the formula for the protoderivative we obtain:

Theorem 2.7 ([7]) *Let $(\bar{x}, \bar{y})$ be a locally optimistic solution of the bilevel programming problem (4), where $\Psi_K(x)$ is given by (16). Assume that the solution set mapping Ψ_K is locally bounded and that the qualification condition (17) holds. Then one has*

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

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

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

Here

$$K_* = \{d \in C_K(\bar{y}) : \nabla_y f(\bar{x}, \bar{y})d = 0\}$$

and

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

provided that

$$K = \{y \in \mathbb{R}^m : a_i^\top y \leq b_i, \quad i = 1, \ldots, p\},$$

where $a_i \in \mathbb{R}^m$ for $i = 1, \ldots, p$ and $b \in \mathbb{R}$ for $i = 1, \ldots, p$. Here, $I(\bar{y})$ denotes the set of active indices at $\bar{y}$.

Optimality conditions for problem (4) using the coderivative of Mordukhovich can be found in the papers [7, 13, 44]. While in the paper [44] the coderivative is applied directly to the graph of the solution set mapping, the attempt in the papers [7, 13] applies the coderivative to the normal cone mapping to the feasible set mapping. We will not go into the details here but refer to the paper [13] in the same volume.

2.4 Global minima

The following sufficient condition for a global optimal solution applies in the case when $X = \mathbb{R}^n$.

Theorem 2.8 ([12]) *Consider the problem (1), (4), let the assumptions of Theorem 2.5 be satisfied at all feasible points $(x, y) \in M$. Let $\overline{x}$ be given and assume that*

$$G'(\widetilde{x}; d) > 0 \ \forall \ d \neq 0, \ \forall \ \widetilde{x} \text{ with } G(\widetilde{x}) = G(\overline{x})$$

Then, $\overline{x}$ is a global minimum of the function $G(x) = F(x, y(x))$.

Proof. First, we show that the point $\overline{x}$ is a strict local minimum of G. If $\overline{x}$ is not a strict local minimum then there exists a sequence $\{x^k\}$ converging to $\overline{x}$ such that $G(x^k) \le G(\overline{x})$. Put $d^k = \dfrac{x^k - \overline{x}}{\|x^k - \overline{x}\|}$. Then, $\{d^k\}_{k=1}^{\infty}$ is a bounded sequence and hence has a convergent subsequence $\{d^k\}$ converging to d^0 (say). If we denote this subsequence again by $\{d^k\}_{k=1}^{\infty}$ we have $x^k = \overline{x} + t_k d^k$ where $t_k = \|x^k - \overline{x}\|$. Hence,

$$G(\overline{x} + t_k d^k) - G(\overline{x}) \le 0.$$

This immediately leads to

$$t_k G'(\overline{x}; d^k) + o(t_k) \le 0.$$

Passing to the limit we obtain a contradiction to the assumption. Hence $\overline{x}$ is a strict local minimum.

Now assume that $\overline{x}$ is not a global minimum. Then, there exists x^0 with $G(x^0) < G(\overline{x})$. Consider the line $Z := \{x : x = \lambda x^0 + (1-\lambda)\overline{x}, \lambda \in [0,1]\}$. Then,

1. G is continuous on Z.
2. $x^0 \in Z,\ \overline{x} \in Z$.

Hence, there exist $\{x^1, \ldots, x^p\} \subseteq Z$ with $G(x^i) = G(\overline{x})$ for all i and $G(x) \ne G(\overline{x})$ for all other points in Z. By the assumption this implies that $G(x) \ge G(\overline{x})$ on Z (remember that Z is homeomorphic to a finite closed interval in $\mathbb{R}$ and that $g(\lambda) := G(\lambda x^0 + (1-\lambda)\overline{x}) : \mathbb{R} \to \mathbb{R}$). But this contradicts $G(x^0) < G(\overline{x})$. □

2.5 Optimality conditions for pessimistic optimal solutions

The pessimistic bilevel programming problem is more difficult than the optimistic one. This may be the reason for attacking the optimistic problem (explicitly or not) in most of the references on bilevel programming problems. The investigations in the paper [15] in this volume indicate that this may not be true for discrete bilevel programming problems.

The approach for using the radial directional derivative for deriving necessary and sufficient optimality conditions for pessimistic optimal solutions has been earlier used for linear bilevel programming problems [5, 9].

Definition 2.3 *Let $U \subseteq \mathbb{R}^m$ be an open set, $\overline{x} \in U$ and $\alpha : U \to \mathbb{R}$. We say that α is radial-continuous at $\overline{x}$ in direction $r \in \mathbb{R}^m$, $\|r\| = 1$, if there exists a real number $\alpha_r(\overline{x})$ such that*

$$\lim_{t \downarrow 0} \alpha(\overline{x} + tr) = \alpha_r(\overline{x}).$$

If the radial limit $\alpha_r(\overline{x})$ *exists for all* $r \in \mathbb{R}^m, \|r\| = 1$, α *is called radial-continuous at* $\overline{x}$.
The function α *is radial-directionally differentiable at* $\overline{x}$, *if there exists a positively homogeneous function* $\mathrm{d}\alpha(\overline{x}; \cdot) : \mathbb{R}^m \to \mathbb{R}$ *such that*

$$\alpha(\overline{x} + tr) - \alpha_r(\overline{x}) = t\mathrm{d}\alpha(\overline{x}; r) + o(\overline{x}, tr)$$

with $\lim\limits_{t \downarrow 0} \frac{o(\overline{x}, tr)}{t} = 0$ *holds for all* $r \in \mathbb{R}^m, \|r\| = 1$, *and all* $t > 0$.

Obviously, the vector $\mathrm{d}\alpha(\overline{x}; \cdot)$ *is uniquely defined and is called the radial-directional derivative of* α *at* $\overline{x}$.

It is not very difficult to show, that, for (mixed-discrete) linear bilevel programming problems, the functions $\varphi_o(\cdot)$ and $\varphi_p(\cdot)$ determined in (3) and (6) are radial-directionally differentiable [5].

A necessary optimality condition is given in the next theorem:

Theorem 2.9 ([9]) *Let* $\alpha : \mathbb{R}^m \to \mathbb{R}$ *be a radial-directionally differentiable function and* $\overline{x} \in \mathbb{R}^m$ *a fixed point. If there exists* $r \in \mathbb{R}^m$ *such that one of the following two conditions is satisfied then* $\overline{x}$ *is not a local optimum of the function* α:

- $\mathrm{d}\alpha(\overline{x}; r) < 0$ *and* $\alpha_r(\overline{x}) \leq \alpha(\overline{x})$
- $\alpha_r(\overline{x}) < \alpha(\overline{x})$.

This optimality condition can be complemented by a sufficient one.

Theorem 2.10 ([9]) *Let* $\alpha : \mathbb{R}^m \to \mathbb{R}$ *be a radial-directionally differentiable function and* $\overline{x}$ *a fixed point which satisfies one of the following two conditions.*

- $\alpha(\overline{x}) < \alpha_r(\overline{x}) \ \forall r \in \mathbb{R}^m$
- $\alpha(\overline{x}) \leq \alpha_r(\overline{x}) \ \forall r$ *and* $\mathrm{d}\alpha(\overline{x}; r) > 0 \ \forall r : \alpha(\overline{x}) = \alpha_r(\overline{x}), \|r\| = 1$.

Then, α *achieves a local minimum at* $\overline{x}$.

3 Relations to set-valued optimization

Closely related to bilevel programming problems are also *set-valued optimization problems* e.g. of the kind

$$\text{“}\min_x\text{”}\{\mathcal{F}(x) : x \in X\}, \tag{18}$$

where $\mathcal{F} : X \to 2^{\mathbb{R}^p}$ is a point-to-set mapping sending $x \in X \subseteq \mathbb{R}^n$ to a subset of $\mathbb{R}^p$. To see this assume that $\mathcal{F}(x)$ corresponds to the set of all possible upper level objective function values

$$\mathcal{F}(x) := \bigcup_{y \in \Psi(x)} F(x, y).$$

Thus, the bilevel programming problem is transformed into (18) in the special case of $\mathcal{F}(x) \subseteq \mathbb{R}$.

An edited volume on the related set-valued optimization problems is [2], while [23] is a survey on that topic.

Definition 3.1 ([3]) *Let an order cone $C \subseteq \mathbb{R}^p$ with nonempty interior be given. A pair $(\overline{x}, \overline{z})$ with $\overline{x} \in X$, $\overline{z} \in \mathcal{F}(\overline{x})$ is called a weak minimizer of problem (18) if $\overline{z}$ is a weak minimal element of the set*

$$\mathcal{F}(X) := \bigcup_{x \in X} \mathcal{F}(x).$$

Here, $\overline{z} \in \mathcal{F}(X)$ is a weak minimal element of the set $\mathcal{F}(X)$ if

$$(\overline{z} + \operatorname{int} C) \cap \mathcal{F}(X) = \emptyset.$$

Let C be a polyhedral cone. Then, there exist a finite number of elements l^i, $i = 1, \ldots, p$ such that the dual cone C^* to C is

$$C^* = \{z : z^\top d \geq 0 \ \forall\, d \in C\} = \left\{ z : \exists\, \mu \in \mathbb{R}^p_+ \text{ with } z = \sum_{i=1}^p \mu_i l^i \right\}.$$

The following theorem is well-known:

Theorem 3.1 *If the set $\mathcal{F}(X)$ is convex then a point $\overline{z} \in \mathcal{F}(X)$ is a weak minimal element of $\mathcal{F}(X)$ if and only if $\overline{z}$ is an optimal solution of*

$$\min_z \left\{ \sum_{i=1}^p \mu_i l^{i\top} z : z \in \mathcal{F}(X) \right\}$$

for some $\mu \in \mathbb{R}^p_+$.

In the case of bilevel programming (i.e. $\mathcal{F}(X) \subseteq \mathbb{R}$) $p = 1, l_1 = 1, \mu_1 = 1$ can be selected and the problem in Theorem 3.1 reduces to

$$\min_z \{z : z \in \mathcal{F}(X)\}.$$

For this, the convexity assumption is not necessary but the set $\mathcal{F}(X)$ is only implicitly given. Possible necessary optimality conditions for this problem reduce to the ones discussed above.

The difficulty with the optimistic definition is the following: Assume that there are two decision makers, the first one is choosing $x^0 \in X$ and the second one selects $y^0 \in \mathcal{F}(x^0)$. Assume that the first decision maker has no control over the selection of the second one but that he intends to determine a solution $x^0 \in X$ such that for each selection $\widehat{y} \in \mathcal{F}(\widehat{x})$ and $\widehat{x}$ close to x^0 of the second one there exists $y^0 \in \mathcal{F}(x^0)$ which is preferable to $\widehat{y}$. In this case, since the selection of the second decision maker is out of control of the first one, the latter cannot evaluate the quality of her selection chosen according to the above definition.

To weaken this definition assume that the first decision maker is able to compute the sets $\mathcal{F}(x)$ for all $x \in X$. Then, he can try to compute a point $x^* \in X$ such that

$$\mathcal{F}(x) \subseteq \mathcal{F}(x^*) + C$$

for all $x \in X$ sufficiently close to x^*. In distinction to Definition 3.1 this reflects a pessimistic point of view in the sense that the first decision maker bounds the damage caused by the selection of the second one. Let

$$\mathcal{F}(x^1) \preccurlyeq_C \mathcal{F}(x^2) \Longleftrightarrow \mathcal{F}(x^2) \subseteq \mathcal{F}(x^1) + C.$$

Definition 3.2 *Let an order cone $C \subseteq \mathbb{R}^p$ be given. A point $\overline{x} \in X$ is called a pessimistic local minimizer of problem (18) if*

$$\mathcal{F}(\overline{x}) \preccurlyeq_C \mathcal{F}(x) \;\forall\; x \in X \cap \{z : \|z - \overline{x}\| < \varepsilon\}$$

for some $\varepsilon > 0$.

Theorem 3.2 *Let $\overline{x} \in X$ be not a pessimistic local minimizer and assume that C and $\mathcal{F}(x)$ are convex sets for all $x \in X$. Then there exist a vector $\hat{k} \in C^* \setminus \{0\}$ and a point $\hat{x} \in X$ such that*

$$\min\{\hat{k}^\top y : y \in \mathcal{F}(\hat{x})\} < \min\{\hat{k}^\top y : y \in \mathcal{F}(\overline{x})\}.$$

Proof: Let $\overline{x} \in X$ be not a pessimistic minimizer. Then, by definition there exists $\hat{x} \in X$ sufficiently close to $\overline{x}$ such that $\mathcal{F}(\overline{x}) \not\preccurlyeq_C \mathcal{F}(\hat{x})$. Then there necessarily exists $\hat{y} \in \mathcal{F}(\hat{x})$ with $\hat{y} \notin \mathcal{F}(\overline{x}) + C$. Since by our assumption both $\mathcal{F}(\overline{x})$ and C are convex there is a vector $\hat{k} \neq 0$ with

$$\min\{\hat{k}^\top y : y \in \mathcal{F}(\hat{x})\} \leq \hat{k}^\top \hat{y} < \min\{\hat{k}^\top y : y \in \mathcal{F}(\overline{x}) + C\}$$

by a strong separation theorem in convex analysis (see e.g. [42]). Now assume that $\hat{k} \notin C^*$. Then, since C is a cone, we get

$$\begin{aligned} &\min\{\hat{k}^\top y : y \in \mathcal{F}(\overline{x}) + C\} \\ &= \min\{\hat{k}^\top (y^1 + ty^2) : y^1 \in \mathcal{F}(\overline{x}), y^2 \in C\} \\ &= \min\{\hat{k}^\top y^1 : y^1 \in \mathcal{F}(\overline{x})\} + t \min\{\hat{k}^\top y^2 : y^2 \in C\} \end{aligned}$$

for all $t \geq 0$. But since $\hat{k} \notin C^*$, the last term tends to minus infinity for increasing t which cannot be true since it is bounded from below by $\hat{k}^\top \hat{y}$. This proves the theorem. □

This implies that, if for all $k \in C^*$

$$\min_{x,y}\{k^\top y : (x, y) \in \operatorname{Gph} \mathcal{F}, x \in X, \|x - \overline{x}\| \leq \varepsilon\} \geq \min_{y}\{k^\top y : y \in \mathcal{F}(\overline{x})\}$$

then $\overline{x}$ is a local pessimistic minimizer. The main difference of this result to Theorem 3.1 is that here this condition needs to be satisfied for all elements $k \in C^*$ whereas there must exist one element $k \in C^*$ with the respective condition in Theorem 3.1.

Applied to bilevel programming, the notions of both the optimistic and the pessimistic minimizer coincide.

4 Relation to mathematical programs with equilibrium conditions

Applying the Karush-Kuhn-Tucker conditions to the lower level problem (1) in (4) we derive the problem

$$\begin{aligned} F(x,y) &\to \min_{x,y,u} \\ \nabla_x L(x,y,u) &= 0 \\ g(x,y) &\leq 0 \\ u &\geq 0 \\ u^\top g(x,y) &= 0 \\ x &\in X \end{aligned} \tag{19}$$

provided that the lower level problem satisfies the (MFCQ) at all feasible points for all $x \in X$ and that it is a convex optimization problem for fixed $x \in X$. Problem (19) is called a mathematical program with equilibrium constraints (MPEC) [34, 37]. There has been many interesting results concerning optimality conditions for MPECs in the recent time, cf. e.g. [16, 17, 18, 19, 40, 43]. Here we are interested in conditions needed for applying such conditions.

Example 4.1 This example shows that the convexity assumption is crucial.

Consider the problem [35]

$$\min_{x,y}\{(x-2)+(y-1)^2 : y \in \Psi(x)\}$$

where $\Psi(x)$ is the set of optimal solutions of the following unconstrained optimization problem on the real axis:

$$-x\exp\{-(y+1)^2\} - \exp\{-(y-1)^2\} \to \min_y$$

Then, the necessary optimality conditions for the lower level problem are

$$x(y+1)\exp\{-(y+1)^2\} + (y-1)\exp\{-(y-1)^2\} = 0$$

which has three solutions for $0.344 \leq x \leq 2.903$. The global optimum of the lower level problem is uniquely determined for all $x \neq 1$ and it has a jump at the point $x = 1$. Here the global optimum of the lower level problem can be found at the points $y = \pm 0.957$. The point $(x^0; y^0) = (1; 0.957)$ is also the global optimum of the optimistic bilevel problem.

But if the lower level problem is replaced with its necessary optimality conditions and the necessary optimality conditions for the resulting problem are solved then three solutions: $(x, y) = (1.99; 0.895)$, $(x, y) = (2.19; 0.42)$, $(x, y) = (1.98; -0.98)$ are obtained. Surprisingly, the global optimal solution of the bilevel problem is not obtained with this approach. The reason for this is that the problem

$$\min\{(x-2)+(y-1)^2 : x(y+1)\exp\{-(y+1)^2\}+(y-1)\exp\{-(y-1)^2\}=0\}$$

has a much larger feasible set than the bilevel problem. And this feasible set has no jump at the point $(x, y) = (1; 0.957)$ but is equal to a certain connected curve in $\mathbb{R}^2$. And on this curve the objective function has no stationary point at the optimal solution of the bilevel problem. △

The following example can be used to illustrate that the above equivalence between problems (2) and (19) is true only if global optima are searched for.

Theorem 4.1 ([5]) *Consider the optimistic bilevel programming problem (1), (2), (3) and assume that, for each fixed y, the lower level problem (1) is a convex optimization problem for which (MFCQ) is satisfied for each fixed x and all feasible points. Then, each locally optimal solution for the problem (1), (2), (3) corresponds to a locally optimal solution for problem (19).*

This theorem implies that it is possible to derive necessary optimality conditions for the bilevel programming problem by applying the known conditions for MPECs. This has been done e.g. in [5]. Using the recent conditions in the papers [16, 17, 18, 19, 43] interesting results can be obtained. It would be a challenging topic for future research to check if the assumptions used in these papers can successfully be interpreted for bilevel programming problems.

But, the application of these results to get strong necessary optimality conditions and also to get sufficient optimality conditions seems to be restricted. This can be seen in the following example.

Example 4.2 We consider a linear bilevel programming problem with an optimistic optimal solution $(\overline{x}, \overline{y})$. Assume that the linear independence constraint qualification is not satisfied at the lower level problem at $(\overline{x}, \overline{y})$ and that there are more than one Lagrange multiplier for the lower level problem at $\overline{y}$. Then, the situation is as depicted in fig. 1: We see a part of the feasible set of the upper level problem (which is the union of faces of the graph of the lower level feasible set $\{(x, y) : Ax \le y\}$) in the right-hand side picture. There is a kink at the point $(\overline{x}, \overline{y})$. The point $(\overline{x}, \overline{y})$ belongs to the intersection of two faces of the graph of the lower level feasible set $\{(x, y) : Ax \le y\}$. The optimal solution of the lower level problem is unique for all x, hence the lower level solution function can be inserted into the upper level objective function. In the left-hand side picture we see the graph of the set of Lagrange multipliers in the lower level problem which is assumed to reduce to a singleton for $x \neq \overline{x}$ and is multivalued for $x = \overline{x}$.

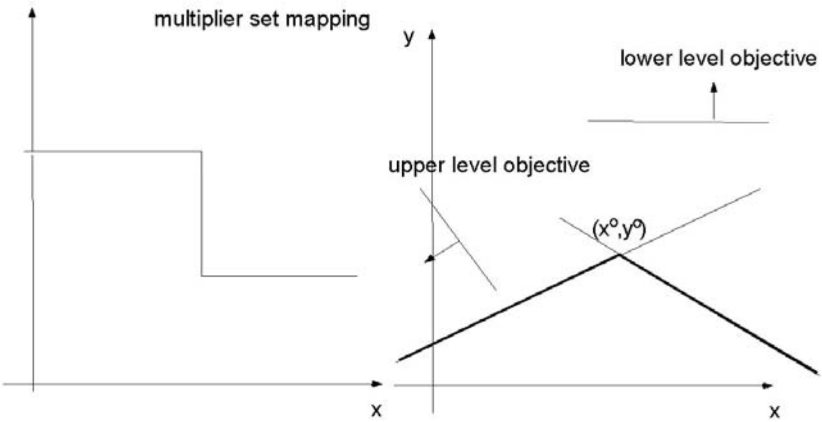

Fig. 1. Feasible set in the upper level problem and Lagrange multiplier mapping of the lower level problem

Now assume that we have used the MPEC corresponding to the bilevel programming problem for deriving necessary optimality conditions. For this we fix a feasible solution $(\overline{x}, \overline{y}, \overline{\lambda})$ of the MPEC, where $\overline{\lambda}$ denotes one Lagrange multiplier of the lower level problem at $\overline{y}$. Then, the necessary optimality conditions for the MPEC are satisfied. They show, that there does not exist a better feasible solution than $(\overline{x}, \overline{y}, \overline{\lambda})$ in a suitable small neighborhood of this point for the MPEC. This neighborhood of $(\overline{x}, \overline{y}, \overline{\lambda})$ restricts the λ-part to a neighborhood of $\overline{\lambda}$. Due to complementarity slackness this restriction implies for the bilevel programming problem that the feasible set of the upper level problem is restricted to one face of the graph of $\{(x, y) : Ax \leq y\}$, see fig. 2, where this face is the right-hand sided one. Hence, the necessary optimality conditions for the corresponding MPEC mean that the point under consideration is a stationary (here optimal) solution of the bilevel programming problem but only with respect to a part (and not with respect to an open neighborhood !) of the feasible set.

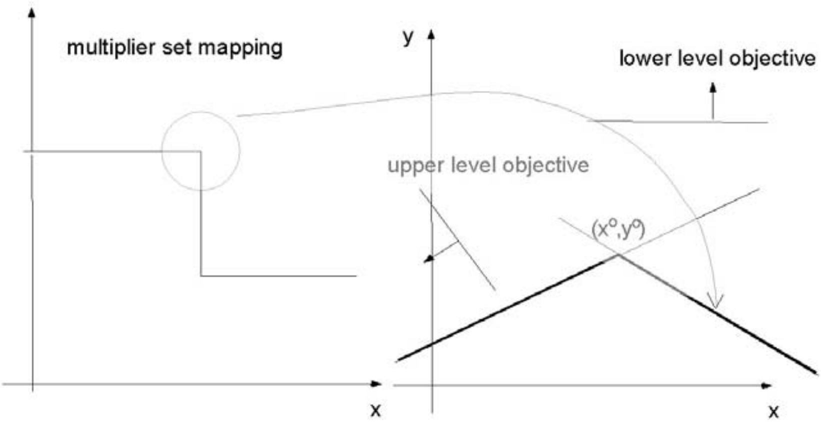

Fig. 2. Neighborhood of a Lagrange multiplier $\overline{\lambda}$ and corresponding part of the feasible set of the bilevel programming problem

5 Variational inequality approach

The problem of solving a mathematical program with variational inequalities or complementarity conditions as constraints arises quite frequently in the analysis of physical and socio-economic systems. According to a remark in the paper [20], the current state-of-the-art for solving such problems is heuristic. The latter paper [20] presents an exterior-point penalty method based on M.J. Smith's optimization formulation of the finite-dimensional variational inequality problem [41]. In the paper by J. Outrata [36], attention is also paid to this type of optimization problems.

An approach to solving the above-mentioned problem consists in a variational re-formulation of the optimization criterion and looking for a solution of the thus obtained variational inequality among the points satisfying the initial variational inequality constraints. This approach works well for the case when both operators involved are monotone and it is enlightened in the first part of the section. Namely, in subsection 5.2, we examine conditions under which the set of the feasible points is non-empty, and compare the conditions with those established previously [22]. Subsection 5.3 describes a penalty function method solving the bilevel problem after having reduced it to a single variational inequality with a penalty parameter.

5.1 Existence theorem

Let X be a non-empty, closed, convex subset of R^n and G a continuous mapping from X into R^n. Suppose that $\mathcal{F}$ is pseudo-monotone with respect to X, i.e.

$$(x-y)^{\mathrm{T}}\mathcal{F}(y) \geq 0 \quad \text{implies} \quad (x-y)^{\mathrm{T}}\mathcal{F}(x) \geq 0 \quad \forall x, y \in X, \tag{20}$$

and that there exists a vector $x^0 \in X$ such that

$$\mathcal{F}(x^0) \ \in \ \text{int}\ (0^+X)^*, \tag{21}$$

where int$(\cdot)$ denotes the interior of the set. Here 0^+X is the recession cone of the set X, i.e. the set of all directions $s \in R^n$ such that $X + s \subset X$; at last, C^* is the dual cone of $C \subset R^n$, i.e.

$$C^* = \{y \in R^n \ : \ y^{\mathrm{T}}x \geq 0 \qquad \forall x \in C\}. \tag{22}$$

Hence, condition (21) implies that the vector $\mathcal{F}(x^0)$ lies within the interior of the dual to the recession cone of the set X.

Under these assumptions, the following result obtains:

Proposition 5.1 ([24]) *The variational inequality problem: to find a vector* $z \in X$ *such that*

$$(x-z)^{\mathrm{T}}\mathcal{F}(z) \geq 0 \qquad \forall x \in X, \tag{23}$$

has a non-empty, compact, convex solution set.

Proof. It is well-known [27] that the pseudo-monotonicity (20) and continuity of the mapping G imply convexity of the solution set

$$Z = \{z \in X \ : \ (x-z)^{\mathrm{T}}\mathcal{F}(z) \geq 0 \qquad \forall x \in X\}, \tag{24}$$

of problem (23) provided that the latter is non-empty. Now we show the existence of at least one solution to this problem. In order to do that, we use the following fact [14]: if there exists a non-empty bounded subset D of X such that for every $x \in X \backslash D$ there is a $y \in D$ with

$$(x-y)^{\mathrm{T}}\mathcal{F}(x) > 0, \tag{25}$$

then problem (23) has a solution. Moreover, the solution set (24) is bounded because $Z \subset D$. Now, we construct the set D as follows:

$$D = \{\ x \in X\!: \ (x-x^0)^{\mathrm{T}}\mathcal{F}(x^0) \leq 0\ \}. \tag{26}$$

The set D is clearly non-empty, since it contains the point x^0. Now we show that D is bounded, even if X is not such. On the contrary, suppose that a sequence $\{x^k\} \subseteq D$ is norm-divergent, i.e. $\|x^k - x^0\| \ \to \ +\infty$ when $k \to \infty$. Without lack of generality, assume that $x^k \neq x^0,\ k = 1, 2, \ldots,$ and consider the inequality

$$\frac{(x^k - x^0)^{\mathrm{T}}\mathcal{F}(x^0)}{\|x^k - x^0\|} \leq 0, \qquad k = 1, 2, \ldots, \tag{27}$$

which follows from definition (26) of the set D. Again without affecting generality, accept that the normed sequence $(x^k - x^0)/\|x^k - x^0\|$ converges to a vector $s \in R^n$, $\|s\| = 1$. It is well-known (cf. [39], Theorem 8.2) that $s \in 0^+X$. From (27), we deduce the limit relation

$$s^{\mathrm{T}}\mathcal{F}(x^0) \ \leq \ 0. \tag{28}$$

Since $0^+X \neq \{0\}$ (as X is unbounded and convex), we have $0 \in \partial(0^+X)^*$, hence $\mathcal{F}(x^0) \neq 0$. Now it is easy to see that inequality (28) contradicts assumption (21). Indeed, the inclusion $\mathcal{F}(x^0) \in \text{int } (0^+X)^*$ implies that $s^{\mathrm{T}}\mathcal{F}(x^0) > 0$ for any $s \in 0^+X, s \neq 0$. The contradiction establishes the boundedness of the set D, and the statement of Proposition 5.1 therewith. Indeed, for a given $x \in X \backslash D$, one can pick $y = x^0 \in D$ with the inequality $(x-y)^{\mathrm{T}}\mathcal{F}(y) > 0$ taking place. The latter, jointly with the pseudo-monotonicity of $\mathcal{F}$, implies the required condition (25) and thus completes the proof. □

Remark 5.1 *The assertion of Proposition 5.1 has been obtained also in [22] under the same assumptions except for inclusion (21), which is obviously invariant with respect to an arbitrary translation of the set X followed by the corresponding transformation of the mapping G. Instead of (21), the authors [22] used another assumption $\mathcal{F}(x^0) \in \mathrm{int}(X^*)$ which is clearly not translation-invariant.*

Now suppose that the solution set Z of problem (23) contains more than one element, and consider the following variational inequality problem: to find a vector $z^* \in Z$ such that

$$(z - z^*)^{\mathrm{T}}\mathcal{G}(z^*) \geq 0 \qquad \text{for all} \quad z \in Z. \tag{29}$$

Here, the mapping $\mathcal{G}\colon X \to R^n$ is continuous and strictly monotone over X; i.e.

$$(x-y)^{\mathrm{T}}\left[\mathcal{G}(x) - \mathcal{G}(y)\right] > 0 \qquad \forall x, y \in X, x \neq y. \tag{30}$$

In this case, the compactness and convexity of the set Z guaranties [14] the existence of a unique (due to the strict monotonicity of $\mathcal{G}$) solution z of the problem (29). We refer to problem (23), (24), (29) as the *bilevel variational inequality* (BVI). In the next subsection, we present a penalty function algorithm solving the BVI without explicit description of the set Z.

5.2 Penalty function method

Fix a positive parameter ε and consider the following parametric variational inequality problem: Find a vector $x_\varepsilon \in X$ such that

$$(x - x_\varepsilon)^{\mathrm{T}}[\mathcal{F}(x_\varepsilon) + \varepsilon\mathcal{G}(x_\varepsilon)] \geq 0 \qquad \text{for all} \quad x \in X. \tag{31}$$

If we assume that the mapping $\mathcal{F}$ is monotone over X, i.e.

$$(x-y)^{\mathrm{T}}[\mathcal{F}(x) - \mathcal{F}(y)] \geq 0 \qquad \forall x, y \in X, \tag{32}$$

and keep intact all the above assumptions regarding $\mathcal{F}, \mathcal{G}$ and Z, then the following result obtains:

Proposition 5.2 ([24]) *For each sufficiently small value $\varepsilon > 0$, problem (31) has a unique solution x_ε. Moreover, x_ε converge to the solution z^* of BVI (23), (24), (29) when $\varepsilon \to 0$.*

Proof. Since $\mathcal{F}$ is monotone and $\mathcal{G}$ is strictly monotone, the mapping $\Phi_\varepsilon = \mathcal{F} + \varepsilon\mathcal{G}$ is strictly monotone on X for any $\varepsilon > 0$. It is also clear that if x^0 satisfies (21) then the following inclusion holds

$$\Phi_\varepsilon(x^0) = \mathcal{F}(x^0) + \varepsilon\mathcal{G}(x^0) \in \text{int } (0^+X)^*, \tag{33}$$

if $\varepsilon > 0$ is small enough. Hence, Proposition 5.1 implies the first assertion of Proposition 5.2; namely, for every $\varepsilon > 0$ satisfying (33), the variational inequality (31) has a unique solution x_ε .

From the continuity of F and G, it follows that each (finite) limit point $\bar{x}$ of the generalized sequence $Q = \{x_\varepsilon\}$ of solutions to problem (31) solves variational inequality (23); that is, $\bar{x} \in Z$. Now we prove that the point $\bar{x}$ solves problem (29), too. In order to do that, we use the following relations valid for any $z \in Z$ due to (23), (29) and (31):

$$(z - x_\varepsilon)^{\mathrm{T}}[\mathcal{F}(z) - \mathcal{F}(x_\varepsilon)] \geq 0, \tag{34}$$

$$(z - x_\varepsilon)^{\mathrm{T}}\mathcal{F}(z) \leq 0, \tag{35}$$

$$(z - x_\varepsilon)^{\mathrm{T}}\mathcal{F}(x_\varepsilon) \geq -\varepsilon(z - x_\varepsilon)^{\mathrm{T}}\mathcal{G}(x_\varepsilon). \tag{36}$$

Subtracting (36) from (35) and using (34), we obtain the following series of inequalities

$$0 \leq (z - x_\varepsilon)^{\mathrm{T}}[\mathcal{F}(z) - \mathcal{F}(x_\varepsilon)] \leq \varepsilon(z - x_\varepsilon)^{\mathrm{T}}\mathcal{G}(x_\varepsilon). \tag{37}$$

From (37) we have $(z - x_\varepsilon)^{\mathrm{T}}\mathcal{G}(x_\varepsilon) \geq 0$ for all $\varepsilon > 0$ and $z \in Z$. Since $\mathcal{G}$ is continuous, the following limit relation holds: $(z - \bar{x})^{\mathrm{T}}\mathcal{G}(\bar{x}) \geq 0$ for each $z \in Z$, which means that $\bar{x}$ solves (29).

Thus we have proved that every limit point of the generalized sequence Q solves BVI (23), (24), (29). Hence, Q can have at most one limit point. To complete the proof of Proposition 5.2, it suffices to establish that the set Q is bounded, and consequently, the limit point exists. In order to do that, consider a norm-divergent sequence $\{x_{\varepsilon_k}\}$ of solutions to parametric problem (31) where $\varepsilon_k \to 0$ as $k \to \infty$. Without loss of generality, suppose that $x_{\varepsilon_k} \neq x^0$ for each k, and $\dfrac{(x_{\varepsilon_k} - x^0)}{\|x_{\varepsilon_k} - x^0\|} \to s \in R^n, \quad \|s\| = 1$; here x^0 is the vector from condition (21). Since $\|x_{\varepsilon_k} - x^0\| \to +\infty$, we get $s \in 0^+X$ (cf. [39]). As the mappings $\mathcal{F}$ and $\mathcal{G}$ are monotone, the following inequalities take place for all $k = 1, 2, \ldots$:

$$(x_{\varepsilon_k} - x^0)^{\mathrm{T}}[\mathcal{F}(x_{\varepsilon_k}) + \varepsilon_k\mathcal{G}(x_{\varepsilon_k})] \leq 0, \tag{38}$$

and hence,

$$(x_{\varepsilon_k} - x^0)^{\mathrm{T}}[\mathcal{F}(x^0) + \varepsilon_k \mathcal{G}(x^0)] \le 0. \tag{39}$$

Dividing inequality (39) by $\|x_{\varepsilon_k} - x^0\|$ we obtain

$$\frac{(x_{\varepsilon_k} - x^0)^{\mathrm{T}}}{\|x_{\varepsilon_k} - x^0\|} \cdot [\mathcal{F}(x^0) + \varepsilon_k \mathcal{G}(x^0)] \le 0, \qquad k = 1, 2, ..., \tag{40}$$

which implies (as $\varepsilon_k \to 0$) the limit inequality $s^{\mathrm{T}}\mathcal{F}(x^0) \le 0$. Since $s \ne 0$, the latter inequality contradicts assumption 21. This contradiction demonstrates the set Q to be bounded which completes the proof. □

Example 5.1 *Let $\Omega \subseteq R^m$, $\Lambda \subseteq R^n$ be subsets of finite-dimensional Euclidean spaces and $g : \Omega \times \Lambda \to R$, $f : \Omega \times \Lambda \to R^n$ be continuous mappings. Consider the following mathematical program with variational inequality constraint:*

$$\min_{(u,v)\in\Omega\times\Lambda} g(u, v), \tag{41}$$

subject to

$$f(u, v)^T(w - v) \ge 0, \qquad \forall w \in \Lambda. \tag{42}$$

If the function g is continuously differentiable, then problem (41)-(42) is obviously tantamount to BVI (23), (24), (29) with the gradient mapping $g'(z)$ used as $\mathcal{G}(z)$ and $\mathcal{F}(u, v) = [0; f(u, v)]$; here $z = (u, v) \in \Omega \times \Lambda$.

As an example, examine the case when

$$g(u, v) = (u - v - 1)^2 + (v - 2)^2; \quad f(u, v) = uv; \quad \Omega = \Lambda = R^1_+. \tag{43}$$

Then it is readily verified that $z^* = (1; 0)$ solves problem (41)-(42) and the parametrized mapping is given by

$$\Phi_\varepsilon(u, v) = \big[\varepsilon(2u - 2v - 2); uv + \varepsilon(-2u + 4v - 2)\big]. \tag{44}$$

Now solving the variational inequality: Find $(u_\varepsilon, v_\varepsilon) \in R^2_+$ such that

$$\Phi_\varepsilon(u_\varepsilon, v_\varepsilon)^T \big[(u, v) - (u_\varepsilon, v_\varepsilon)\big] \ge 0 \quad \forall (u, v) \in R^2_+, \tag{45}$$

we obtain

$$u_\varepsilon = v_\varepsilon + 1; \quad v_\varepsilon = -\frac{1}{2} - \varepsilon + \sqrt{(\frac{1}{2} + \varepsilon)^2 + 4\varepsilon}. \tag{46}$$

Clearly $(u_\varepsilon, v_\varepsilon) \to z^*$ when $\varepsilon \to 0$. △

Acknowledgment: The second and third authors' research was financially supported by the PAICyT project CA71033-05 at the UANL (Mexico), by the Russian Foundation for Humanity Research project RGNF 04-02-00172, and by the CONACyT-SEP project SEP-2004-C01-45786 (Mexico).

References

1. J.F. Bard (1998): *Practical Bilevel Optimization: Algorithms and Applications*, Dordrecht: Kluwer Academic Publishers.
2. G.Y. Chen and J. Jahn (eds.) (1998), *Mathematical Methods of Operations Research*, vol. 48.
3. G. Y. Chen and J. Jahn (1998). Optimality conditions for set-valued optimization problems. *Mathematical Methods of Operations Research*, 48: 187-200.
4. S. Dempe (1992). A necessary and a sufficient optimality condition for bilevel programming problems. *Optimization*, 25: 341–354.
5. S. Dempe (2002): *Foundations of Bilevel Programming*, Dordrecht: Kluwer Academic Publishers, 2002.
6. S. Dempe (2003). Annotated Bibliography on Bilevel Programming and Mathematical Programs with Equilibrium Constraints, *Optimization*, 52:, 333-359.
7. S. Dempe, J. Dutta, S. Lohse (2005) Optimality conditions for bilevel programming problems, Preprint, TU Bergakademie Freiberg.
8. S. Dempe, S. Lohse: Inverse linear programming, In: *Recent Advances in Optimization. Proceedings of the 12th French-German-Spanish Conference on Optimization held in Avignon, September 20-24, 2004,* Edited by A. Seeger. Lectures Notes in Economics and Mathematical Systems, Vol. 563, Springer-Verlag Berlin Heidelberg, 2006. pp. 19-28.
9. S. Dempe, T. Unger (1999). Generalized PC^1–functions, *Optimization*, 46:311-326.
10. V.F. Dem'yanov and A.M. Rubinov (1986): *Quasidifferential Calculus*, New York: Optimization Software.
11. A. Dontchev and R. T. Rockafellar (2001). Ample parametrization of variational inclusions, *SIAM Journal on Optimization*, 12: 170-187.
12. J. Dutta and S. Dempe (2004). *Global minima of locally Lipschitz functions*, KANGAL Research Report No. 2004/018, I.I.T. Kanpur, India, 2004.
13. J. Dutta, S. Dempe (2006). Bilevel programming with convex lower level problems, In: S. Dempe, V. Kalashnikov (eds.): *Optimization with Multivalued Mappings: Theory, Applications and Algorithms,* Springer Verlag, Berlin.
14. B.C. Eaves (1971). On the basic theorem of complementarity, *Mathematical Programming*, 1: 68-75.
15. D. Fanghänel (2006). Optimality criteria for bilevel programming problems using the radial subdifferential, In: S. Dempe, V. Kalashnikov (eds.): *Optimization with Multivalued Mappings: Theory, Applications and Algorithms,* Springer Verlag, Berlin.
16. M. L. Flegel and C. Kanzow (2003). A Fritz John approach to first order optimality conditions for mathematical programs with equilibrium constraints, *Optimization*, 52: 277-286.
17. M. L. Flegel and C. Kanzow. On the Guignard constraint qualification for mathematical programs with equilibrium constraints, *Optimization*, to appear.
18. M. L. Flegel and C. Kanzow. On M-stationarity for mathematical programs with equilibrium constraints, *Journal of Mathematical Analysis and Applications*, to appear.
19. M. L. Flegel and C. Kanzow (2005). Abadie-type constraint qualification for mathematical programs with equilibrium constraints, *Journal of Optimization Theory and Applications*, 124: 595–614.

20. P.T. Harker and S.-C. Choi (1991). A penalty function approach for mathematical programs with variational inequality constraints, *Information and Decision Technologies*, 17: 41-50.
21. P.T. Harker and J.-S. Pang (1988). Existence of optimal solutions to mathematical programs with equilibrium constraints, *Operations Research Letters*, 7: 61-64.
22. P.T. Harker and J.-S. Pang (1990). Finite-dimensional variational inequality and nonlinear complementarity problems: a survey of theory, algorithms and applications, *Mathematical Programming*, 48: 161-220.
23. J. Jahn (2000). *Set-valued optimization : a survey*, Preprint, Universität Erlangen-Nürnberg. Institut für Angewandte Mathematik.
24. G.Isac, V.A. Bulavsky and V.V. Kalashnikov (2002). *Complementarity, Equilibrium, Efficiency and Economics*, Kluwer Academic Publishers, Dordrecht/Boston/London, 468 p.
25. V.V. Kalashnikov and N.I. Kalashnikova (1996). Solving two-level variational inequality, *Journal of Global Optimization*, 8: 289-294.
26. S. Karamardian (1976). Complementarity over cones with monotone and pseudo-monotone maps, *Journal of Optimization Theory and Applications*, 18: 445-454.
27. S. Karamardian (1976). An existence theorem for the complementarity problem, *Journal of Optimization Theory and Applications*, 18: 455-464.
28. M. Kojima (1980). Strongly stable stationary solutions in nonlinear programs. In: S.M. Robinson (ed.): *Analysis and Computation of Fixed Points*, New York: Academic Press, pp. 93–138.
29. D. Klatte and B. Kummer (2002). *Nonsmooth Equations in Optimization; Regularity, Calculus, Methods and Applications*, Dordrecht: Kluwer Academic Publishers.
30. A. B. Levy and R. T. Rockafellar (1995). Sensitivity of solutions in nonlinear programming problems with nonunique multipliers. In: D.-Z. Du and L. Qi and R. S. Womersley (eds.): *Recent Advances in Nonsmooth Optimization*, World Scientific Publishing Co., Singapore et al., 215–223.
31. S. Lohse: Komplexitätsbeweis. Technical Report, TU Bergakademie Freiberg, 2005. see `http://www.mathe.tu-freiberg.de/~lohses/Dateien/komplex.pdf`
32. P. Loridan and J. Morgan (1996). Weak via strong Stackelberg problem: New results. *Journal of Global Optimization*, 8: 263-287.
33. R. Lucchetti and F. Mignanego and G. Pieri(1987). Existence theorem of equilibrium points in Stackelberg games with constraints. *Optimization*, 18: 857-866.
34. Z.-Q. Luo and J.-S. Pang and D. Ralph (1996). *Mathematical Programs with Equilibrium Constraints*. Cambridge University Press, Cambridge.
35. J. A. Mirrlees (1999). The theory of moral hazard and unobservable bevaviour: part I. *Review of Economic Studies*, 66: 3-21.
36. J.V. Outrata (1994). On optimization problems with variational inequality constraints, *SIAM Journal on Optimization*, 4: 340-357.
37. J. Outrata and M. Kočvara and J. Zowe (1998). *Nonsmooth Approach to Optimization Problems with Equilibrium Constraints*. Kluwer Academic Publishers, Dordrecht.
38. D. Ralph and S. Dempe (1995). Directional Derivatives of the Solution of a Parametric Nonlinear Program. *Mathematical Programming*, 70: 159-172.
39. R.T. Rockafellar (1970). *Convex Analysis*, Princeton University Press, Princeton, New Jersey.

40. H. Scheel and S. Scholtes (2000). Mathematical programs with equilibrium constraints: stationarity, optimality, and sensitivity. *Mathematics of Operations Research*, 25: 1-22.
41. M.J. Smith (1984). A descent algorithm for solving monotone variational inequalities and monotone complementarity problems, *Journal of Optimization Theory and Applications*, 44: 485-496.
42. H. Tuy (1998). *Convex Analysis and Global Optimization.* Kluwer Academic Publishers, Dordrecht.
43. J. J. Ye (2005). Necessary and sufficient optimality conditions for mathematical programs with equilibrium constraints, *Journal of Mathematical Analysis and Applications*, 307: 350-369.
44. R. Zhang (1994). Problems of hierarchical optimization in finite dimensions, *SIAM Journal on Optimization*, 4: 521-536.