

Bilevel programming: A survey

Benoît Colson¹, Patrice Marcotte², and Gilles Savard³

¹ Department of Mathematics, The University of Namur, 5000 Namur, BELGIUM*

² Département d'Informatique et de Recherche Opérationnelle, Université de Montréal, C.P. 6128 Succursale Centre-ville Montreal (QC), CANADA H3C 3J7
(e-mail: marcotte@iro.umontreal.ca)

³ Département de mathématiques et de génie industriel, Ecole Polytechnique de Montréal, C.P. 6079 Succursale Centre-ville Montreal (QC), CANADA H3C 3A7
(e-mail: gilles.savard@polymtl.ca)

Received: April 2005 / Revised version: May 2005

Abstract. This paper provides an introductory survey of a class of optimization problems known as bilevel programming. We motivate this class through a simple application, and then proceed with the general formulation of bilevel programs. We consider various cases (linear, linear-quadratic, nonlinear), describe their main properties and give an overview of solution approaches.

Key words: Bilevel programming, Mathematical programs with equilibrium constraints, Nonlinear programming, Optimal pricing

MSC classification: 90C05, 90C11, 90C20, 90C27, 90C3D, 65K05

1 Introduction

This paper is dedicated to two-level – or *bilevel* – optimization and our purpose is to provide the reader with the key concepts and solution methods associated with this particular class of hierarchical mathematical programs.

Although a wide range of applications fit the bilevel programming framework (see das and Pardalos(2001) or das and Pardalos(2001)l, for domains of application), real-life implementations are scarce, due mainly to the lack of efficient algorithms for tackling large-scale problems. This state of affairs motivated us to use as an

* *Present address:* SAMTECH s.a., Liège, Rue des Chasseurs Ardennais, B-4031 Liège, Belgium
(e-mail: benoit.colson@samcef.com)

introductory example one exception that has been proposed for modelling revenue maximization situations (see é et al.(2003)Côté, Marcotte, and Savardo é et al.(2003)Côté, Marcotte, and Savard^).

Our example lies in the framework of transportation modelling and more precisely it concerns a *toll-setting problem*. The latter arises from a situation where an authority or the owners of a highway system are allowed to set tolls on (a subset of) the links of the network. Taking into account that network users wish to minimize their travel costs, an optimal toll schedule will be such that toll levels are not too high – otherwise the users may be deterred from using the infrastructure – though still generating “large” revenues. Once the network managers have set tolls, travellers react to these values and select their itinerary in such a way that total travel cost, i.e. standard costs (time, distance, etc.) plus tolls, is minimized. An important feature of this problem – and more generally of bilevel programs – is the hierarchical relationship between two classes of decision makers. This will be reflected in the mathematical formulation that we derive below.

As previously stated, the network manager’s objective is to maximize revenue. Denoting by \mathcal{A} the set of links of the network and by $\bar{\mathcal{A}}$ the subset of tolled links, this corresponds to the mathematical program

$$\max_{T,x} \sum_{a \in \bar{\mathcal{A}}} T_a x_a \quad (1.1a)$$

$$\text{s.t. } l_a \leq T_a \leq u_a \quad \forall a \in \bar{\mathcal{A}}, \quad (1.1b)$$

where T_a and x_a denote the toll and the flow on link a respectively, and l_a (respectively u_a) is a lower (respectively upper) bound on the toll¹.

The situation of network users is modelled by means of an equilibrium in terms of a path-flow vector f , as in classical tools for network analysis and traffic flow assignment (see e.g., fi(1985)e fi(1985)h, for more details). In the simplest situation, e.g., in a congestion-free environment, such *user equilibrium* coincides with a flow assignment that minimizes total system cost. It follows that the path-flow vector f , together with the link-flow vector x , is solution of the linear program:

$$\min_{f,x} \sum_{a \in \mathcal{A}} c_a x_a + \sum_{a \in \bar{\mathcal{A}}} T_a x_a \quad (1.2a)$$

$$\text{s.t. } \sum_{p \in \mathcal{P}_{rs}} f_p^{rs} = d_{rs} \quad \forall (r, s) \in \Theta, \quad (1.2b)$$

$$x_a = \sum_{(r,s) \in \Theta} \sum_{p \in \mathcal{P}_{rs}} \delta_{a,p}^{rs} f_p^{rs} \quad \forall a \in \mathcal{A}, \quad (1.2c)$$

$$f_p^{rs} \geq 0 \quad \forall p \in \mathcal{P}_{rs}, \quad \forall (r, s) \in \Theta. \quad (1.2d)$$

¹ While it seems natural to have $l_a = 0$, it is sometimes advantageous to set tolls to negative values. This corresponds to subsidies.

The objective (1.2a) is the sum of the cost resulting from tolls T_a ($a \in \bar{\mathcal{A}}$) and the other costs (duration, length, etc.) which are aggregated in a measure c_a for each link. Constraint (1.2b) expresses demand satisfaction in the sense that, for a given origin-destination pair (r, s) (the set of all such pairs is denoted by Θ), the sum of the flows f_p^{rs} on all paths p connecting r to s (these paths being regrouped in \mathcal{P}_{rs}) equals the travel demand, d_{rs} , from r to s . Constraint (1.2c) links path flows f_p^{rs} and link flows x_a , where

$$\delta_{a,p}^{rs} = \begin{cases} 1 & \text{if path } p \in \mathcal{P}_{rs} \text{ uses link } a, \\ 0 & \text{otherwise.} \end{cases}$$

Clearly, mathematical programs (1.1a) and (1.2a) are connected since both of them use the same set of variables, namely the tolls T_a ($a \in \bar{\mathcal{A}}$) and the flows x_a ($a \in \mathcal{A}$). Also, the profit of the network manager (see (1.1a)) cannot be computed until the flows are known, these flows not being in the direct control of the manager, but the solution of a mathematical program parameterized in the toll vector T . This yields the bilevel formulation²

$$\begin{aligned} \max_{T, f, x} \quad & \sum_{a \in \bar{\mathcal{A}}} T_a x_a \\ \text{s.t.} \quad & l_a \leq T_a \leq u_a \quad \forall a \in \bar{\mathcal{A}}, \\ (f, x) \in \arg \min_{f', x'} \quad & \sum_{a \in \mathcal{A}} c_a x'_a + \sum_{a \in \bar{\mathcal{A}}} T_a x'_a \\ \text{s.t.} \quad & \sum_{p \in \mathcal{P}_{rs}} f_p'^{rs} = d_{rs} \quad \forall (r, s) \in \Theta, \\ & x'_a = \sum_{(r,s) \in \Theta} \sum_{p \in \mathcal{P}_{rs}} \delta_{a,p}^{rs} f_p'^{rs} \quad \forall a \in \mathcal{A}, \\ & f_p'^{rs} \geq 0 \quad \forall p \in \mathcal{P}_{rs}, \quad \forall (r, s) \in \Theta. \end{aligned}$$

The hierarchical relationship is here reflected in that the mathematical program related to the users' behaviour is part of the manager's constraints. This is the major feature of bilevel programs: they include two mathematical programs within a single instance, one of these problems being part of the constraints of the other one. Because of this hierarchical relationship, the program (1.1) is called the *upper-level problem* while (1.2) corresponds to the *lower-level problem*.

We will now leave the particular framework of toll-setting problems (the interested reader is referred to e.g., [é et al.\(1998\)](#) [Labbé, Marcotte, and Savard](#) [é](#)

² In the sequel, we will simply write down the upper and lower level problems, dispensing with the "prime" and "arg min" notation. The resulting "vertical" format is indeed less heavy and more transparent.

et al.(1998)Labbé, Marcotte, and Savarda, corne et al.(2001)Brotcorne, Labbé, Marcotte, and Savardo corne et al.(2001)Brotcorne, Labbé, Marcotte, and Savardr, for further details). The next section describes bilevel programs from a more general point of view. This will be followed by a survey of existing methods for solving various types of bilevel programs (Sect. 3). Mathematical programs with equilibrium constraints, which are very similar to bilevel programs, will be the subject of Sect. 4. We will conclude this paper with a review of some perspectives and challenges for future research in the field of bilevel programming.

2 General formulation and basic concepts

The general formulation of a bilevel programming problem (BLPP) is

$$\min_{x \in X, y} F(x, y) \quad (2.1a)$$

$$\text{s.t. } G(x, y) \leq 0 \quad (2.1b)$$

$$\min_y f(x, y) \quad (2.1c)$$

$$\text{s.t. } g(x, y) \leq 0, \quad (2.1d)$$

where $x \in \mathbb{R}^{n_1}$ and $y \in \mathbb{R}^{n_2}$. The variables of problem (2.1) are divided into two classes, namely the *upper-level variables* $x \in \mathbb{R}^{n_1}$ and the *lower-level variables* $y \in \mathbb{R}^{n_2}$. Similarly, the functions $F : \mathbb{R}^{n_1} \times \mathbb{R}^{n_2} \rightarrow \mathbb{R}$ and $f : \mathbb{R}^{n_1} \times \mathbb{R}^{n_2} \rightarrow \mathbb{R}$ are the *upper-level* and *lower-level objective functions* respectively, while the vector-valued functions $G : \mathbb{R}^{n_1} \times \mathbb{R}^{n_2} \rightarrow \mathbb{R}^{m_1}$ and $g : \mathbb{R}^{n_1} \times \mathbb{R}^{n_2} \rightarrow \mathbb{R}^{m_2}$ are called the *upper-level* and *lower-level constraints* respectively. Upper-level constraints involve variables from both levels (in contrast with the constraints specified by the set X) and play a very specific role. Indeed, they must be enforced indirectly, as they do not bind the lower-level decision-maker.

The previous section illustrated the usefulness of bilevel programs through a specific example. Several other transportation issues may be modelled by bilevel programs – see alas(1995)g (i) for a review – but, more generally, real-world problems involving a hierarchical relationship between two decision levels are frequently encountered in fields such as management (facility location, environmental regulation, credit allocation, energy policy, hazardous materials), economic planning (social and agricultural policies, electric power pricing, oil production), engineering (optimal design, structures and shape), chemistry, environmental sciences, optimal control, etc. For instance, the upper level may represent decision-makers whose policies lead to some reaction within a particular market or social entity, the latter corresponding to the lower level of the system under study.

From a historical point of view, multilevel optimization is closely related to the economic problem of kelberg(1952)a (t) in the field of game theory, which we briefly describe now. To this end, we consider an economic planning process involving interacting agents at two distinct levels: some of the individuals – collectively called

the *leader* – issue directives to the remaining agents – called the *followers*. In the particular framework of Stackelberg games, the leader is assumed to anticipate the reactions of the followers; this allows him to choose his best – or *optimal* – strategy accordingly. More precisely, the leader chooses a strategy x in a set $X \subseteq \mathbb{R}^n$, and every follower i has a strategy set $Y_i(x) \subseteq \mathbb{R}^{m_i}$ corresponding to each $x \in X$. The sets $Y_i(x)$ are assumed to be closed and convex. Any follower i also has a cost function depending on both the leader’s and *all* followers’ strategies and which may be expressed as

$$\theta_i(x, \cdot) : \prod_{j=1}^M \mathbb{R}^{m_j} \rightarrow \mathbb{R},$$

where M is the number of followers. It is further assumed that for fixed values of $x \in X$ and y_j ($j \neq i$) the function θ_i is convex and continuously differentiable in $y_i \in Y_i(x)$. The followers behave collectively according to the *noncooperative principle* of (1951)s (a) which means that, for each $x \in X$, they will choose a joint response vector

$$y^{opt} \equiv (y_i^{opt})_{i=1}^M \in C(x),$$

where $C(x) = \prod_{i=1}^M Y_i(x)$, such that, for every $i = 1, \dots, M$, there holds

$$y_i^{opt} \in \operatorname{argmin}\{\theta_i(x, y_i, y_{j \neq i}^{opt}) : y_i \in Y_i(x)\}.$$

In the above setting, considered by ali et al.(1983)Sherali, Soyster, and Murphye (h) in an oligopolistic situation, Stackelberg problems possess a hierarchical structure similar to that of BLPP, although the lower-level program is an equilibrium rather than an optimization problem. This class of problems will be discussed in more details in Sect. 4.

Bilevel programs were initially considered by Bracken and McGill in a series of papers – see ken and McGill(1973)a (r), ken and McGill(1974)a (r), ken and McGill(1978)a (r) – that dealt with applications in the military field as well as in production and marketing decision making. By that time, such problems were called *mathematical programs with optimization problems in the constraints*, which exactly reflects the situation formulated in (2.1), the terms *bilevel* and *multilevel programming* being introduced later by ler and Norton(1977)n (a). Notice however that the problems studied in the latter paper did not involve joint upper-level constraints, that is, constraints depending on both x and y . To our knowledge, the general formulation with $G(x, y) \leq 0$ as upper-level constraints first appeared in shi and Shimizu(1981)y (i).

We now return to problem (2.1) to introduce some further concepts of bilevel programming. The *relaxed* problem associated with (2.1) is

$$\begin{aligned} & \min_{x \in X, y} F(x, y) \\ & \text{s.t. } G(x, y) \leq 0, \\ & g(x, y) \leq 0, \end{aligned} \tag{2.2}$$

and its optimal value is a lower bound for the optimal value of (2.1). The *relaxed feasible region* (or *constraint region*) is

$$\Omega = \{(x, y) \in \mathbb{R}^{n_1} \times \mathbb{R}^{n_2} : x \in X, G(x, y) \leq 0 \text{ and } g(x, y) \leq 0\}.$$

For a given (fixed) vector $\bar{x} \in X$, the *lower-level feasible set* is defined by

$$\Omega(\bar{x}) = \{y \in \mathbb{R}^{n_2} : g(\bar{x}, y) \leq 0\}$$

while the *lower-level reaction set*³ (or *rational reaction set*) is

$$R(\bar{x}) = \{y \in \mathbb{R}^{n_2} : y \in \operatorname{argmin} \{f(\bar{x}, \hat{y}) : \hat{y} \in \Omega(\bar{x})\}\}.$$

Every $y \in R(\bar{x})$ is a *rational response*. For a given x , $R(x)$ is an implicitly defined multi-valued function of x that may be empty for some values of its argument. Finally, the set

$$\mathcal{IR} = \{(x, y) \in \mathbb{R}^{n_1} \times \mathbb{R}^{n_2} : x \in X, G(x, y) \leq 0, y \in R(x)\},$$

that regroups the feasible points of the BLPP, corresponds to the feasible set of the leader, and is known as the *induced region* (or *inducible region*). This set is usually nonconvex and it can even be disconnected or empty in presence of upper-level constraints.

We conclude this section with a short discussion on two modelling approaches to bilevel programming. In the case of *optimistic bilevel programming*, it is assumed that, whenever the reaction set $R(x)$ is not a singleton, the leader is allowed to select the element in $\Omega(x)$ that suits him best. In this situation, a point $(x^*, y^*) \in \mathbb{R}^{n_1} \times \mathbb{R}^{n_2}$ is said to be a *local optimistic solution* for problem (2.1) if

$$\begin{aligned} x^* &\in X, \\ G(x^*, y^*) &\leq 0, \\ y^* &\in R(x^*), \\ F(x^*, y^*) &\leq F(x^*, y) \quad \text{for all } y \in R(x^*) \end{aligned}$$

and there exists an open neighbourhood $V(x^*; \delta)$ of x^* (with radius $\delta > 0$) such that

$$\phi_o(x^*) \leq \phi_o(x) \quad \text{for all } x \in V(x^*; \delta) \cap X,$$

where $\phi_o(x) = \min_y \{F(x, y) : y \in R(x)\}$. It is called a *global optimistic solution* if $\delta = \infty$ can be selected, corresponding to $V(x^*) = X$.

When cooperation of the leader and the follower is not allowed, or if the leader is risk-averse and wishes to limit the “damage” resulting from an undesirable selection

³ According to the definition of a bilevel program, the lower level problem must be solvable for global minima. In practice, the lower-level program is assumed to be convex.

of the follower, then a point $(x^*, y^*) \in \mathbb{R}^{n_1} \times \mathbb{R}^{n_2}$ is said to be a *local pessimistic solution* for problem (2.1) if

$$\begin{aligned} x^* &\in X, \\ G(x^*, y^*) &\leq 0, \\ y^* &\in R(x^*), \\ F(x^*, y^*) &\geq F(x^*, y) \quad \text{for all } y \in R(x^*) \end{aligned}$$

and there exists an open neighbourhood $V(x^*; \delta)$ of x^* (with $\delta > 0$) such that

$$\phi_p(x^*) \leq \phi_p(x) \quad \text{for all feasible } x \in V(x^*; \delta),$$

where this time $\phi_p(x) = \max_y \{F(x, y) : y \in R(x)\}$. It is called a *global pessimistic solution* if $\delta = \infty$ can be selected. Note that the difference between the optimistic and pessimistic approaches can also be explained from the follower viewpoint: the optimistic solution results from a friendly or cooperative behaviour while an aggressive follower produces a pessimistic solution. A more complete discussion of these issues may be found in dan and Morgan(1996)r (o) and e(2002)m (e).

3 A survey of existing methods

Although early work on bilevel programming dates back to the nineteen seventies, it was not until the early nineteen eighties that the usefulness of these mathematical programs in modelling hierarchical decision processes and engineering design problems prompted researchers to pay close attention to bilevel programs. A first bibliographical survey on the subject was written in 1985 by Kolstad. Bilevel programming problems being intrinsically difficult (see Sect. 3.1 below), it is not surprising that most algorithmic research to date has focused on the simplest cases of bilevel programs, that is problems having nice properties such as linear, quadratic or convex objective and/or constraint functions. In particular, the most studied instance of bilevel programming problems has been for a long time the *linear* BLPP – in which all functions are linear – and therefore this subclass is the subject of several dedicated surveys, such as those by nd Wen(1989)u (s), nd Hsu(1991)n (e) and Ayed(1993)n (e). Over the years, more complex bilevel programs were studied and even those including discrete variables received some attention, as in nte et al.(1996)Vicente, Savard, and Júdicec (i). Hence more general surveys appeared, such as those by rd(1989)v (a), dalingam and Friesz(1992)a (n) and nte and Calamai(1994)c (i). on(1999)l (o) deals with both nonlinear bilevel programming problems and mathematical programs with equilibrium constraints and recently e(2003)m (e) wrote an annotated bibliography on these same topics. The combinatorial nature of bilevel programming has been reviewed in otte and Savard(2005)r (a).

Following the proliferation of research devoted to bilevel programming, a number of dedicated textbooks have also been published in the late nineteen nineties.

Among them, those by Izu et al. (1997) Shimizu, Ishizuka, and Bardi (1998) and (1998) (a) are authored by some of the early protagonists in the field. Another monograph on the subject is that of Alas et al. (1997) Migdalas, Pardalos, and Värbrand (1998) (i), who consider the more general case of multilevel programming. The most recent book on bilevel programming, as of May 2005, is that of Ekeland (2002) (e).

3.1 Properties

Bilevel programming problems are intrinsically hard. Even the “simplest instance”, the *linear* BLPP, was shown to be \mathcal{NP} -hard by Slowinski (1985) (e), while Ekeland et al. (1992) Hansen, Jaumard, and Savard (1992) (a) proved strong \mathcal{NP} -hardness, using a reduction from KERNEL (see Yeh and Johnson (1979) (r) and Johnson (1979) (a)). Nte et al. (1994) Vicente, Savard, and Júdice (1994) (i) strengthened these results and proved that merely checking strict or local optimality is also \mathcal{NP} -hard, based on reductions from 3-SAT.

A number of authors have proposed *optimality conditions* for bilevel programming problems. Among them are those by Ekeland (1992a) (m) (e), Ekeland (1992b) (m) (e), and Florian (1991) (e) (h) and Ata (1993) (t) (u), who use tools from nonsmooth analysis (see e.g., Kiefer (1990) (a) Kiefer (1990) (l)). Rüdiger and Gauvin (1994) (v) (a) and Nte and Calamai (1995) (c) (i) developed optimality conditions taking the geometry of the induced region into account: the former is achieved by adapting the notion of *steepest descent* to the case of bilevel programs, while the latter generalizes first- and second-order optimality conditions to the case of bilevel programs with quadratic strictly convex lower-level problems. The main result in Nte and Calamai (1995) (c) (i) is that, at each point of the induced region, there exists a *finite* number of convex cones of induced region directions.

However, due to the inherent difficulty of manipulating the mathematical objects involved in all these optimality conditions, they have few practical uses and do not provide convenient stopping criteria for numerical algorithms.

3.2 Extreme-point approaches for the linear case

An important property of *linear* bilevel programs, i.e., programs where all functions involved are linear and the set X is polyhedral, is that their solution set, whenever it is nonempty, contains at least one vertex of the constraint region defined by the polyhedron

$$\Omega = \{(x, y) : x \in X, G(x, y) \leq 0 \text{ and } g(x, y) \leq 0\}.$$

Hence a wide class of methods for solving linear BLPPs is based on vertex enumeration.

The first method using such an approach was proposed by Leri and Townsley (1982) (n) (a) for solving BLPPs with no upper-level constraints and with unique

lower-level solutions. Their algorithm explores a decreasing number of bases of the lower-level problem, but was shown to be relatively slow in subsequent numerical tests. as and Karwan(1984)a (i) introduced the *K-th best* method, which considers bases of the relaxed problem (2.2) sorted in increasing order of upper-level objective function values. The method stops at the lowest index *K* corresponding to a *rational basis*. Such a basis is clearly globally optimal.

Similar vertex enumeration methods were introduced by vassilopoulos(1982)p (a), with the difference that all extreme points considered belong to the induced region *IR*, and that separation techniques are used to explore the adjacent vertices.

Related contributions are those by and Florian(1992)e (h), et al.(1992)Chen, Florian, and Wue (h) and t al.(1993)Tuy, Migdalas, and Värbrandy (u).

3.3 Branch-and-bound

When the lower-level problem is convex and regular, it can be replaced by its Karush-Kuhn-Tucker (KKT) conditions, yielding the single-level reformulation of problem (2.1):

$$\min_{x \in X, y, \lambda} F(x, y) \tag{3.1a}$$

$$\text{s.t. } G(x, y) \leq 0, \tag{3.1b}$$

$$g(x, y) \leq 0, \tag{3.1c}$$

$$\lambda_i \geq 0, i = 1, \dots, m_2 \tag{3.1d}$$

$$\lambda_i g_i(x, y) = 0, i = 1, \dots, m_2 \tag{3.1e}$$

$$\nabla_y \mathcal{L}(x, y, \lambda) = 0, \tag{3.1f}$$

where

$$\mathcal{L}(x, y, \lambda) = f(x, y) + \sum_{i=1}^{m_2} \lambda_i g_i(x, y)$$

is the Lagrangean function associated with the lower-level problem.

Even under suitable convexity assumptions on the functions *F*, *G* and the set *X*, the above mathematical program is not easy to solve, due mainly to the non-convexities that occur in the complementarity and Lagrangean constraints. While the Lagrangean constraint is linear in certain important cases (linear or convex quadratic functions), the complementarity constraint is intrinsically combinatorial, and is best addressed by enumeration algorithms, such as branch-and-bound.

In the branch-and-bound scheme, the root node of the tree corresponds to problem (3.1) from which constraint (3.1e) is removed. At a generic node of the branch-and-bound tree that does not satisfy the complementarity constraints, separation is performed in the following manner: two children nodes are constructed, one with $\lambda_i = 0$ as an additional constraint, and the other with the constraint $g_i(x, y) = 0$.

The optimal values of these problems yield lower bounds valid for the corresponding subtree.

In the absence of upper-level constraints, a rational solution can be computed by solving the lower-level problem resulting from setting x to the partial optimal solution of the relaxed problem. Note that, in contrast with standard branch-and-bound implementations, feasible (i.e., rational) solutions are then generated at *every* node of the implicit enumeration tree. The upper bound is updated accordingly.

Algorithms based on this idea were proposed by and Falk(1982)r (a) and uny-Amat and McCarl(1981)r (o) for solving linear bilevel programming problems. The approach was adapted by and Moore(1990)r (a) to linear-quadratic problems and by hayal et al.(1992)Al-Khayal, Horst, and Pardalos- (l), (1988)r (a) and nds and Bard(1991)m (d) to the quadratic case.

Combining branch-and-bound, monotonicity principles and penalties similar to those used in mixed-integer programming, en et al.(1992)Hansen, Jaumard, and Savardn (a) have developed a code capable of solving medium-sized linear bilevel programs⁴. i et al.(2002)Thoi, Yamamoto, and Yoshiseo (h) have developed a similar scheme for mathematical programs with linear complementarity constraints.

3.4 Complementary pivoting

The first approach using *complementary pivots* is that of as et al.(1980)Bialas, Karwan, and Shawa (i) for solving linear BLPPs. Their algorithm – named *Parametric Complementary Pivot (PCP) Algorithm* – is based on the reformulation (3.1) of a linear bilevel program using the KKT optimality conditions for the lower-level problem. At each iteration, the algorithm computes a feasible point (x, y) for the original problem such that the upper-level objective $F(x, y)$ takes a value at most equal to α , and where constraint (3.1f) is perturbed by adding a term εHy , where H is a negative definite matrix and ε is sufficiently small so that the solution to the original problem is not modified. The parameter α is updated after each iteration and the process is repeated until no feasible (x, y) can be found. However Ayed and Blair(1990)n (e) showed that this algorithm does not always converge to the optimal solution.

Let us also mention the contributions of ce and Faustino(1988)d (ú), ice and Faustino(1992)u (´), ce and Faustino(1994)d (ú), who introduced the so-called *sequential linear complementarity problem* (LCP) for solving linear and linear-quadratic bilevel programming problems. Note that their approach may actually be viewed as a combination of the techniques described in Sects. 3.2 and 3.3, namely vertex enumeration and branch-and-bound methods.

⁴ I.e., of the order of 100 variables and 100 constraints.

3.5 Descent methods

Assuming that, for any x , the optimal solution of the lower-level problem is unique and defines y as an implicit function $y(x)$ of x , problem (2.1) may be viewed solely in terms of the upper-level variables $x \in \mathbb{R}^{n_1}$. Given a feasible point x , an attempt is made to find a feasible (rational) direction $d \in \mathbb{R}^{n_1}$ along which the upper-level objective decreases. A new point $x + \alpha d$ ($\alpha > 0$) is computed so as to ensure a reasonable decrease in F while maintaining feasibility for the bilevel problem. However, a major issue is the availability of the gradient (or a sub-gradient) of the upper-level objective, $\nabla_x F(x, y(x))$, at a feasible point. Applying the chain rule of differentiation, we have, whenever $\nabla_x y(x)$ is well defined:

$$\nabla_x F(x, y(x)) = \nabla_x F(x, y) + \nabla_y F(x, y) \nabla_x y(x),$$

where the functions are evaluated at the current iterate. and Lasdon(1990)l (o) have proposed a method for approximating this gradient.

Another line of attack is that of rd and Gauvin(1994)v (a), for problems where no upper-level constraints are present and where the lower-level constraints are rewritten as:

$$\begin{aligned} g_i(x, y) &\leq 0, & i \in I, \\ g_j(x, y) &= 0, & j \in J. \end{aligned}$$

The authors first show that an upper-level descent direction at a given point x is a vector $d \in \mathbb{R}^{n_1}$ such that

$$\nabla_x F(x, y^*) d + \nabla_y F(x, y^*) w(x, d) < 0, \tag{3.2}$$

where $y^* = y(x)$ and $w \in \mathbb{R}^{n_2}$ is a solution of the program

$$\begin{aligned} \min_w & (d^T, w^T) \nabla_{xy}^2 \mathcal{L}(x, y^*, \lambda) (d, w) \\ \text{s.t.} & \nabla_y g_i(x, y^*) w \leq -\nabla_x g_i(x, y^*) d, \quad i \in I(x), \\ & \nabla_y g_j(x, y^*) w = -\nabla_x g_j(x, y^*) d, \quad j \in J, \\ & \nabla_y f(x, y^*) w = -\nabla_x f(x, y^*) d + \nabla_x \mathcal{L}(x, y^*, \lambda) d, \end{aligned} \tag{3.3}$$

with $I(x) = \{i \in I : g_i(x, y^*) = 0\}$ and

$$\mathcal{L}(x, y, \lambda) = f(x, y) + \sum_{i \in I(x) \cup J} \lambda_i g_i(x, y)$$

is the Lagrangean of the lower-level problem with respect to the active constraints. The steepest descent then coincides with the optimal solution of the *linear-quadratic bilevel program*

$$\begin{aligned} \min_d & \nabla_x F(x, y^*) d + \nabla_y F(x, y^*) w(x, d) \\ \text{s.t.} & \|d\| \leq 1, \\ & w(x, d) \text{ solves the quadratic program (3.3),} \end{aligned} \tag{3.4}$$

for which exact algorithms exist, such as those by and Moore(1990)r (a) or ard et al.(2000)Jaumard, Savard, and Xiongu (a).

Alternatively, nte et al.(1994)Vicente, Savard, and Júdece (i) proposed a descent method for convex quadratic bilevel programs, i.e., problems where both objectives are quadratic, and where constraints are linear. They extend the work of rd and Gauvin(1994)v (a) by solving problem (3.4) using the sequential LCP method of ce and Faustino(1994)d (ú), and propose a way to compute exact stepsizes. Motivated by the fact that checking local optimality in the sequential LCP approach is very difficult, nte et al.(1994)Vicente, Savard, and Júdece (i) have designed a hybrid algorithm using both the abovementioned features and a pivot step strategy that enforces the complementarity constraints.

Finally, let us mention the work of and Liu(1995)l (a), who present a bundle method where the decrease of the upper-level objective is monitored according to subdifferential information obtained from the lower-level problem. They call the resulting setup a *leader predominate algorithm*, according to the role played by the leader in the sequential decision making process.

3.6 Penalty function methods

Penalty methods constitute another important class of algorithms for solving *non-linear* BLPPs, although they are generally limited to computing stationary points and local minima.

An initial step in this direction was achieved by shi and Shimizu(1981)y (i), shi and Shimizu(1984)y (i) and izu and Aiyoshi(1981)i (h). Their approach consists in replacing the lower-level problem (2.1c)-(2.1d) by the penalized problem

$$\min_y p(x, y, r) = f(x, y) + r \phi(g(x, y)), \quad (3.5)$$

where r is a positive scalar, ϕ is a continuous penalty function that satisfies

$$\begin{aligned} \phi(g(x, y)) &> 0 && \text{if } y \in \text{int } S(x), \\ \phi(g(x, y)) &\rightarrow +\infty && \text{if } y \rightarrow \text{bd } S(x), \end{aligned} \quad (3.6)$$

and $\text{int } S(x)$ and $\text{bd } S(x)$ denote the relative interior and the relative boundary of $S(x) = \{y : g(x, y) \leq 0\}$, respectively. Problem (2.1) is then transformed into:

$$\begin{aligned} \min_{x \in X, y} F(x, y^*(x, r)) \\ \text{s.t. } G(x, y^*(x, r)) &\leq 0, \\ p(x, y^*(x, r), r) &= \min_y p(x, y, r). \end{aligned} \quad (3.7)$$

izu and Aiyoshi(1981)i (h) proved that the sequence $\{(x^k, y^*(x^k, r^k))\}$ of optimal solutions to (3.7) converges to the solution of (2.1). The main drawback of this

method is that solving (3.7) for a fixed value of r requires the global solution at every update of the upper-level variables. Each subproblem is not significantly easier to solve than the original bilevel program.

zuka and Aiyoshi(1992)h (s) proposed a double penalty method in which both objective functions (2.1a) and (2.1c) are penalized. They still use the augmented lower-level objective (3.5) and the penalty function ϕ characterized by (3.6) but replace the lower-level problem by its stationarity condition $\nabla_y p(x, y, r) = 0$, thus transforming (2.1) into the single-level program

$$\begin{aligned} \min_{x \in X, y} \quad & F(x, y) \\ \text{s.t.} \quad & G(x, y) \leq 0, \\ & \nabla_y p(x, y, r) = 0, \\ & g(x, y) \leq 0. \end{aligned} \tag{3.8}$$

Note that the last constraint restricts the domain of the function p . For a given r , problem (3.8) is solved using a second penalty function applied to the constraints.

A more recent contribution, by (1999)s (a), follows up on ideas of al.(1991)Bi, Calamai, and Conne (i), who themselves extend a technique proposed in al.(1989)Bi, Calamai, and Conne (i) for linear bilevel programs. Their approach is based on (3.1), that is, a bilevel program for which the lower-level problem has been replaced by its Karush-Kuhn-Tucker conditions. Their method involves a penalty function of the form

$$p(x, y, \lambda, \mu) = F(x, y) + \mu v(x, y, \lambda),$$

where μ is a positive *penalty parameter* and the upper-level objective $F(x, y)$ is augmented by a weighted, nonnegative penalty function associated with the current iterate. More precisely, (1999)s (a) builds a penalty function $v(x, y, \lambda)$ with respect to the ℓ_1 norm, defined as the sum of the terms associated with each constraint of the single-level problem (3.1). The resulting algorithm involves the minimization of the penalty function $p(x, y, \lambda, \mu)$ for a fixed value of μ . In view of the complex structure of the latter function, the authors develop a *trust-region method*, where the *model* for p (see Sect. 3.7) is obtained by replacing each component function of $p(x, y, \lambda, \mu)$ by its second-order Taylor expansion around the current iterate.

3.7 Trust-region methods

Trust-region algorithms are iterative methods based on the approximation of the original problem by a *model* around the current iterate. More specifically, let us consider the unconstrained problem

$$\min_x f(x).$$

Given the iterate x_k obtained at iteration k , one constructs a model m_k that approximates the objective function within a *trust region* usually defined as a ball (according to some norm) of radius Δ_k centered at x_k . The solution s_k to the *trust-region subproblem*

$$\begin{aligned} \min_s \quad & m_k(x_k + s) \\ \text{s.t.} \quad & \|s\| \leq \Delta_k \end{aligned}$$

is then computed. One then evaluates the quality of the model through the ratio of the *actual reduction* over the *predicted reduction*⁵

$$\rho_k = \frac{f(x_k) - f(x_k + s_k)}{m_k(x_k) - m_k(x_k + s_k)}. \quad (3.9)$$

If ρ_k is large enough ($\rho_k > \eta_2$ for some $0 < \eta_2 < 1$), the trial point is accepted as the next iterate ($x_{k+1} = x_k + s_k$), and the trust region radius may be increased. The trial point is also accepted for smaller values of ρ_k that satisfy the condition $\rho_k \in [\eta_1, \eta_2)$, with $0 < \eta_1 < \eta_2$, but in this case the trust-region radius is *not* increased (it might be decreased). Finally, if ρ_k is too small, the trial point is dismissed ($x_{k+1} = x_k$) and the trust-region radius is decreased. The rules for modifying Δ_k are as follows:

$$\Delta_{k+1} \in \begin{cases} [\Delta_k, \infty) & \text{if } \rho_k \geq \eta_2, \\ [\gamma_2 \Delta_k, \Delta_k] & \text{if } \rho_k \in [\eta_1, \eta_2), \\ [\gamma_1 \Delta_k, \gamma_2 \Delta_k] & \text{if } \rho_k < \eta_1, \end{cases}$$

where $0 < \gamma_1 \leq \gamma_2 < 1$ are predefined parameters. For an in-depth study and a comprehensive reference on trust-region methods we refer the reader to the monograph of et al.(2000)Conn, Gould, and Tointn (o).

A trust-region algorithm was recently developed by on et al.(2005)Colson, Marcotte, and Savardl (o) for solving nonlinear bilevel programs where the function G depends solely on the upper-level vector x . This is not the first attempt to solve bilevel programs by means of a trust-region methods. Indeed, a related approach has been proposed by t al.(1998)Liu, Han, and Wangu (i) for problems that do not involve upper-level constraints, and where the lower-level program is strongly convex and linearly constrained. Under suitable assumptions, convergence to a Clarke stationary point may be proved. No computational experience has been reported.

The algorithm in on et al.(2005)Colson, Marcotte, and Savardl (o) is an iterative method which, given the current iterate or incumbent solution (\bar{x}, \bar{y}) , is based on the *linear-quadratic bilevel* model

$$\min_{x \in X, y} F_m(x, y)$$

⁵ Note that, if the model is not accurate, there could be a deterioration of the objective, i.e., the ratio could be negative.

$$\begin{aligned}
& \text{s.t. } G_m(x) \leq 0 & (3.10) \\
& \min_y f_m(x, y) \\
& \text{s.t. } g_m(x, y) \leq 0,
\end{aligned}$$

of problem (2.1), where F_m, G_m and g_m are linear models of F, G, g at (\bar{x}, \bar{y}) respectively, while f_m is a quadratic model of f at (\bar{x}, \bar{y}) . The bilevel problem (3.10) thus defines the trust-region subproblem. This subproblem can be solved for its global solution either by using a specialized algorithm – e.g., ardet al.(2000)Jaumard, Savard, and Xiongu (a) –, either by reformulating it as a mixed integer program (see otte and Savard(2005)rotte and Savard(2005)a) and resorting to an off-the-shelf software.

Let (x^m, y^m) denote the solution of the subproblem, that may fail to be rational. In order to evaluate the true value of this solution, one must compute the lower-level reaction to x^m , i.e., the optimal solution of

$$\begin{aligned}
& \min_y f(x^m, y) \\
& \text{s.t. } g(x^m, y) \leq 0, & (3.11)
\end{aligned}$$

which is denoted by y^* . After computation of the ratio (3.9) of achieved versus predicted reduction

$$\rho_k = \frac{F(x_k, y_k) - F(x^m, y^*)}{F_m(x_k, y_k) - F_m(x^m, y^m)},$$

the algorithm updates both the current iterate and the trust-region radius, and the process is repeated until convergence occurs.

This algorithm has been tested on a set of test problems, including toll-setting problems described in Sect. 1. The good performance of the algorithm in terms of the quality of the solution (a global solution is frequently reached) is due to the accuracy of the model approximation, itself a bilevel program that can be solved for its global solution.

4 Mathematical programs with equilibrium constraints

Having reviewed the major developments in the field of bilevel programming, we would like to complete our survey by considering another important class of related problems, namely *Mathematical Programs with Equilibrium Constraints*, or MPECs. Actually, relationships between BLPPs and MPECs are so strong that some authors use the same terminology for both classes of problems, which may sometimes lead to confusion.

MPECs may be viewed as bilevel programs where the lower-level problem consists in a *variational inequality*. For a given function $\psi : \mathbb{R}^n \rightarrow \mathbb{R}^n$ and convex

set $C \subseteq \mathbb{R}^n$, the vector $x^* \in C$ is said to be a solution of the variational inequality $\text{VI}(\psi, C)$ if it satisfies

$$(x - x^*)^T \psi(x^*) \geq 0 \quad \text{for all } x \in C. \quad (4.1)$$

Note that variational inequalities are mathematical programs that allow the modelling of many *equilibrium* phenomena encountered in engineering, physics, chemistry or economics, hence the origin of the name of MPEC. The reader is referred to the monograph of t al.(1996)Luo, Pang, and Ralpho (u) for a description of fields of application.

The general formulation of an MPEC is as follows:

$$\begin{aligned} \min_{x,y} \quad & F(x, y) \\ \text{s.t.} \quad & (x, y) \in Z \text{ and } y \in \mathcal{S}(x), \end{aligned} \quad (4.2)$$

where $Z \subseteq \mathbb{R}^{n_1+n_2}$ is a nonempty closed set and $\mathcal{S}(x)$ is the solution set of the parameterized variational inequality

$$\begin{aligned} y \in \mathcal{S}(x) \Leftrightarrow y \in C(x) \\ \text{and } (v - y)^T \psi(x, y) \geq 0 \text{ for all } v \in C(x) \end{aligned} \quad (4.3)$$

defined over the closed convex set $C(x) \subset \mathbb{R}^{n_2}$. As for bilevel problems, the terms *upper-level* and *lower-level* variables are used to designate x and y respectively.

The relationship between bilevel programming problems and MPECs may be illustrated by considering two particular cases. To this end, let us first assume that the mapping $\psi(x, \cdot)$ is the partial gradient map of a real-valued continuously differentiable function $f : \mathbb{R}^{n_1+n_2} \rightarrow \mathbb{R}$, that is,

$$\psi(x, y) = \nabla_y f(x, y).$$

Then, for any fixed x , the VI (4.3) characterizes the set of stationarity conditions of the optimization problem

$$\begin{aligned} \min_y \quad & f(x, y) \\ \text{s.t.} \quad & y \in C(x). \end{aligned} \quad (4.4)$$

Moreover, if the parameterized set $C(x)$ takes the form

$$C(x) = \{y : g(x, y) \leq 0\},$$

then problem (4.4) is nothing but the lower-level problem (2.1c)–(2.1d). This shows that MPECs subsume bilevel programs provided the latter involves a *convex* and differentiable lower-level problem. Conversely, an MPEC can be formulated as a bilevel program by replacing the lower-level VI by an optimization problem.

The latter can, for instance, be constructed around the *gap function* as defined by einder(1976)s (u) ($\alpha = 0$) or shima(1992)k (u) ($\alpha > 0$), that is

$$H_\alpha(x, y) = \max_{z \in C(x)} \phi(x, y, z) \quad (4.5)$$

where

$$\phi(x, y, z) = \langle \psi(x, y), y - z \rangle - \frac{1}{2} \alpha \|y - z\|^2.$$

Alternatively, one may minimize the sum of the complementarity slackness terms associated with the KKT formulation of the variational inequality, under suitable regularity conditions on the lower-level constraints.

Many *active set* approaches have been proposed for solving MPECs, see (1998)r (a). More recent methods deal with *constraint regularization* (e.g., hinei et al.(1996)Facchinei, Jiang, and Qic hinei et al.(1996)Facchinei, Jiang, and Qia, shima and Pang(1999)k shima and Pang(1999)u, ltes(2001)h ltes(2001)c), *implicit programming techniques* (ata(1994)t ata(1994)u, ata et al.(1998)Oustrata, Kočvara, and Zowet ata et al.(1998)Oustrata, Kočvara, and Zoweu), or techniques borrowed from constrained nonlinear programming. For instance, *filter methods* have been used by cher and Leyffer(2002)e (l), while *sequential quadratic programming* approaches have been proposed by and Ralph(1997)a (i), h(1998)l (a) and cher et al.(2002)Fletcher, Leyffer, Ralph, and Scholtese (l). otte and Zhu(1996)r (a) discuss algorithms based on *penalty functions*, exact or inexact, constructed around gap functions (4.5). A *trust-region* scheme has been developed by ltes and Stöhr(1999)h (c), which is closely related to the algorithm of on et al.(2005)Colson, Marcotte, and Savardl (o) for nonlinear bilevel programming as presented in the previous section. The reader interested in a comprehensive analysis of MPECs is referred to the monographs of t al.(1996)Luo, Pang, and Ralpho (u) and ata et al.(1998)Oustrata, Kočvara, and Zowet (u).

5 Perspectives and challenges

As evidenced in this survey, bilevel programming is the subject of important research efforts from the mathematical programming and operations research communities. Many classes of bilevel programs now have dedicated solution algorithms, and researchers have started to study more complicated instances – like bilevel programs with integer variables or without derivatives – which to our view is an indication that some maturity has been reached in the field.

It is nevertheless the case that challenges remain to be tackled, in particular concerning nonlinear bilevel problems. Besides the improvement of existing methods and derivation of proper convergence results, our feeling is that a promising approach would be to develop tools similar to those by ltes(2002)h (c) allowing to take advantage of the inherent combinatorial structure of bilevel problems. These

ideas, combined with well-tried tools from nonlinear programming like sequential quadratic programming, should allow the development of a new generation of solution methods.

From a more practical point of view, we feel that the set of available bilevel programming test problems is relatively poor compared to those existing for other classes of mathematical programs. There exist some collections, like the MacMPEC collection maintained by fer(2000)y (e) or the problems presented in on(2002)l (o), but no modelling language currently allows a user-friendly embedding of the two-level structure (in MacMPEC, for instance, problems are reformulated as single-level programs using optimality conditions). This issue, together with the development of a suitable modelling language, might trigger advances in the numerical solution of BLPPs.

References

- Aiyoshi E, Shimizu K (1981) Hierarchical decentralized systems and its new solution by a barrier method. *IEEE Transactions on Systems, Man, and Cybernetics* 11: 444–449
- Aiyoshi E, Shimizu K (1984) A solution method for the static constrained Stackelberg problem via penalty method. *IEEE Transactions on Automatic Control* 29: 1111–1114
- Al-Khayal FA, Horst R, Pardalos PM (1992) Global optimization of concave functions subject to quadratic constraints : an application in nonlinear bilevel programming. *Annals of Operations Research* 34: 125–147
- Anandalingam G, Friesz T (1992) Hierarchical optimization: an introduction. *Annals of Operations Research* 34: 1–11
- Auslender A (1976) *Optimisation: Méthodes numériques*. Masson, Paris
- Bard JF (1988) Convex two-level optimization. *Mathematical Programming* 40: 15–27
- Bard JF (1988) *Practical Bilevel Optimization*, vol 30 of *Nonconvex optimization and its applications*. Kluwer Academic Publishers, Dordrecht, The Netherlands
- Bard JF, Falk J (1982) An explicit solution to the multi-level programming problem. *Computers and Operations Research* 9: 77–100
- Bard JF, Moore JT (1990) A branch and bound algorithm for the bilevel programming problem. *SIAM Journal of Scientific and Statistical Computing* 11(2): 281–292
- Ben-Ayed O (1993) Bilevel linear programming. *Computers and Operations Research* 20: 485–501
- Ben-Ayed O, Blair C (1990) Computational difficulties of bilevel linear programming. *Operations Research* 38: 556–560
- Bi Z, Calamai PH, Conn AR (1989) An exact penalty function approach for the linear bilevel programming problem. Technical Report #167-O-310789, Department of Systems Design Engineering, University of Waterloo
- Bi Z, Calamai PH, Conn AR (1991) An exact penalty function approach for the nonlinear bilevel programming problem. Technical Report #180-O-170591, Department of Systems Design Engineering, University of Waterloo
- Bialas W, Karwan M (1984) Two-level linear programming. *Management Science* 30: 1004–1020
- Bialas W, Karwan M, Shaw J (1980) A parametric complementarity pivot approach for two-level linear programming. Technical Report 80-2, State University of New York at Buffalo, Operations Research Program
- Bracken J, McGill J (1973) Mathematical programs with optimization problems in the constraints. *Operations Research* 21: 37–44
- Bracken J, McGill J (1974) Defense applications of mathematical programs with optimization problems in the constraints. *Operations Research* 22: 1086–1096
- Bracken J, McGill J (1978) Production and marketing decisions with multiple objectives in a competitive environment. *Journal of Optimization Theory and Applications* 24: 449–458
- Brotcorne L, Labbé M, Marcotte P, Savard G (2001) A bilevel model for toll optimization on a multi-commodity transportation network. *Transportation Science* 35: 1–14
- Candler W, Norton R (1977) Multilevel programming. Technical Report 20, World Bank Development Research Center, Washington DC, USA
- Candler W, Townsley R (1982) A linear two-level programming problem. *Computers and Operations Research* 9: 59–76
- Case LM (1999) *An ℓ_1 penalty function approach to the nonlinear bilevel programming problem*. PhD thesis, University of Waterloo, Ontario, Canada
- Chen Y, Florian M (1991) The nonlinear bilevel programming problem: a general formulation and optimality conditions. Technical Report CRT-794, Centre de Recherche sur les Transports, Université de Montréal, Montréal, QC, Canada
- Chen Y, Florian M (1992) On the geometric structure of linear bilevel programs: a dual approach. Technical Report CRT-867, Centre de Recherche sur les Transports, Université de Montréal, Montréal, QC, Canada
- Chen Y, Florian M, Wu S (1992) A descent dual approach for linear bilevel programs. Technical Report CRT-866, Centre de Recherche sur les Transports, Université de Montréal, Montréal, QC, Canada
- Clarke FH (1990) *Optimization and Nonsmooth Analysis*. SIAM Publications, Philadelphia, Pennsylvania, USA
- Colson B (1999) Mathematical programs with equilibrium constraints and nonlinear bilevel programming problems. Master's thesis, Department of Mathematics, FUNDP, Namur, Belgium

- Colson B (2002) BIPA (Bilevel Programming with Approximation methods): software guide and test problems. Technical Report CRT-2002-38, Centre de Recherche sur les Transports, Université de Montréal, Montréal, QC, Canada
- Colson B, Marcotte P, Savard G (2005) A trust-region method for nonlinear programming: algorithm and computational experience. *Computational Optimization and Applications* 30(3): 211–227
- Conn AR, Gould NIM, Toint PL (2000) *Trust-Region Methods*. SIAM Publications, Philadelphia, Pennsylvania, USA
- Côté JP, Marcotte P, Savard G (2003) A bilevel modeling approach to pricing and fare optimization in the airline industry. *Journal of Revenue and Pricing Management* 2: 23–36
- Dempe S (1992a) A necessary and a sufficient optimality condition for bilevel programming problems. *Optimization* 25: 341–354
- Dempe S (1992b) Optimality conditions for bilevel programming problems. In: Kall P (ed) *System modelling and optimization*. Springer-Verlag, Heidelberg, Berlin, New York, pp 17–14
- Dempe S (2002) *Foundations of Bilevel Programming*, vol 61 of *Nonconvex optimization and its applications*. Kluwer Academic Publishers, Dordrecht, The Netherlands
- Dempe S (2003) Annotated bibliography on bilevel programming and mathematical programs with equilibrium constraints. *Optimization* 52: 333–359
- Edmunds T, Bard JF (1991) Algorithms for nonlinear bilevel mathematical programs. *IEEE transactions on Systems, Man, and Cybernetics* 21(1): 83–89
- Facchinei F, Jiang H, Qi L (1996) A smoothing method for mathematical programs with equilibrium constraints. Technical Report R03-96, Università di Roma “La Sapienza”, Dipartimento di Informatica e Sistemistica
- Falk JE, Liu J (1995) On bilevel programming, Part I : general nonlinear cases. *Mathematical Programming* 70(1): 47–72
- Fletcher R, Leyffer S (2002) Numerical experience with solving MPECs by nonlinear programming methods. Numerical Analysis Report NA/210, Department of Mathematics, University of Dundee, Dundee, Scotland
- Fletcher R, Leyffer S, Ralph D, Scholtes S (2002) Local convergence of SQP methods for Mathematical Programs with Equilibrium Constraints. Numerical Analysis Report NA/209, Department of Mathematics, University of Dundee, Dundee, Scotland
- Floudas CA, Pardalos PM, eds (2001) *Encyclopedia of Optimization*. Kluwer Academic Publishers, Dordrecht, The Netherlands
- Fortuny-Amat J, McCarl B (1981) A representation and economic interpretation of a two-level programming problem. *Journal of the Operational Research Society* 32: 783–792
- Fukushima M (1992) Equivalent differentiable optimization problems and descent methods for asymmetric variational inequality problems. *Mathematical Programming* 53: 99–110
- Fukushima M, Pang JS (1999) Complementarity constraint qualifications and simplified B-stationarity conditions for mathematical programs with equilibrium constraints. *Computational Optimization and Applications* 13: 111–136
- Garey MR, Johnson DS (1979) *Computers and intractability: A guide to the theory of NP-completeness*. W.H. Freeman, New York
- Hansen P, Jaumard B, Savard G (1992) New branch-and-bound rules for linear bilevel programming. *SIAM Journal on Scientific and Statistical Computing* 13(5): 1194–1217
- Hsu S, Wen U (1989) A review of linear bilevel programming problems. In *Proceedings of the National Science Council, Republic of China, Part A: Physical Science and Engineering*, vol 13, pp 53–61
- Ishizuka Y, Aiyoshi E (1992) Double penalty method for bilevel optimization problems. *Annals of Operations Research* 34: 73–88
- Jaumard B, Savard G, Xiong J (2000) A new algorithm for the convex bilevel programming problem. Draft paper, Ecole Polytechnique de Montréal
- Jeroslow RG (1985) The polynomial hierarchy and a simple model for competitive analysis. *Mathematical Programming* 32: 146–164
- Jian H, Ralph D (1997) Smooth SQP methods for mathematical programs with nonlinear complementarity constraints. Manuscript, Department of Mathematics and Statistics, University of Melbourne
- Júdice JJ, Faustino A (1988) The solution of the linear bilevel programming problem by using the linear complementarity problem. *Investigação Operacional* 8: 77–95
- Júdice JJ, Faustino A (1992) A sequential LCP method for bilevel linear programming. *Annals of Operations Research* 34: 89–106
- Júdice JJ, Faustino A (1994) The linear-quadratic bilevel programming problem. *Information Systems and Operational Research* 32: 87–98

- Kolstad CD (1985) A review of the literature on bi-level mathematical programming. Technical Report LA-10284-MS, Los Alamos National Laboratory, Los Alamos, New Mexico, USA
- Kolstad CD, Lasdon LS (1990) Derivative estimation and computational experience with large bilevel mathematical programs. *Journal of Optimization Theory and Applications* 65: 485–499
- Labbé M, Marcotte P, Savard G (1998) A bilevel model of taxation and its applications to optimal highway pricing. *Management Science* 44: 1595–1607
- Leyffer S (2000) MacMPEC – AMPL collection of Mathematical Programs with Equilibrium Constraints. <http://www-unix.mcs.anl.gov/leyffer/MacMPEC/>
- Liu G, Han J, Wang S (1998) A trust region algorithm for bilevel programming problems. *Chinese Science Bulletin* 43(10): 820–824
- Loridan P, Morgan J (1996) Weak via strong Stackelberg problem: New results. *Journal of Global Optimization* 8: 263–287
- Luo ZQ, Pang JS, Ralph D (1996) *Mathematical programs with equilibrium constraints*. Cambridge University Press, Cambridge, England
- Marcotte P, Savard G (2005) Bilevel programming: A combinatorial perspective. In: Avis D, Hertz A, Marcotte O (eds) *Graph Theory and Combinatorial Optimization*. Kluwer Academic Publishers, Boston
- Marcotte P, Zhu DL (1996) Exact and inexact penalty methods for the generalized bilevel programming problem. *Mathematical Programming* 74: 141–157
- Migdalas A (1995) Bilevel programming in traffic planning: models, methods and challenge. *Journal of Global Optimization* 7: 381–405
- Migdalas A, Pardalos PM, Värbrand P (1997) *Multilevel Optimization: Algorithms and Applications*, vol 20 of *Nonconvex optimization and its applications*. Kluwer Academic Publishers, Dordrecht, The Netherlands
- Nash J (1951) Non-cooperative games. *Annals of Mathematics* 54: 286–295
- Outrata J (1993) Necessary optimality conditions for Stackelberg problems. *Journal of Optimization Theory and Applications* 76: 305–320
- Outrata J (1994) On optimization problems with variational inequality constraints. *SIAM Journal on Optimization* 4(2): 340–357
- Outrata J, Kočvara M, Zowe J (1998) *Nonsmooth Approach to Optimization Problems with Equilibrium Constraints: Theory, Applications and Numerical Results*, vol 28 of *Nonconvex optimization and its applications*. Kluwer Academic Publishers, Dordrecht, The Netherlands
- Papavassilopoulos G (1982) Algorithms for static Stackelberg games with linear costs and polyhedral constraints. In *Proceedings of the 21st IEEE Conference on Decisions and Control*, pp 647–652
- Ralph D (1998) Optimization with equilibrium constraints: A piecewise SQP approach. Manuscript, Department of Mathematics and Statistics, University of Melbourne
- Savard G (1989) *Contribution à la programmation mathématique à deux niveaux*. PhD thesis, Ecole Polytechnique de Montréal, Université de Montréal, Montréal, QC, Canada
- Savard G, Gauvin J (1994) The steepest descent direction for the nonlinear bilevel programming problem. *Operations Research Letters* 15: 265–272
- Scholtes S (2001) Convergence properties of a regularisation scheme for mathematical programs with complementarity constraints. *SIAM Journal on Optimization* 11: 918–936
- Scholtes S (2002) Combinatorial structures in nonlinear programming. *Optimization Online* http://www.optimization-online.org/DB_HTML/2002/05/477.html
- Scholtes S, Stöhr M (1999) Exact penalization of mathematical programs with equilibrium constraints. *SIAM Journal on Control and Optimization* 37(2): 617–652
- Sheffi Y (1985) *Urban transportation networks: equilibrium analysis with mathematical programming methods*. Prentice-Hall, Englewood Cliffs, New Jersey, USA
- Sherali HD, Soyster AL, Murphy FH (1983) Stackelberg-Nash-Cournot equilibria: Characterizations and computations. *Operations Research* 31: 253–276
- Shimizu K, Aiyoshi E (1981) A new computational method for Stackelberg and min-max problems by use of a penalty method. *IEEE Transactions on Automatic Control* 26: 460–466
- Shimizu K, Ishizuka Y, Bard JF (1997) *Nondifferentiable and two-level mathematical programming*. Kluwer Academic Publishers, Dordrecht, The Netherlands
- Stackelberg H (1952) *The Theory of Market Economy*. Oxford University Press, Oxford, UK
- Thoai NV, Yamamoto Y, Yoshise A (2002) Global optimization method for solving mathematical programs with linear complementarity constraints. Discussion Paper No. 987, Institute of Policy and Planning Sciences, University of Tsukuba, Japan

- Tuy H, Migdalas A, Värbrand P (1993) A global optimization approach for the linear two-level program. *Journal of Global Optimization* 3: 1–23
- Vicente LN, Calamai PH (1994) Bilevel and multilevel programming: a bibliography review. *Journal of Global Optimization* 5(3): 291–306
- Vicente LN, Calamai PH (1995) Geometry and local optimality conditions for bilevel programs with quadratic strictly convex lower levels. In: Du D, Pardalos PM (eds) *Minimax and Applications*, vol 4 of *Nonconvex optimization and its applications*. Kluwer Academic Publishers, Dordrecht, The Netherlands, pp 141–151
- Vicente LN, Savard G, Júdice JJ (1994) Descent approaches for quadratic bilevel programming. *Journal of Optimization Theory and Applications* 81: 379–399
- Vicente LN, Savard G, Júdice JJ (1996) The discrete linear bilevel programming problem. *Journal of Optimization Theory and Applications* 89: 597–614
- Wen U, Hsu S (1991) Linear bi-level programming problems – a review. *Journal of the Operational Research Society* 42: 125–133