Chapter 1

BILEVEL PROGRAMMING - A SURVEY

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

Abstract Bilevel programming problems are hierarchical optimization problems where an objective function is to be minimized over the graph of the solution set mapping of a second parametric optimization problem. It is the aim of the paper to give a survey for this living research area indicating main recent approaches to solve such problems and to describe optimality conditions as well as to touch main particularities of this problem class.

1. Introduction

Bilevel programming problems are hierarchical mathematical problems where the set of all variables is partitioned between vectors x and y. Given y, the vector x is to be chosen as an optimal solution $x = x(y)$ of an optimization problem parametrized in y:

$$x(y) \in \Psi(y) := \operatorname*{Argmin}_x \; \{f(x, y) : g(x, y) \leq 0\}. \tag{1.1}$$

This problem is the so-called *lower level problem* or the *follower's problem*. Here, $f : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$, $g : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^p$. Equality constraints can be added without serious problems. Sometimes, integrality conditions are added to the lower level problem. The point-to-set mapping $\Psi : \mathbb{R}^m \to 2^{\mathbb{R}^n}$ is the solution set mapping of the problem (1.1). The solution $x(y)$ is called the rational reaction of the follower on the leader's choice y. Knowing this reaction, the bilevel problem reads as follows:

$$\text{“}\min_y\text{''}\{F(x, y) : G(y) \leq 0,\; x \in \Psi(y)\}. \tag{1.2}$$

This problem is the *leader's problem* or the *bilevel problem.* Quotation marks have been added in (1.2) due to an unclear definition of the objective function value $F(x, y)$ from the leader's point-of-view (who has control over y only) if the set of optimal solutions of (1.1) does not reduce to a singleton. Throughout the paper it is assumed that all the functions F, G_j, f, g_i are sufficiently smooth.

Bilevel programming has many potential applications in very different fields, as transportation, economics, ecology, engineering and others, see Dempe (2003). One problem in chemistry from Dempe (2002) is the following. If a certain mixture y of chemical substances is put into a chemical reactor running with a certain temperature T and a given pressure p a chemical equilibrium x arises. This equilibrium depends on y, T, p but in general it cannot be influenced directly by the engineer. One can compute this equilibrium by solving a convex optimization problem

$$\begin{aligned} &\sum_{i=1}^{N} c_i(p,T)x_i + \sum_{i=1}^{G} x_i \ln \frac{x_i}{z} \to \min_x \\ &z = \sum_{j=1}^{G} x_j,\ Ax = \overline{A}y,\ x \geq 0, \end{aligned} \tag{1.3}$$

where $G \leq N$ denotes the number of gaseous and N the total number of reacting substances. Each row of the matrix A corresponds to a chemical element, each column to a substance. Hence, a column gives the amount of the different elements in the substances; x is the vector of the masses of the substances in the resulting chemical equilibrium whereas y denotes the initial masses of substances put into the reactor. The value of $c_i(p, T)$ gives the chemical potential of a substance which depends on the pressure p and on the temperature T. Let $x(p, T, y)$ denote the (unique) optimal solution of this problem. Then, since there exists some desire on the result of the chemical reactions, another problem arises. Namely, the question of how to determine the mixture of substances, the temperature and the pressure such that the resulting chemical equilibrium has a desired quality as containing a large amount of needed substances and a small amount of others. This goal can be expressed as another optimization problem whose feasible set depends on the optimal solution of the problem (1.3):

$$\begin{aligned} \langle c, x\rangle \quad &\to \min_{p,T,y} \\ (p,T,y) \in Y, \quad &x = x(p,T,y). \end{aligned}$$

This latter problem is the bilevel programming problem.

Bilevel programming has been the topic of a large number of papers including master's and PhD thesis, at least two monographs Bard (1998); Dempe (2002) and edited volumes Anandalingam and Friesz (1992); Migdalas et al. (1998), see the annotated bibliography Dempe

(2003). These publications present both theoretical results as well as solution approaches and a large number of applications.

2. Nonunique lower level optimal solutions

In the definition of the bilevel programming problem (1.2) quotation marks have been used to express the ambiguity in the definition of the problem in case of multiple optimal solutions in the lower level problem (1.1) for some parameter values. This is illustrated by the following example:

EXAMPLE 1.1 (LUCCHETTI AT AL. (1987)) *Let the lower level problem be given as*

$$\Psi(y) = \operatorname*{Argmin}_x \{-xy : 0 \le x \le 1\}$$

and consider the bilevel problem

$$\text{“}\min_y\text{”}\{x^2 + y^2 : x \in \Psi(y),\ 0 \le y \le 1\}.$$

Then, evaluating the lower level problem and inserting the optimal solution of this problem into the objective function of the upper level one results in

$$\Psi(y) = \begin{cases} \{0\}, & y > 0, \\ \{1\}, & y < 0, \\ [0,1], & y = 0. \end{cases} \qquad F(x(y), y) \begin{cases} = y^2, & y > 0, \\ = 1 + y^2, & y < 0, \\ \in [0,1], & y = 0. \end{cases}$$

Unclear is the function value of the function $y \mapsto F(x(y), y)$ at the point $y = 0$. The infimal function value of $F(x(y), y)$ is equal to zero but this value is attained only if $F(x(0), 0) = 0$. This situation is called the optimistic position in what follows. If this is not the case then the bilevel problem has no solution.

Note that in problem (1.2) the leader is not allowed to force the follower to take the one or the other of his optimal solutions. Hence, the leader cannot predict the true value of his objective function until the follower has communicated his choice. To overcome this ambiguity at least two approaches have been suggested.

2.1 The optimistic position

The first one is the optimistic position. The leader can use this approach if he supposes that the follower is willing to support him, i.e. that the follower will select a solution $x(y) \in \Psi(y)$ which is a best one from the point-of-view of the leader. Let

$$\varphi_o(y) := \min_x \{F(x,y) : x \in \Psi(y)\} \tag{1.4}$$

denote the optimistic objective function value in the upper level. Then, the optimistic position of the bilevel programming problem reduces to

$$\min\{\varphi_o(y) : G(y) \leq 0\}. \tag{1.5}$$

A pair $(x(\overline{y}), \overline{y})$ such that $\overline{y}$ solves (1.5) and $x(\overline{y})$ solves (1.4) for $y = \overline{y}$ is an optimal solution for the problem

$$\min_{x,y}\{F(x,y) : G(y) \leq 0,\ x \in \Psi(y)\}, \tag{1.6}$$

too, and vice versa, provided that the latter one has an optimal solution. Most of the contributions to bilevel programming treat this problem. One of the reasons can be that this problem has an optimal solution under quite reasonable assumptions.

To formulate an existence result for this problem a regularity condition is needed. For a problem

$$\min\{\alpha(x) : \beta(x) \leq 0,\ \gamma(x) = 0\} \tag{1.7}$$

with equality and inequality constraints ($\alpha,\ \beta_i,\ \gamma_j : \mathbb{R}^n \to \mathbb{R},\ i = 1,\ldots,p,\ j = 1,\ldots,q$) the *Mangasarian-Fromowitz constraint qualification* at a point x^0 reads as

(MFCQ) there exists a direction d with

$$\begin{array}{l} \nabla\beta_i(x^0)d < 0\ \forall\ i : \beta_i(x^0) = 0, \\ \nabla\gamma_j(x^0)d = 0\ \forall\ j \end{array}$$

and the gradients $\{\nabla\gamma_j(x^0)\ \forall\ j\}$ are linearly independent.

THEOREM 1.2 (DEMPE (2002)) *If*

(C) *the set $\{(x,y) : g(x,y) \leq 0, G(y) \leq 0\}$ is non-empty and compact*

and, for each y with $G(y) \leq 0$, the (MFCQ) is satisfied, then problem (1.6) has an optimal solution.

Related results can be found in Harker and Pang (1988).

2.2 The pessimistic position

The optimistic position seems not to be possible without any trouble at least in the cases when cooperation is not allowed respectively not possible or when the follower's seriousness of keeping the agreement is not granted. Then, one way out of this unpleasant situation for the leader is to bound the damage resulting from an undesirable selection of the follower. This leads to the problem

$$\min\{\varphi_p(y) : G(y) \leq 0\}, \tag{1.8}$$

where

$$\varphi_p(y) := \max_x \{F(x, y) : x \in \Psi(y)\} \tag{1.9}$$

denotes the worst upper level objective function value achievable on the set of optimal solutions of the lower level problem.

THEOREM 1.3 (DEMPE (2002)) *Let the point-to-set mapping $\Psi(\cdot)$ be lower semicontinuous at all points y with $G(y) \leq 0$ and let assumption (C) be satisfied. Then, problem (1.8) has an optimal solution.*

Here, the point-to-set mapping $\Psi(\cdot)$ is lower semicontinuous at a point $y = y^0$ if, for each open set A with $\Psi(y^0) \cap A \neq \emptyset$ there is an open neighborhood B of y^0 with $\Psi(y) \cap A \neq \emptyset$ for all $y \in B$.

Similar results can be found in the paper Lucchetti at al. (1987). Note that the assumptions in this theorem are much more restrictive than in Theorem 1.2. The bilevel programming problem has an implicitly determined set of feasible solutions and, at least looking at (1.5) or (1.8), it is a nonconvex optimization problem with a nondifferentiable objective function. Hence local optimal solutions may occur. To define the notion of a locally optimal solution or of a stationary solution usual tools from optimization applied to the problems (1.5) or (1.8) can be used.

DEFINITION 1.4 *A point (x^0, y^0) with $G(y^0) \leq 0$ and $x^0 \in \Psi(y^0)$ is called a locally optimistic (pessimistic) optimal solution of the bilevel programming problem (1.1), (1.2) if*

$$F(x^0, y^0) \leq F(x, y^0) \quad (\textit{resp. } F(x^0, y^0) \geq F(x, y^0)) \ \forall\, x \in \Psi(y^0)$$

and there is an $\varepsilon > 0$ such that

$$\varphi_o(y^0) \leq \varphi_o(y) \quad (\textit{resp. } \varphi_p(y^0) \leq \varphi_p(y)) \ \forall\, y : G(y) \leq 0,\ \|y - y^0\| < \varepsilon.$$

2.3 Weak solutions

If the lower level problem is nonconvex (for fixed parameter value y) the computation of a globally optimal solution in the lower level problem can be computationally intractable (especially tracing the global solution set mapping for varying parameter). In this case it can be considered as being helpful to modify the bilevel problem such that a locally optimal solution in the lower level problem is searched for instead of a globally optimal solution. But, as it is shown in an example in Vogel (2002), this can completely change the existence of an optimal solution of the bilevel programming problem.

Example 1.5 (Vogel (2002)) *Consider the bilevel problem*

$$\text{“}\min_y\text{”}\{(x+1)^2 : -3 \le y \le 2,\ x \in \Psi_s(y)\},$$

with the lower level problem

$$\min_x\{x^3 - 3x : x \ge y\}.$$

Then, if $\Psi_s(y)$ denotes the set of global optimal solutions of the last problem and if the optimistic approach is used, then an optimal solution is $y^* = -2$, an optimal solution in the pessimistic approach does not exist. But, if $\Psi_s(y)$ denotes the set of locally optimal solutions of the lower level problem the following results are obtained:

$$\varphi_o^s(y) := \min\{F(x,y) : x \in \Psi_s(y)\} = \begin{cases} (y+1)^2, & \text{if } y \in [-3,-1) \cup [1,2] \\ 4, & \text{if } y \in [-1,1) \end{cases}$$

and

$$\varphi_p^s(y) := \max\{F(x,y) : x \in \Psi_s(y)\} = \begin{cases} (y+1)^2, & \text{if } y \in (1,2] \\ 4, & \text{if } y \in [-3,1] \end{cases}$$

Hence, $\inf \varphi_o^s(y) = 0$ and an optimal solution of this problem does not exist. On the other hand, $\inf \varphi_p^s(y) = 4$ and all points $y \in [-3,1]$ are optimal solutions. The reason for this behavior is it that the point-to-set mapping of locally optimal solutions of a parametric optimization problem is generally not upper semicontinuous if the convexity assumption is dropped in the assumptions of Theorem 1.2.

To circumvent this unpleasant situation, Vogel (2002) has defined a weaker notion of an optimistic and a pessimistic solutions. For this, let the point-to-set mapping $\overline{\Psi}_s := \text{cl}\ \Psi_s$ be defined via the closure of the graph of the point-to-set mapping Ψ_s:

$$\text{grph}\ \overline{\Psi}_s := \text{cl grph}\ \Psi_s.$$

Consider the problem

$$\text{“}\min_y\text{”}\{F(x,y) : G(y) \le 0,\ x \in \Psi_s(y)\} \tag{1.10}$$

and define

$$
\begin{aligned}
\overline{\varphi}_o(y) &:= \inf_{x\in\Psi_s(y)} F(x,y)\\
\overline{F}_o &:= \inf_{y:G(y)\le 0} \varphi_o(y)\\
\overline{\varphi}_p(y) &:= \sup_{x\in\Psi_s(y)} F(x,y)\\
\overline{F}_p &:= \inf_{y:G(y)\le 0} \varphi_p(y),
\end{aligned}
$$

where Ψ_s is a point-to-set mapping defined by "solutions" of the lower level problem (e.g. local optimal solutions or global optimal solutions or stationary points).

Definition 1.6 (Vogel (2002)) *Consider the problem (1.10).*

1 *A point $\overline{y}$ with $G(\overline{y}) \le 0$ is a weak optimistic solution of the bilevel programming problem if there is some $\overline{x} \in \overline{\Psi}_s(\overline{y})$ such that $F(\overline{x},\overline{y}) = \overline{F}_o$.*

2 *A point $\overline{y}$ with $G(\overline{y}) \le 0$ is a weak pessimistic solution of the bilevel programming problem if there is some $\overline{x} \in \overline{\Psi}_s(\overline{y})$ and a sequence $\{y^k\}_{k=1}^{\infty} \subset \operatorname{dom}\Psi_s$ such that $\lim_{k\to\infty} y^k = \overline{y}$ and $\lim_{k\to\infty} \overline{\varphi}_p(y^k) = \overline{F}_p$ as well as $F(\overline{x},\overline{y}) = \overline{F}_p$.*

The definitions of weak optimistic and pessimistic solutions are different since, if a weak optimistic solution exists, it can be shown that the additional property similar to the one formulateded for the pessimistic solution is satisfied. This additional property is needed to guarantee that the obtained solution is not an isolated one which is far from all feasible solutions for slightly perturbed problems.

Theorem 1.7 (Vogel (2002)) *Let $\{(x,y) : x \in \Psi_s(y), G(y) \le 0\}$ be nonempty and bounded. Then, the bilevel programming problem (1.10) has a weak optimistic and a weak pessimistic solutions.*

The assumption of this theorem is satisfied if $\Psi_s(y)$ denotes the set of globally or locally optimal solutions or the set of Fritz John points and also if it denotes the set of generalized critical points (Guddat et al. (1990)) of the lower level problem (1.1).

3. Relations to other problems

The bilevel programming problem is closely related to other optimization problems which is often used to solve this problem. In the papers

Audet et al. (1997); Frangioni (1995) it is shown that every mixed-discrete optimization problem can be formulated as bilevel programming problem. This of course implies $\mathcal{NP}$–hardness of bilevel programming. The latter is also shown in

THEOREM 1.8 (DENG (1998)) *For any $\varepsilon > 0$ it is $\mathcal{NP}$–hard to find a feasible solution to the linear bilevel programming problem with no more than ε times the optimal value.*

Related results can also be found in Hansen et al. (1992).

The relations to bicriterial optimization have been investigated e.g. in the papers Fliege and Vicente (2003); Haurie et al. (1990); Marcotte and Savard (1991). On the one hand it is easy to see that at least one feasible point of the bilevel programming problem (1.1), (1.6) is Pareto optimal for the problem

$$\begin{gathered}\left.\begin{array}{c}F(x,y)\\ f(x,y)\end{array}\right\} \to \text{“}\min_{x,y}\text{”}\\ G(y) \le 0, g(x,y) \le 0.\end{gathered}$$

But this, in general, is not true for a (local) optimal solution of the bilevel problem. Hence, attempts to solve the bilevel programming problem via bicriterial optimization with the ordering cone $\mathbb{R}^2_+$ will in general not work. On the other hand, Fliege and Vicente (2003) shows that bicriterial optimization can indeed be used to prove optimality for the bilevel programming problem. But, for doing so, another more general ordering cone has to be used. Closely related to bilevel programming problems are also the problems of minimizing a function over the efficient set of some multicriterial optimization problem (see Fülöp (1993); Muu (2000)).

One tool often used to reformulate the bilevel programming problem as an one-level problem are the Karush-Kuhn-Tucker conditions. If a regularity condition is satisfied for the lower level problem (1.1), then the Karush-Kuhn-Tucker conditions are necessary optimality conditions. They are also sufficient in the case when (1.1) is a convex optimization problem in the x–variables for fixed parameters y. This suggests to replace problem (1.1), (1.6) by

$$\begin{aligned}&F(x,y) \to \min_{x,y,\lambda}\\ &G(y) \le 0\\ &\nabla_x f(x,y) + \lambda^\top \nabla_x g(x,y) = 0\\ &g(x,y) \le 0,\ \lambda \ge 0,\ \lambda^\top g(x,y) = 0.\end{aligned} \tag{1.11}$$

The relations between (1.1), (1.4), (1.5) and (1.11) are highlighted in the following theorem.

THEOREM 1.9 (DEMPE (2002)) *Consider the optimistic bilevel programming problem (1.1), (1.4), (1.5) and assume that, for each fixed y, the lower level problem (1.1) is a convex optimization problem for which (MFCQ) is satisfied for each fixed y and all feasible points. Then, each local optimal solution for the problem (1.1), (1.4),(1.5) corresponds to a local optimal solution for problem (1.11).*

Note that the opposite implication is not true in general. This can be seen in the following example.

EXAMPLE 1.10 *Consider the simple linear bilevel programming problem*

$$\text{“}\min_y\text{”}\{y : x \in \Psi(y),\ -1 \leq y \leq 1\},$$

where

$$\Psi(y) := \operatorname*{Argmin}_x \{xy : 0 \leq x \leq 1\}$$

at the point $(x, y) = (0, 0)$. Then, this point is a local minimum of problem (1.11), i.e. there exists an open neighborhood $W_\varepsilon(0,0) = (-\varepsilon, \varepsilon) \times (-\varepsilon, \varepsilon)$ with $0 < \varepsilon < 1$ such that $y \geq 0$ for all $(x, y) \in W_\varepsilon(0,0)$ with $x \in \Psi(y)$ and $-1 \leq y \leq 1$. The simple reason for this is it that there is no $-\varepsilon < x < \varepsilon$ with $x \in \Psi(y)$ for $y < 0$ since $\Psi(y) = \{1\}$ for $y < 0$. But a closer look at the definition of a local optimistic optimal solution shows that $y^0 = 0$ is not a local optimistic optimal solution since it is not a local minimum of the function $\varphi_o(y) = y$.

It is a first implication of these considerations that the problems (1.1), (1.4), (1.5) and (1.1), (1.6) are not equivalent if local optimal solutions are considered and a second one that not all local optimal solutions of the problem (1.11) correspond in general to local optimal solutions of the problem (1.1), (1.4), (1.5). It should be noted that, under the assumptions of Theorem 1.9 and if the optimal solutions of the lower level problem are strongly stable in the sense of Kojima (1980), then problems (1.1), (1.2) and (1.11) are equivalent.

The following example from Mirrlees (1999) shows that this result is no longer valid if the convexity assumption is dropped.

EXAMPLE 1.11 *Consider the problem*

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

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

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

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

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

which has three solutions for $0.344 \le y \le 2.903$. *The global optimum of the lower level problem is uniquely determined for all* $y \neq 1$ *and it has a jump at the point* $y = 1$. *Here the global optimum of the lower level problem can be found at the points* $x = \pm 0.957$. *The point* $(x^0; y^0) = (0.957; 1)$ *is also the global optimum of the 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: $(x, y) = (0.895; 1.99)$, $(x, y) = (0.42; 2.19)$, $(x, y) = (-0.98; 1.98)$ *are obtained. Surprisingly, the global optimal solution of the bilevel problem is not obtained with this approach. The reason for this is it that the problem*

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

has a much larger feasible set than the bilevel problem. And this feasible set has no jump at the point $(x, y) = (0.957; 1)$ *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.*

A surprising result is also the dependence of bilevel programming problems on irrelevant constraints, cf. Macal and Hurter (1997). This means that dropping some lower level constraints which are not active at an optimal solution (x^0, y^0) of the bilevel programming problem can change the problem drastically such that (x^0, y^0) will not remain optimal.

Moreover, the location of constraints is essential. Moving one constraint from the lower to the upper levels (or vice versa) generally changes the problem significantly; it is even possible that one of the problems has an optimal solution whereas the other has not.

4. Optimality conditions

4.1 Implicit functions approach

The formulation of optimality conditions for bilevel programming problems usually starts with a suitable reformulation of the problem as a one-level one. First conditions are based on strong stability of optimal solutions for problem (1.1) and replace the implicit constraint $x \in \Psi(y)$ by the implicitly determined function $x = x(y)$ with $\{x(y)\} = \Psi(y)$.

Let $L(x, y, \lambda) := f(x, y) + \lambda^\top g(x, y)$ denote the Lagrange function for problem (1.1) and

$$\Lambda(x, y) := \{\lambda \ge 0 : \nabla_x L(x, y, \lambda) = 0,\ g(x, y) \le 0,\ \lambda^\top g(x, y) = 0\}$$

denote its set of regular Lagrange multipliers.

THEOREM 1.12 (KOJIMA (1980)) *Consider problem (1.1) and let x^0 be a locally optimal solution of this problem at $y = y^0$. Assume that (MFCQ) and*

(SSOC) *for all $\lambda^0 \in \Lambda(x^0, y^0)$ and for all $d \neq 0$ with*

$$\nabla_x g_i(x^0, y^0)d = 0 \ \forall i \in J(\lambda^0) := \{j : \lambda^0_j > 0\}$$

the inequality $d^\top \nabla^2_{xx} L(x^0, y^0, \lambda^0)d > 0$ holds

are satisfied. Then, the solution x^0 is strongly stable, i.e. there exist open neighborhoods U of x^0 and V of y^0 and a uniquely determined function $x : V \to U$ being the unique locally optimal solution of (1.1) in U for all $y \in V$.

If the assumptions of Theorem 1.12 are satisfied for $x^0 \in \Psi(y^0)$ and problem (1.1) is a convex optimization problem for fixed y then, problem (1.1), (1.2) can locally equivalently be replaced with

$$\min_y \{F(x(y), y) : G(y) \leq 0\}. \tag{1.12}$$

Using the chain rule, necessary and sufficient optimality conditions can now be derived, provided it is possible to compute, say, a directional derivative for the function $x(y)$ in the point y^0.

THEOREM 1.13 (RALPH AND DEMPE (1995)) *Consider problem (1.1) and let x^0 be a locally optimal solution of this problem at $y = y^0$. Assume that (MFCQ), (SSOC) together with*

(CRCQ) *there exists an open neighborhood W of (x^0, y^0) such that, for each subset $I \subseteq I(x^0, y^0) := \{j : g_j(x^0, y^0) = 0\}$ the family of gradient vectors $\{\nabla_x g_i(x, y) : i \in I\}$ has the same rank on W*

are satisfied at (x^0, y^0). Then, the by Theorem 1.12 locally uniquely determined function $x(y)$ is directionally differentiable at the point $y = y^0$ with the directional derivative in direction r being the unique optimal solution of the problem

$$0.5 d^\top \nabla^2_{xx} L(x^0, y^0, \overline{\lambda})d + d^\top \nabla^2_{yx} L(x^0, y^0, \overline{\lambda})r \to \min_d$$

$$\nabla_x g_i(x^0, y^0) \begin{cases} = 0, & \text{if } i \in J(\overline{\lambda}) \\ \leq 0, & \text{if } i \in I(x^0, y^0) \setminus J(\overline{\lambda}) \end{cases}$$

for all $\overline{\lambda} \in \operatorname*{Argmax}_\lambda \{\nabla_y L(x^0, y^0, \lambda)r : \lambda \in \Lambda(x^0, y^0)\}$.

Denote the directional derivative of $x(y)$ at $y = y^0$ in direction r by $x'(y^0; r)$. Then, necessary and sufficient optimality conditions for the bilevel programming problem can be formulated.

THEOREM 1.14 (DEMPE (1992)) *Consider the problem (1.1), (1.2) at a point (x^0, y^0) with $G(y^0) \le 0,\ x^0 \in \Psi(y^0)$ and let (MFCQ), (SSOC) and (CRCQ) be satisfied for the lower level problem. Moreover assume that (1.1) is a convex optimization problem parametrized in y. Then,*

1 if y^0 is a locally optimal solution of this problem then the following optimization problem has the optimal objective function value zero:

$$\begin{array}{rcl} \alpha & \to & \min\limits_{\alpha,r} \\ \nabla_x F(x^0,y^0)x'(y^0;r) + \nabla_y F(x^0,y^0)r & \le & \alpha \\ \nabla G_i(y^0) & \le & \alpha,\ \forall\ i : G_i(y^0) = 0 \\ \|r\| & \le & 1. \end{array} \tag{1.13}$$

2 If the optimal function value v of the problem

$$\begin{array}{rcl} \nabla_x F(x^0,y^0)x'(y^0;r) + \nabla_y F(x^0,y^0)r & \to & \min\limits_{r} \\ \nabla G_i(y^0) & \le & 0,\ \forall\ i : G_i(y^0) = 0 \\ \|r\| & = & 1. \end{array}$$

is greater than zero ($v > 0$) then y^0 is a strict local optimal solution of the problem (1.1), (1.2), i.e. for each $0 < z < v$ there is $\varepsilon > 0$ such that

$$F(x,y) \ge F(x^0,y^0) + z\|y - y^0\|$$

for all (x,y) with $G(y) \le 0,\ x \in \Psi(y),\ \|y - y^0\| \le \varepsilon$.

Put $\mathcal{F}(y) := F(x(y), y)$. If the (MFCQ) is satisfied at the point y^0 for the problem (1.12) then the necessary optimality condition of first order in Theorem 1.14 means that

$$\mathcal{F}'(y^0; r) \ge 0\ \forall\ r \text{ satisfying } \nabla G_i(y^0)r \le 0,\ i : G_i(y^0) = 0.$$

This property is usually called Bouligand stationarity (or B-stationarity) of the point y^0.

4.2 Using the KKT conditions

If the Karush-Kuhn-Tucker conditions are applied to replace the lower level problem by a system of equations and inequalities, problem (1.11) is obtained. The Example 1.11 shows that it is possible to obtain necessary optimality conditions for the bilevel programming problem by this approach only in the case when the lower level problem is a convex parametric one and also only using the optimistic position. But even in this case this is not so easy since the familiar regularity conditions are not satisfied for this problem.

THEOREM 1.15 (SCHEEL AND SCHOLTES (2000)) *For problem (1.11) the Mangasarian-Fromowitz constraint qualification (MFCQ) is violated at every feasible point.*

To circumvent the resulting difficulties for the construction of Karush-Kuhn-Tucker type necessary optimality conditions for the bilevel programming problem, in Scheel and Scholtes (2000) a nonsmooth version of the KKT reformulation of the bilevel programming problem in the optimistic case is constructed:

$$\begin{array}{rcl} F(x,y) & \to & \min\limits_{x,y,\lambda} \\ G(y) & \le & 0 \\ \nabla_x L(x,y,\lambda) & = & 0 \\ \min\{-g(x,y),\lambda\} & = & 0. \end{array} \tag{1.14}$$

Here, for $a, b \in \mathbb{R}^n$, the formula $\min\{a, b\} = 0$ is understood component wise.

For problem (1.14) the following generalized variant of the linear independence constraint qualification can be defined (Scholtes and Stöhr (2001)):

(PLICQ) The *piecewise linear independence constraint qualification* is satisfied for the problem (1.14) at a point (x^0, y^0, λ^0) if the gradients of all the vanishing components of the constraint functions $G(y), \nabla_x L(x, y, \lambda), g(x, y), \lambda$ are linearly independent.

Problem (1.14) can be investigated by considering the following patchwork of nonlinear programs for fixed sets I:

$$\begin{array}{rcl} F(x,y) & \to & \min\limits_{x,y,\lambda} \\ G(y) & \le & 0 \\ \nabla_x L(x,y,\lambda) & = & 0 \\ g_i(x,y) & = & 0, \text{ for } i \in I \\ g_i(x,y) & \le & 0, \text{ for } i \notin I \\ \lambda_i & = & 0, \text{ for } i \notin I \\ \lambda_i & \ge & 0, \text{ for } i \in I. \end{array} \tag{1.15}$$

Then, the piecewise linear independence constraint qualification is valid for problem (1.14) at some point (x^0, y^0, λ^0) if and only if it is satisfied for each of the problems (1.15) for all sets $J(\lambda^0) \subseteq I \subseteq I(x^0, y^0)$.

The following theorem says that the (PLICQ) is generically satisfied. For this define the set

$$\mathcal{H}_B^l = \{ \ (F, G, f, g) \in C^l(\mathbb{R}^{m+n}, \mathbb{R}^{1+s+1+p}) : \ \text{(PLICQ) is satisfied at each feasible point of (1.14) with } \|\lambda\|_\infty \leq B\}$$

for an arbitrary constant $0 < B < \infty$, $\|\lambda\|_\infty = \max\{|\lambda_i| : 1 \leq i \leq p\}$ is the L_∞-norm of a vector $\lambda \in \mathbb{R}^p$ and $l \geq 2$.

THEOREM 1.16 (SCHOLTES AND STÖHR (2001)) *For $2 \leq k \leq l$, the set $\mathcal{H}_B^l$ is open in the C^k-topology (Hirsch (1994)). Moreover, for $l > m$, the set $\mathcal{H}_B^l$ is also dense in the C^k-topology for all $2 \leq k \leq l$.*

Now, after this excursion to regularity, the description of necessary optimality conditions for the bilevel programming problem with convex lower level problems using the optimistic position is continued. For the origin of the following theorem for mathematical programs with equilibrium constraints see Scheel and Scholtes (2000). There a relaxation of problem (1.11) is considered:

$$\begin{array}{rcl} F(x,y) & \rightarrow & \min\limits_{x,y,\lambda,\mu} \\ \nabla_x L(x,y,\lambda,\mu) & = & 0 \\ g_i(x,y) & = & 0 \text{ for } g_i(x^0,y^0) = 0 \\ \lambda_i & = & 0 \text{ for } \lambda_i^0 = 0 \\ g_i(x,y) & \leq & 0 \text{ for } g_i(x^0,y^0) < 0 \\ \lambda_i & \geq & 0 \text{ for } \lambda_i^0 > 0 \\ G(y) & \leq & 0. \end{array} \tag{1.16}$$

In the following theorem, the following more restrictive regularity condition than (MFCQ) is needed

(SMFCQ) The *strict Mangasarian-Fromowitz constraint qualification* (SMFCQ) is satisfied at x^0 for problem (1.7) if there exists a Lagrange multiplier (λ, μ),

$$\lambda \geq 0, \lambda^\top \beta(x^0) = 0, \nabla \alpha(x^0) + \lambda^\top \nabla \beta(x^0) + \mu^\top \nabla \gamma(x^0) = 0,$$

as well as a direction d satisfying

$$\begin{array}{ll} \nabla \alpha_i(x^0)d < 0, & \text{for each } i \text{ with } \beta_i(x^0) = \lambda_i = 0, \\ \nabla \beta_i(x^0)d = 0, & \text{for each } i \text{ with } \lambda_i > 0, \\ \nabla \gamma_j(x^0)d = 0, & \text{for each } j \end{array}$$

and $\{\nabla \beta_i(x^0) : \lambda_i > 0)\} \cup \{\nabla_x \gamma_j(x^0) : j = 1, \ldots, q\}$ are linearly independent.

Note that this condition is implied by (PLICQ).

THEOREM 1.17 *Let (x^0, y^0, λ^0) be a local minimizer of problem (1.14) and use $z^0 = (x^0, y^0)$.*

- *If the (MFCQ) is valid for problem (1.16) at (x^0, y^0, λ^0), then there exist multipliers $(\kappa, \omega, \zeta, \xi)$ satisfying*

$$\begin{aligned}\nabla F(z^0) + \kappa^\top(0, \nabla_y G(y^0)) + \nabla(\nabla_x L(z^0, \lambda^0)\omega) + \zeta^\top \nabla g(z^0) &= 0\\ \nabla_x g(z^0)\omega - \xi &= 0\\ g_i(z^0)\zeta_i &= 0,\ \forall i\\ \lambda_i^0 \xi_i &= 0,\ \forall i\\ \zeta_i \xi_i &\geq 0,\ i \in K\\ \kappa^\top G(y^0) &= 0,\\ \kappa &\geq 0,\end{aligned}$$

where $K = \{i : g_i(x^0, y^0) = \lambda_i^0 = 0\}$.

- *If the (SMFCQ) is fulfilled for the problem (1.16), then there exist unique multipliers $(\kappa, \omega, \zeta, \xi)$ solving the last system of equations and inequalities with $\zeta_i \xi_i \geq 0,\ i \in K$ being replaced by*

$$\zeta_i \geq 0, \xi_i \geq 0,\ i \in K.$$

For related optimality conditions see e.g. Flegel and Kanzow (2002); Flegel and Kanzow (2003).

5. Solution algorithms

5.1 Implicit function approach

Again, to solve the bilevel programming problem, it is reformulated as a one-level problem. The first approach again uses the implicit determined solution function of the convex lower level problem $x(y)$ provided this function is uniquely determined. If the assumptions (C), (MFCQ), (SSOC), and (CRCQ) are satisfied at every point y with $G(y) \leq 0$, then the resulting problem

$$\min\{F(x(y), y) : G(y) \leq 0\}$$

has an objective function being piecewise continuously differentiable (see Ralph and Dempe (1995)). The pieces of the solution function $x(y)$ are obtained by replacing some of the active inequalities $g_i(x, y) \leq 0, i \in \overline{I}$ in the lower level problem by equations $g_i(x, y) = 0, i \in \overline{I}$, where

$$J(\lambda^0) := \{i : \lambda_i^0 > 0\} \subseteq \overline{I} \subseteq I(x(y^0), y^0) := \{j : g_j(x(y^0), y^0) = 0\}$$

and λ^0 is a Lagrange multiplier vector in the lower level problem corresponding to the optimal solution $x(y^0)$ for $y = y^0$. If the constraints $g_i(x, y) \leq 0$ in problem (1.1) are locally replaced by $g_i(x, y) = 0,\ i \in \overline{I}$, the resulting lower level problems are

$$\min_x \{f(x, y) : g_i(x, y) = 0,\ \forall\ i \in \overline{I}\}. \tag{1.17}$$

If the gradients $\{\nabla_x g_i(x(y^0), y^0) : i \in \overline{I}\}$ are moreover linearly independent (which can be guaranteed for small sets $\overline{I} \supset J(\lambda^0)$ with λ^0 being a vertex in $\Lambda(x(y^0), y^0)$), then the optimal solution function $x^{\overline{I}}(\cdot)$ of the problem (1.17) is differentiable (Fiacco (1983)). Let $\mathcal{I}$ denote the family of all index sets determined by the above two demands for all vertices $\lambda^0 \in \Lambda(x(y^0), y^0)$.

Theorem 1.18 (Dempe and Pallaschke (1997)) *Consider problem (1.1) at the point $(x(y^0), y^0)$ and let (MFCQ), (SSOC) and (CRCQ) be satisfied there. If the condition*

(FRR) *For each vertex $\lambda^0 \in \Lambda(x(y^0), y^0)$ the matrix*

$$\begin{pmatrix} \nabla^2_{xx} L(x(y^0), y^0, \lambda^0) & \nabla_x^\top g_{J(\lambda^0)}(z^0) & \nabla^2_{yx} L(x(y^0), y^0, \lambda^0) \\ \nabla_x g_{I(x(y^0), y^0)}(x(y^0), y^0) & 0 & \nabla_y g_{I(x(y^0), y^0)}(x(y^0), y^0) \end{pmatrix}$$

has full row rank $n + |I(x(y^0), y^0)|$

is valid, then the generalized derivative of the function $x(\cdot)$ at the point $y = y^0$ in the sense of Clarke (1983) is

$$\partial x(y^0) = \mathrm{conv} \bigcup_{I \in \mathcal{I}} \nabla x^I(y^0).$$

Using this formula, a bundle algorithm (cf. Outrata at al. (1998)) can be derived to solve the resulting problem (1.12).

Since the full description of bundle algorithms is rather lengthy, the interested reader is referred e.g. to Outrata at al. (1998). Repeating the results in Schramm (1989) (cf. also Outrata at al. (1998)) the following result is obtained:

Theorem 1.19 (Dempe (2002)) *If the assumptions (C), (MFCQ), (CRCQ), (SSOC), and (FRR) are satisfied for the convex lower level problem (1.1) at all points (x, y), $x \in \Psi(y)$, $G(y) \equiv 0$, and the sequence of iteration points $\{(x(y^k), y^k, \lambda^k)\}_{k=1}^\infty$ in the bundle algorithm remains bounded, then this algorithm computes a sequence $\{(x(y^k), y^k \lambda^k)\}_{k=1}^\infty$ having at least one accumulation point $(x(y^0), y^0, \lambda^0)$ with*

$$0 \in \nabla_x F(x(y^0), y^0) \partial x(y^0) + \nabla_y F(x(y^0), y^0).$$

If assumption (FRR) is not satisfied, then the point $(x(y^0), y^0)$ is pseudostationary in the sense of Mikhalevich et al. (1987).

Hence, under suitable assumptions the bundle algorithm computes a Clarke stationary point. Such points are in general not Bouligand stationary.

5.2 A smoothing method

If the lower level problem (1.1) is a convex prarametric problem for which, at every feasible parameter value y with $G(y) \leq 0$, a constraint qualification is satisfied, then the optimistic problem (1.6) can be replaced equivalently by the problem (1.11). By Theorem 1.9 every optimistic optimal solution of the bilevel programming problem correspondes to an optimal solution of each of the problems (1.11). To solve this problem several authors (e.g. Fukushima and Pang (1999)) use a NCP function approach to replace the complementarity constraints. This results in the nondifferentiable problem

$$\begin{array}{rcl} F(x,y) & \to & \min\limits_{x,y,\lambda} \\ G(y) & \leq & 0 \\ \nabla_x L(x,y,\lambda) & = & 0 \\ \Phi(-g_i(x,y),\lambda_i) & = & 0, \ \forall\, i = 1,\ldots,p, \end{array} \tag{1.18}$$

where a function $\Phi(\cdot,\cdot)$ satisfying

$$\Phi(a,b) = 0 \Longleftrightarrow a \geq 0,\ b \geq 0,\ ab = 0$$

is called a NCP function. Examples and properties of NCP functions can be found in the book of Geiger and Kanzow (2003). NCP functions are inherently nondifferentiable, and algorithms solving problem (1.18) use smoothed NCP functions. Fukushima and Pang (1999) use the function

$$\Phi_\varepsilon(a,b) = a + b - \sqrt{a^2 + b^2 + \varepsilon},\ \varepsilon > 0$$

and solve the resulting problems

$$\begin{array}{rcl} F(x,y) & \to & \min\limits_{x,y,\lambda} \\ G(y) & \leq & 0 \\ \nabla_x L(x,y,\lambda) & = & 0 \\ \Phi_\varepsilon(-g_i(x,y),\lambda_i) & = & 0, \ \forall\, i = 1,\ldots,p, \end{array} \tag{1.19}$$

for $\varepsilon \to 0$ with suitable standard algorithms. Hence selecting an arbitrary sequence $\{\varepsilon_k\}_{k=1}^\infty$ they compute a sequence of solutions

$\{(x^k, y^k, \lambda^k)\}_{k=1}^{\infty}$ and investigate the properties of the accumulation points of this sequence.

To formulate their convergence result, the assumption of week nondegeneracy is needed. To formulate this assumption consider the Clarke derivative of the function $\Phi(-g_i(x,y), \lambda_i)$. This Clarke derivative exists and is contained in the set

$$C_i(x,y,\lambda) = \{r = -\xi_i(\nabla g_i(x,y),0)+\chi_i(0,0,1) : (1-\xi_i)^2+(1-\chi_i)^2 \leq 1\}.$$

Let the point $(\overline{x}, \overline{y}, \overline{\lambda})$ be an accumulation point of the sequence $\{(x^k, y^k, \lambda^k)\}_{k=1}^{\infty}$. It is then easy to see that, for each i such that $g_i(\overline{x}, \overline{y}) = \overline{\lambda}_i = 0$ any accumulation point of the sequence

$$\{\nabla \Phi_{\varepsilon_k}(-g_i(x^k, y^k), \lambda_i^k)\}_{k=1}^{\infty}$$

belongs to $C_i(\overline{x}, \overline{y}, \overline{\lambda})$, hence is of the form

$$r = -\xi_i(\nabla g_i(\overline{x}, \overline{y}), 0) + \chi_i(0,0,1)$$

with $(1-\xi_i)^2 + (1-\chi_i)^2 \leq 1$. Then, it is said that the sequence $\{(x^k, y^k, \lambda^k)\}_{k=1}^{\infty}$ is asymptotically weakly nondegenerate, if in this formula neither ξ_i nor χ_i vanishes for any accumulation point of $\{(x^k, y^k, \lambda^k)\}_{k=1}^{\infty}$. Roughly speaking this means that both $g_i(x^k, y^k)$ and λ_i^k approach zero in the same order of magnitude (see Fukushima and Pang (1999)).

Theorem 1.20 (Fukushima and Pang (1999)) *Let for each point (x^k, y^k, λ^k) the necessary optimality conditions of second order for problem (1.19) be satisfied. Suppose that the sequence $\{(x^k, y^k, \lambda^k)\}_{k=1}^{\infty}$ converges to some $(\overline{x}, \overline{y}, \overline{\lambda})$ for $k \to \infty$. If the (PLICQ) holds at the limit point and the sequence $\{(x^k, y^k, \lambda^k)\}_{k=1}^{\infty}$ is asymptotically weakly nondegenerate, then $(\overline{x}, \overline{y}, \overline{\lambda})$ is a Bouligand stationary solution for problem (1.11).*

5.3 SQP methods

In recent times several authors have reported (in view of the violated regularity condition rather surprisingly) a good behavior of SQP methods for solving mathematical programs with equilibrium constraints (see Anitescu (2002); Fletcher et al. (2002); Fletcher and Leyffer (2002)).

To scetch these results consider a bilevel programming problem (1.6) with a convex parametric lower level problem (1.1) and assume that a regularity assumption is satisfied for each fixed parameter value y with $G(y) \leq 0$. Then, by Theorem 1.9, a locally optimal solution of the bilevel programming problem corresponds to a locally optimal for the problem

(1.11). Consequently, in order to compute local minima of the bilevel problem, problem (1.11) can be solved.

In doing this, Anitescu (2002) uses the elastic mode approach in a sequantial quadratic programming algorithm solving (1.11). This means that if a quadratic programming problem minimizing a quadratic approximation of the objective function of problem (1.11) subject to a linear approximation of the constraints of this problem has a feasible solution with bounded Lagrange multipliers then the solution of this problem is used as a search direction. And if not, a regularized quadratic programming problem is used to compute this search direction.

For simplicity, this idea is described for problem (1.7). Then this means that the following problem is used to compute this seach direction:

$$\begin{aligned}\nabla\alpha(x)d + d^\top W d &\to \min_d \\ \beta_i(x) + \nabla\beta(x)d \le 0,\ &\forall\ i = 1,\ldots,p \\ \gamma_j(x) + \nabla\gamma(x)d = 0,\ &\forall\ j = 1,\ldots,q.\end{aligned}$$

Here, W can be the Hessian matrix of the Lagrange function of the problem (1.7) or another positive definite matrix approximating this Hessian. If this problem has no feasible solution or unbounded Lagrange multipliers the solution of problem (1.7) (or accordingly the solution process for the problem (1.11)) with the sequential quadratic programming approach is replaced by the solution of the following problem by the same approach:

$$\min_{x,\zeta}\{\alpha(x)+c\zeta : \beta_i(x) \le \zeta, \forall\, i = 1,\ldots,p,\ -\zeta \le \gamma_j(x) \le \zeta, \forall\, j = 1,\ldots,q\},$$

where c is a sufficiently large constant. This is the elastic mode SQP method.

To implement the idea of Anitescu (2002) assume that the problem (1.16) satisfies the (SMFCQ) and that the *quadratic growth condition* at a point $x = x^0$

(QGC) There exists $\sigma > 0$ satisfying

$$\max\{\alpha(x) - \alpha(x^0), \beta_i(x),\ \forall\ i, |\gamma_j(x)|,\ \forall\ j\} \ge \sigma\|x - x^0\|$$

for all x in some open neighborhood of x^0

is valid for problem (1.11) at a locally optimal solution of this problem.

THEOREM 1.21 (ANITESCU (2002)) *If the above two assumtions are satisfied then the elastic mode sequantial quadratic programming algorithm computes a locally optimal solution of the problem (1.11) provided it is started sufficiently close to that solution and the constant c is sufficiently large.*

Using stronger assumptions Fletcher et al. (2002) have even been able to prove local Q-quadratic convergence of sequential quadratic programming algorithms to solutions of (1.11).

6. Discrete bilevel programming

If discreteness demands are added to the lower or upper levels of a bilevel programming problem the investigation becomes more difficult and the number of references is rather small, see Dempe (2003). With respect to the existence of optimal solutions the location of the discreteness demand is important Vicente et al. (1996). Most difficult is the situation when the lower level problem is a parametric discrete one and the upper level problem is a continuous problem. Then the graph of the solution set mapping $\Psi(\cdot)$ is in general neither closed nor open. The situation is more or less similar to the continuous case if the lower level problem is a continuous parametric optimization problem or if the discreteness demands are situated in both levels.

One way to solve discrete optimization problems (and also bilevel programming problems) is branch-and-bound. Here a second difficulty for the solution of discrete bilevel programming problems appeares: the usual fathomimg procedure becomes wrong, see Moore and Bard, (1990). If the integrality conditions in both levels are dropped at the beginning and are introduced via the branching procedure, then a global optimal solution of the relaxed problem, which occasionally proves to be feasible for the bilevel problem is in general not an optimal solution for the bilevel programming problem.

Known solution methods include one being based on the investigation of the solution set mapping if the lower level problem is a right-hand side parametrized Boolean knapsack problem and another one using cutting planes in the discrete lower level problem with parameters in the objective function only (see Dempe (2002)).

To describe a third approach consider a linear bilevel programming problem with upper level discreteness demands only:

$$
\begin{array}{l}
c_1^\top x + d_1^\top y \to \max\limits_{x,y} \\
A_1 x \le b_1, x \ge 0, \text{ integer} \\
\text{where } y \text{ solves} \\
\qquad d_2^\top y \to \max_y \\
\qquad A_2 x + B_2 y = b_2 \\
\qquad y \ge 0
\end{array}
\tag{1.20}
$$

Then, an idea of White and Anandalingam (1993) can be used to transform this problem into a mixed discrete optimization problem. For this, apply the Karush-Kuhn-Tucker conditions to the lower level problem. This transforms problem (1.20) into

$$\begin{aligned} & c_1^\top x + d_1^\top y \to \max_{x,y,\lambda} \\ & A_1 x \leq b_1, x \geq 0, \text{ integer} \\ & B_2^\top \lambda \geq b_2, \\ & A_2 x + B_2 y = b_2 \\ & y \geq 0, \ y^\top (B_2^\top \lambda - d_2) = 0. \end{aligned} \tag{1.21}$$

Now use a penalty function approach to get rid of the complementarity constraint resulting in the problem

$$\begin{aligned} & c_1^\top x + d_1^\top y - K y^\top (B_2^\top \lambda - d_2) \to \max_{x,y,\lambda} \\ & A_1 x \leq b_1, x \geq 0, \text{ integer} \\ & A_2 x + B_2 y = b_2, y \geq 0 \\ & B_2^\top \lambda \geq b_2. \end{aligned} \tag{1.22}$$

By application of the results in White and Anandalingam (1993) the following is obtained:

Theorem 1.22 *Assume that problem (1.22) has an optimal solution for some positive K_0. Then, the problem (1.22) describes an exact penalty function approach for problem (1.20), i.e. there is a number K^* such that the optimal solutions of the problems (1.22) and (1.20) for all $K \geq K^*$ coincide.*

This idea has been used in Dempe and Kalashnikov, (2002) to solve an application problem in gas industry. Moreover, the implications of a movement of the discreteness condition from the lower to the upper level problems has been touched there.

7. Conclusion

In the paper a survey of results in bilevel programming has been given. It was not the intention of the author to give a detailled description of one or two results but rather to give an overview over different directions of research and to describe some of the challenges of this topic. Since bilevel programming is a very living area a huge number of questions remain open. These include optimality conditions as well as solution algorithms for problems with nonconvex lower level problems, discrete bilevel programming problems in every context, many questions related to the investigation of pessimistic bilevel programming problems to call only some of them. Also, one implication from $\mathcal{NP}$-hardness often used in theory is it that such problems should also be solved with approximation algorithms which, if possible, should be complemented by a bound on the accuracy of the computed solution. One example for such an approximation algorithm can be found in Marcotte (1986) but in general the description of such algorithms is a challenging task for future research.

References

G. Anandalingam and T. Friesz (eds.). *Hierarchical optimization.* Annals of Operations Research, vol. 24, 1992.

M. Anitescu. *On solving mathematical programs with complementarity constraints as nonlinear programs.* Technical Report Nr. ANL/NCS-P864-1200, Department of Mathematics, University of Pittsburgh, 2002.

C. Audet, P. Hansen, B. Jaumard and G. Savard. Links between linear bilevel and mixed 0-1 programming problems, *Journal of Optimization Theory and Applications*, 93: 273-300, 1997.

J.F. Bard. *Practical Bilevel Optimization: Algorithms and Applications.* Kluwer Academic Publishers, Dordrecht, 1998.

F.H. Clarke. *Optimization and Nonsmooth Analysis.* John Wiley & Sons, New York, 1983.

S. Dempe. A necessary and a sufficient optimality condition for bilevel programming problems. *Optimization*, 25: 341–354, 1992.

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

S. Dempe. Annotated Bibliography on Bilevel Programming and Mathematical Programs with Equilibrium Constraints. *Optimization*, 52: 333-359, 2003.

S. Dempe and V. Kalashnikov. *Discrete bilevel programming: application to a gas shipper's problem.* Preprint Nr. 2002-02, TU Bergakademie Freiberg, Fakultät für Mathematik und Informatik, 2002.

S. Dempe and D. Pallaschke. Quasidifferentiability of optimal solutions in parametric nonlinear optimization. *Optimization*, 40: 1-24, 1997.

X. Deng. Complexity issues in bilevel linear programming. In: A. Migdalas, P.M. Pardalos and P. Värbrand (eds.), *Multilevel Optimization: Algorithms and Applications*, pp. 149-164, Kluwer Academic Publishers, Dordrecht, 1998.

A. V. Fiacco. *Introduction to Sensitivity and Stability Analysis in Nonlinear Programming.* Academic Press, New York, 1983.

M. L. Flegel and C. Kanzow. *Optimality conditions for mathematical programs with equilibrium constraints: Fritz John and Abadie-type approaches.* Report, Universität Würzburg, Germany, 2002.

M. L. Flegel and C. Kanzow. A Fritz John approach to first order optimality conditions for mathematical programs with equilibrium constraints. *Optimization*, 52: 277-286, 2003.

R. Fletcher and S. Leyffer. *Numerical experience with solving MPECs as NLPs.* Numerical Analysis Report Nr. NA/210, Department of Mathematics, University of Dundee, UK, 2002.

R. Fletcher, S. Leyffer, D. Ralph and S. Scholtes. *Local convergence of SQP methods for mathematical programs with equilibrium constraints.* Numerical Analysis Report NA/209, Department of Mathematics, University of Dundee, UK, 2002.

J. Fliege and L. N. Vicente. *A bicriteria approach to bilevel optimization*, Technical Report, Fachbereich Mathematik, Universität Dortmund, Germany, 2003.

A. Frangioni. On a new class of bilevel programming problems and its use for reformulating mixed integer problems, *European Journal of Operational Research*, 82: 615-646, 1995.

M. Fukushima and J.-S. Pang. Convergence of a smoothing continuation method for mathematical programs with complementarity constraints. In: M. Thera and R. Tichatschke (eds.), *Ill-posed Variational Problems and Regularization Techniques.* Number 477 in Lecture Notes in Economics and Mathematical Systems. Berlin et al.: Springer, 1999.

J. Fülöp. *On the equivalence between a linear bilevel programming problem and linear optimization over the efficient set.* Working Paper No. WP 93-1, Laboratory of Operations Research and Decision Systems, Computer and Automation Institute, Hungarian Academy of Sciences, 1993.

C. Geiger and C. Kanzow. *Theorie und Numerik restringierter Optimierungsaufgaben.* Berli et al.: Springer, 2002.

J. Guddat, F. Guerra Vasquez and H.Th. Jongen. *Parametric Optimization: Singularities, Pathfollowing and Jumps.* John Wiley & Sons, Chichester and B.G. Teubner, Stuttgart, 1990.

P. Hansen, B. Jaumard and G. Savard. New branch-and-bound rules for linear bilevel programming. *SIAM Journal on Scientific and Statistical Computing.* 13: 1194–1217, 1992.

P.T. Harker and J.-S. Pang. Existence of optimal solutions to mathematical programs with equilibrium constraints. *Operations Research Letters*, 7: 61-64, 1988.

A. Haurie, G. Savard and D. White. A note on: an efficient point algorithm for a linear two-stage optimization problem. *Operations Research*, 38: 553–555, 1990.

M. W. Hirsch. *Differential Topology*, Springer-Verlag, Berlin et al., 1994.

M. Kojima. Strongly stable stationary solutions in nonlinear programs. In: S.M. Robinson (ed.), *Analysis and Computation of Fixed Points*, pp. 93–138, Academic Press, New York, 1980.

R. Lucchetti, F. Mignanego and G. Pieri, Existence theorem of equilibrium points in Stackelberg games with constraints. *Optimization*, 18: 857-866, 1987.

C.M. Macal and A.P. Hurter. Dependence of bilevel mathematical programs on irrelevant constraints. *Computers and Operations Research*, 24: 1129-1140, 1997.

P. Marcotte. Network design problem with congestion effects: a case of bilevel programming. *Mathematical Programming*, 34: 142–162, 1986.

P. Marcotte and G. Savard. A note on the Pareto optimality of solutions to the linear bilevel programming problem. *Computers and Operations Research*, 18: 355–359, 1991.

A. Migdalas, P.M. Pardalos and P. Värbrand (eds.) *Multilevel Optimization: Algorithms and Applications*. Kluwer Academic Publishers, Dordrecht, 1998.

V. S. Mikhalevich and A. M. Gupal and V. I. Norkin. *Methods of Nonconvex Optimization*. Nauka, Moscow, 1987 (in Russian).

J. A. Mirrlees. The theory of moral hazard and unobservable bevaviour: part I. *Review of Economic Studies*, 66: 3-21, 1999.

J. Moore and J. F. Bard. The mixed integer linear bilevel programming problem. *Operations Research*, 38: 911–921, 1990.

L. D. Muu. On the construction of initial polyhedral convex set for optimization problems over the efficient set and bilevel linear programs. *Vietnam Journal of Mathematics*, 28: 177-182, 2000.

J. Outrata and M. Kočvara and J. Zowe. *Nonsmooth Approach to Optimization Problems with Equilibrium Constraints*. Kluwer Academic Publishers, Dordrecht, 1998.

D. Ralph and S. Dempe. Directional Derivatives of the Solution of a Parametric Nonlinear Program. *Mathematical Programming*, 70: 159-172, 1995.

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

S. Scholtes and M. Stöhr. How stringent is the linear independence assumption for mathematical programs with stationarity constraints?. *Mathematics of Operations Research*, 26: 851-863, 2001.

H. Schramm. *Eine Kombination von bundle- und trust-region-Verfahren zur Lösung nichtdifferenzierbarer Optimierungsprobleme*. Bayreuther Mathematische Schriften, Bayreuth, No. 30, 1989.

L. N. Vicente, G. Savard and J. J. Judice. The discrete linear bilevel programming problem. *Journal of Optimization Theory and Applications*, 89: 597-614, 1996.

S. Vogel. *Zwei-Ebenen-Optimierungsaufgaben mit nichtkonvexer Zielfunktion in der unteren Ebene: Pfadverfolgung und Sprünge*. PhD thesis, Technische Universität Bergakademie Freiberg, 2002.

D. J. White and G. Anandalingam. A penalty function approach for solving bi-level linear programs. *Journal of Global Optimization*, 3: 397-419, 1993.