Energy Systems

Stephan Dempe
Vyacheslav Kalashnikov
Gerardo A. Pérez-Valdés
Nataliya Kalashnykova

Bilevel Programming Problems

Theory, Algorithms and Applications to Energy Networks

Springer

Energy Systems

Series editor

Panos M. Pardalos, Gainesville, USA

More information about this series at http://www.springer.com/series/8368

Stephan Dempe · Vyacheslav Kalashnikov
Gerardo A. Pérez-Valdés
Nataliya Kalashnykova

Bilevel Programming Problems

Theory, Algorithms and Applications to Energy Networks

Springer

Stephan Dempe
Fak. für Mathematik und Informatik
TU Bergakademie Freiberg
Freiberg, Sachsen
Germany

Vyacheslav Kalashnikov
Department of Systems and Industrial Engineering
Tecnologico de Monterrey (ITESM)
Monterrey, NL
Mexico

and

Department of Social Modeling
Central Economics and Mathematics Institute (CEMI)
Russian Academy of Sciences
Moscow
Russian Federation

and

Deparment of Computing
Sumy State University
Sumy
Ukraine

Gerardo A. Pérez-Valdés
Department of Industrial Economics
Norwegian University of Science and Technology (NTNU)
Trondheim
Norway

Nataliya Kalashnykova
Facultad Ciencias Físico-Mathemáticas (FCFM)
Universidad Autónoma de Nuevo León (UANL)
San Nicolás de los Garza
Mexico

ISSN 1867-8998 ISSN 1867-9005 (electronic)
Energy Systems
ISBN 978-3-662-45826-6 ISBN 978-3-662-45827-3 (eBook)
DOI 10.1007/978-3-662-45827-3

Library of Congress Control Number: 2014958578

Springer Heidelberg New York Dordrecht London

Printed on acid-free paper

Springer-Verlag GmbH Berlin Heidelberg is part of Springer Science+Business Media
(www.springer.com)

Preface

Bilevel optimization is a vital field of active research. Depending on its formulation it is part of nonsmooth or nondifferentiable optimization, conic programming, optimization with constraints formulated as generalized equations, or set-valued optimization. The investigation of many practical problems as decision making in hierarchical structures, or situations where the reaction of nature on selected actions needs to be respected, initiated modeling them as bilevel optimization problems. In this way, new theories have been developed with new results obtained.

A first attempt was the use of the Karush-Kuhn-Tucker conditions in situations when they are necessary and sufficient optimality conditions for the lower level problem, or dual problems in case strong duality holds to model the bilevel optimization problem. The result is a special case of the mathematical program with equilibrium constraints (MPEC), or complementarity constraints (MPCC). The latter has motivated the investigation of optimality conditions and the development of algorithms solving such problems. Unfortunately, it has been shown very recently that stationary points of an MPEC need not be related to stationary solutions of the bilevel optimization problem. Because of that, the solution algorithms must select the Lagrange multipliers associated with the lower level problem very carefully. Another option is to avoid the explicit use of Lagrange multipliers resulting in the so-called primal KKT transformation, which is an optimization problem with a generalized equation as the constraint. Violation of the constraint qualifications, often used to verify the optimality conditions and convergence of the solution algorithms, at every feasible point are other challenges for research.

The idea of using the optimal value function of the lower level problem to model the bilevel optimization problem is perhaps self-explanatory. The result yet is a nondifferentiable equality constraint. One promising approach here is based on variational analysis, which is also exploited to verify the optimality conditions for the MPCC. So, bilevel optimization initiated some advances in variational analysis, too.

Applications often force the use of integer variables in the respective models. Besides suitable formulations, mixed-integer bilevel optimization problems renew the question of existence of an optimal solution, leading to the notion of a weak solution. Surprisingly, adding some constraints that are inactive at a global optimum of the continuous bilevel problem, as well as replacing a discrete bilevel problem with its continuous relaxation can destroy the global optimality of a feasible point.

These and other questions are the topic of the first part of the monograph. In the second part, certain applications are carefully investigated, especially a natural gas cash-out problem, an equilibrium problem in a mixed oligopoly, and a toll assignment problem. For these problems, besides the formulation of solution algorithms, results of the first numerical experiments with them are also reported.

Bilevel optimization is a quickly developing field of research with challenging and promising contributions from different topics of mathematics like optimization, as well as from other sciences like economics, engineering, or chemistry. It was not a possible aim of the authors to provide an overview of all the results available in this area. Rather than that, we intended to show some interactions with other topics of research, and to formulate our opinion about some directions for explorations in the future.

Stephan Dempe
Vyacheslav Kalashnikov
Gerardo A. Pérez-Valdés
Nataliya Kalashnykova

Acknowledgment

Kalashnikov V. and Kalashnykova N., was supported by the SEP-CONACYT grant CB-2013-01-221676 (Mexico).

Contents

Chapter 1
Introduction

1.1 The Bilevel Optimization Problem

Since its first formulation by Stackelberg in his monograph on market economy in 1934 [294] and the first mathematical model by Bracken and McGill in 1972 [27] there has been a steady growth in investigations and applications of bilevel optimization. Formulated as a hierarchical game (the Stackelberg game), two decision makers act in this problem. The so-called leader minimizes his objective function subject to conditions composed (in part) by optimal decisions of the so-called follower. The selection of the leader influences the feasible set and the objective function of the follower's problem, who's reaction has strong impact on the leader's payoff (and feasibility of the leader's initial selection). Neither player can dominate the other one completely. The bilevel optimization problem is the leader's problem, formulated mathematically using the graph of the solution set of the follower's problem.

The bilevel optimization has been shown to be $\mathcal{NP}$-hard, even verification of local optimality for a feasible solution is in general $\mathcal{NP}$-hard, often used constraint qualifications are violated in every feasible point. This makes the computation of an optimal solution a challenging task.

Bilevel optimization problems are nonconvex optimization problems, tools of variational analysis have successfully been used to investigate them. The results are a larger number of necessary optimality conditions, some of them are presented in Chap. 3 of this monograph.

A first approach to investigate bilevel optimization problems is to replace the lower level problem by its (under certain assumptions necessary and sufficient) optimality conditions, the Karush-Kuhn-Tucker conditions. This replaces the bilevel optimization problem by a so-called mathematical program with complementarity conditions (MPCC). MPCCs are nonconvex optimization problems, too. Algorithms solving them compute local optimal solutions or stationary points. Recently it has been shown that local optimal solutions of an MPCC need not to be related to local optimal solutions of the corresponding bilevel optimization problem, new attempts

S. Dempe et al., *Bilevel Programming Problems*, Energy Systems,
DOI 10.1007/978-3-662-45827-3_1

for the development of solution approaches for the bilevel problem are necessary. Some results can be found in different chapters of this monograph.

The existence of an optimal solution, verification of necessary optimality conditions, and convergence of solution algorithms are strongly related to continuity of certain set-valued mappings. These properties can often not be guaranteed for mixed-discrete bilevel optimization problems. This is perhaps one reason for the small number of references on those class of problems. But, applied problems do often lead to mixed-integer bilevel problems. One such problem is investigated in Chap. 6. Focus on mixed-discrete bilevel optimization including the notion of a weak optimal solution and some ideas for solving these problems is in Chap. 5.

The solution set of an optimization problem does in general not reduce to a singleton, leading to the task of selecting a "good" optimal solution. If the quality of an optimal solution is measured by a certain function, this function needs to be minimized on the solution set of a second optimization problem. This is the so-called simple bilevel optimization problem, investigated in Chap. 4.

Interest in bilevel optimization is largely driven by applications. Two of them are investigated in details in Chaps. 6 and 7. The gas cash-out problem is a bilevel optimization problem with one Boolean variable, formulated using nondifferentiable functions. Applying results of the previous chapters, after some transformations and the formulation of an approximate problem, a model is obtained which can efficiently be solved. The obtained solutions have successfully be used in practice.

Due to its complexity, the dimension of bilevel optimization models is of primarily importance for solving them. Large-scale problems can perhaps not be solved in reasonable time. But, e.g. the investigation of stochastic bilevel optimization problems using methods to approximate the probability distributions leads to large-scale problems and not all data are deterministic ones on many applications. This makes ideas to reduce the number of variables important. Such ideas are the topic of Chap. 8.

1.2 Possible Transformations into a One-Level Problem

Bilevel optimization problems are optimization problems where the feasible set is determined (in part) using the graph of the solution set mapping of a second parametric optimization problem. This problem is given as

$$\min_y \{f(x, y) : g(x, y) \le 0,\ y \in T\}, \tag{1.1}$$

where $f : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R},\ g : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^p,\ T \subseteq \mathbb{R}^m$ is a (closed) set.

Let $Y : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ denote the *feasible set mapping*:

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

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

the *optimal value function*, and $\Psi : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ the *solution set mapping* of the problem (1.1) for a fixed value of x:

$$\Psi(x) := \{y \in Y(x) \cap T : f(x, y) \le \varphi(x)\}.$$

Let

$$\mathbf{gph}\,\Psi := \left\{(x, y) \in \mathbb{R}^n \times \mathbb{R}^m : y \in \Psi(x)\right\}$$

be the graph of the mapping Ψ. Then, the bilevel optimization problem is given as

$$\text{“}\min_x\text{”}\{F(x, y) : G(x) \le 0,\ (x, y) \in \mathbf{gph}\,\Psi,\ x \in X\}, \tag{1.2}$$

where $F : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R},\ G : \mathbb{R}^n \to \mathbb{R}^q,\ X \subseteq \mathbb{R}^n$ is a closed set.

Problem (1.1), (1.2) can be interpreted as an hierarchical game of two decision makers (or players) which make their decisions according to an hierarchical order. The first player (which is called the *leader*) makes his selection first and communicates it to the second player (the so-called *follower*). Then, knowing the choice of the leader, the follower selects his *response* as an optimal solution of problem (1.1) and gives this back to the leader. Thus, the leader's task is to determine a best decision, i.e. a point $\widehat{x}$ which is feasible for the problem (1.2): $G(\widehat{x}) \le 0,\ \widehat{x} \in X$, minimizing together with the response $\widehat{y} \in \Psi(\widehat{x})$ the function $F(x, y)$. Therefore, problem (1.1) is called the follower's problem and (1.2) the leader's problem. Problem (1.2) is the *bilevel optimization problem*.

Example 1.1 In case of a linear bilevel optimization problem with only one upper and one lower level variables, where all functions F, f, g_i are (affine) linear functions, the bilevel optimization problem is illustrated in Fig. 1.1. Here, $G(x) \equiv 0$ and the set $\{(x, y) : g(x, y) \le 0\}$ of feasible points for all values of x is the hatched area. If x is fixed to $x0$ the feasible set of the lower level problem (1.1) is the set of points $(x0, y)$ above $x0$. Now, if the lower level objective function $f(x, y) = -y$ is minimized on this set, the optimal solution of the lower level problem on the thick line is obtained. Then, if x is moved along the x-axis, the thick line as the set of feasible solutions of the upper level problem arises. In other words, the thick line equals the **gph** Ψ of the solution set mapping of the lower level problem. This is the feasible set of the upper level (or bilevel) optimization problem. Then, minimizing the upper level objective function on this set, the (in this case unique) optimal solution of the bilevel optimization problem is obtained as indicated in Fig. 1.1. □

It can be seen in Fig. 1.1 that the bilevel optimization problem is a nonconvex (since **gph** Ψ is nonconvex) optimization problem. Hence, local optimal solutions and also stationary points can appear.

Example 1.2 Consider the problem

$$\text{“}\min_x\text{”}\{x^2 + y : y \in \Psi(x)\},$$

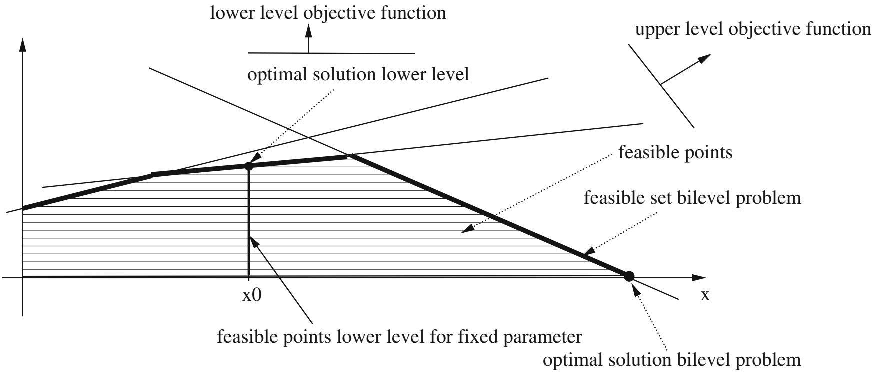

Fig. 1.1 Illustration of the linear bilevel optimization problem

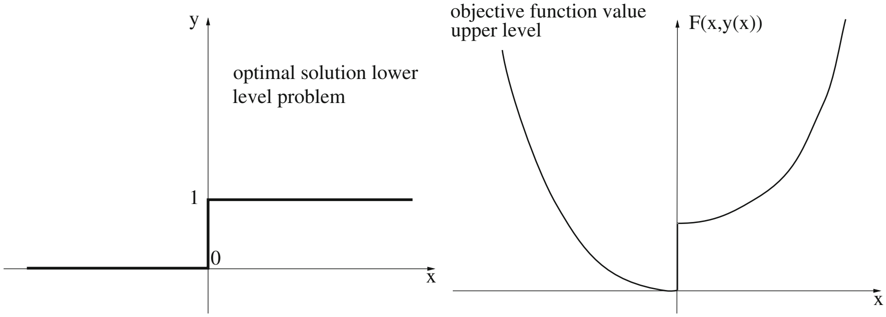

Fig. 1.2 Mapping to be "minimized" in Example 1.2

where

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

Then, the graph of the mapping Ψ is given in the figure on the left hand side of Fig. 1.2 and the graph of the mapping $x \mapsto F(x, \Psi(x))$ of the upper level objective function is plotted in the figure on the right-hand side. Note, that this is not a function and that its minimum is unclear since its existence depends on the response $y \in \Psi(x)$ of the follower on the leader's selection at $x = 0$. If the solution $y = 0$ is taken for $x = 0$, an optimal solution of the bilevel optimization problem exists. This is the *optimistic bilevel optimization problem* introduced below. In all other cases, the minimum does not exist, the infimum function value of the upper level objective function is again zero but it is not attained. If $y = 1$ is taken, the so-called *pessimistic bilevel optimization problem* arises. $\square$

Hence, strictly speaking, the problem (1.2) is not well-posed in the case that the set $\Psi(x)$ is not a a singleton for some x, the mapping $x \mapsto F(x, y(x))$ is not a function. This is implied by an ambiguity in the computation of the upper level objective function value, which is rather an element in the set $\{F(x, y) : y \in \Psi(x)\}$. We have used quotation marks in (1.2) to indicate this ambiguity. To overcome such an unpleasant situation, the leader has a number of possibilities:

1. The leader can assume that the follower is willing (and able) to cooperate. In this case, the leader can take that solution within the set $\Psi(x)$ which is a best one with respect to the upper level objective function. This leads then to the function

$$\varphi_o(x) := \min\{F(x, y) : y \in \Psi(x)\} \tag{1.3}$$

to be minimized on the set $\{x : G(x) \leq 0,\ x \in X\}$. This is the optimistic approach leading to the *optimistic bilevel optimization problem*. The function $\varphi_o(x)$ is called *optimistic solution function*. Roughly speaking, this problem is closely related to the problem

$$\min_{x,y}\{F(x, y) : G(x) \leq 0,\ (x, y) \in \mathbf{gph}\,\Psi,\ x \in X\}. \tag{1.4}$$

If the point $\overline{x}$ is a local minimum of the function $\varphi_o(\cdot)$ on the set

$$\{x : G(x) \leq 0,\ x \in X\}$$

and $\overline{y} \in \Psi(\overline{x})$, then the point $(\overline{x}, \overline{y})$ is also a local minimum of problem (1.4). The opposite implication is in general not correct, as the following example shows:

Example 1.3 Consider the problem of minimizing the function $F(x, y) = x$ subject to $x \in [-1, 1]$ and $y \in \Psi(x) := \underset{y}{\mathrm{Argmin}}\{xy : y \in [0, 1]\}$. Then,

$$y(x) \in \begin{cases} [0, 1] & \text{for } x = 0, \\ \{1\} & \text{for } x < 0, \\ \{0\} & \text{for } x > 0. \end{cases}$$

Hence, the point $(\overline{x}, \overline{y}) = (0, 0)$ is a local minimum of the problem

$$\min_{x,y}\{x : x \in [-1, 1], y \in \Psi(x)\}$$

since, for each feasible point (x, y) with $\|(x, y) - (\overline{x}, \overline{y})\| \leq 0.5$ we have $x \geq 0$. But, the point $\overline{x}$ does not minimize the function $\varphi_o(x) = x$ on $[-1, 1]$ locally. □

For more information about the relation between both problems, the interested reader is referred to Dempe [52].

2. The leader has no possibility to influence the follower's selection neither he/she has an intuition about the follower's choice. In this case, the leader has to accept

the follower's opportunity to take a worst solution with respect to the leader's objective function and he/she has to bound the damage resulting from such an unpleasant selection. This leads to the function

$$\varphi_p(x) := \max\{F(x, y) : y \in \Psi(x)\} \tag{1.5}$$

to be minimized on the set $\{x : G(x) \leq 0,\ x \in X\}$:

$$\min\{\varphi_p(x) : G(x) \leq 0,\ x \in X\} \tag{1.6}$$

This is the pessimistic approach resulting in the *pessimistic bilevel optimization problem*. The function $\varphi_p(x)$ is the *pessimistic solution function*. This problem is often much more complicated than the optimistic bilevel optimization problem, see Dempe [52].

In the literature there is also another pessimistic bilevel optimization problem. To describe this problem consider the bilevel optimization problem with connecting upper level constraints and an upper level objective function depending only on the upper level variable x:

$$\text{“}\min_x\text{”}\{F(x) : G(x, y) \leq 0,\ y \in \Psi(x)\}. \tag{1.7}$$

In this case, a point x is feasible if there exists $y \in \Psi(x)$ such that $G(x, y) \leq 0$, which can be written as

$$\min_x\{F(x) : G(x, y) \leq 0 \text{ for some } y \in \Psi(x)\}.$$

Now, if the quantifier $\exists$ is replaced by $\forall$ we derive a second pessimistic bilevel optimization problem

$$\min_x\{F(x) : G(x, y) \leq 0 \text{ for all } y \in \Psi(x)\}. \tag{1.8}$$

This problem has been investigated in Wiesemann et al. [316]. The relations between (1.8) and (1.6) should to be investigated in future.

3. The leader is able to predict a selection of the follower: $y(x) \in \Psi(x)$ for all x. If this function is inserted into the upper level objective function, this leads to the problem

$$\min_x\{F(x, y(x)) : G(x) \leq 0,\ x \in X\}.$$

Such a function $y(\cdot)$ is called a selection function of the point-to-set mapping $\Psi(\cdot)$. Hence, we call this approach the *selection function approach*. One special case of this approach arises if the optimal solution of the lower level problem is unique for all values of x. It is obvious that the optimistic and the pessimistic problems are special cases of the selection function approach.

Even under restrictive assumptions (as in the case of linear bilevel optimization or if the follower's problem has a unique optimal solution for all x), the function $y(\cdot)$ is in general not differentiable. Hence, the bilevel optimization problem is a nonsmooth optimization problem.

Definition 1.1 A point $\overline{z} \in Z$ is a local optimal solution of the optimization problem

$$\min\{w(z) : z \in Z\}$$

provided that there is a positive number $\varepsilon > 0$ such that

$$w(z) \geq w(\overline{z}) \ \forall\, z \in Z \text{ satisfying } \|z - \overline{z}\| \leq \varepsilon.$$

$\overline{z}$ is a global optimal solution of this problem if ε can be taken arbitrarily large.

This well-known notion of a (local) optimal solution can be applied to the bilevel optimization problems and, using e.g. Weierstraß Theorem we obtain that problem (1.4) has a global optimal solution if the function F is continuous and the set $Z := \{(x, y) : G(x) \leq 0,\ (x, y) \in \textbf{gph}\,\Psi,\ x \in X\}$ is not empty and compact. If this set is not bounded but only a nonempty and closed set and the function F is continuous and coercive (i.e. $F(x, y)$ tends to infinity for $\|(x, y)\|$ tending to infinity) problem (1.4) has a global optimal solution, too. For closedness of the set Z we need closedness of the graph of the solution set mapping of the lower level problem. We will come back to this issue in Chap. 3, Theorem 3.3.

With respect to problem

$$\min\{\varphi_0(x) : G(x) \leq 0,\ x \in X\} \tag{1.9}$$

existence of an optimal solution is guaranteed if the function $\varphi_0(\cdot)$ is lower semicontinuous (which means that $\liminf_{x\to x^0} \varphi_0(x) \geq \varphi_0(x^0)$ for all x^0) and the set Z is not empty and compact by an obvious generalization of the Weierstraß Theorem. Again boundedness of this set can be replaced by coercivity. Lower semicontinuity of the function is again an implication of upper semicontinuity of the mapping $x \mapsto \Psi(x)$, see for instance Bank et al. [8] in combination with Theorem 3.3. It is easy to see that a function $w(\cdot)$ is lower semicontinuous if and only if its epigraph $\operatorname{epi} w := \{(z, \alpha) : w(z) \leq \alpha\}$ is a closed set.

Example 1.2 showed already that an optimal solution of the problem (1.6) does often not exist. Its existence is guaranteed e.g. if the function $\varphi_p(\cdot)$ is lower semicontinuous and the set Z is not empty and compact (Lucchetti et al. [207]). But, for lower semicontinuity of the function $\varphi_p(\cdot)$ lower semicontinuity of the solution set mapping $x \mapsto \Psi(x)$ is needed which can often only be shown if the optimal solution of the lower level problem is unique (see Bank et al. [8]).

If an optimal solution of problem (1.6) does not exist we can aim to find a weak (global) optimum by replacing the epigraph of the objective function by its closure: Let $\overline{\varphi}_p$ be defined such that

$$\text{epi } \overline{\varphi}_p = \text{cl epi } \varphi_p.$$

Then, a local or global optimal solution of the problem

$$\min\{\overline{\varphi}_p(x) : G(x) \le 0, \; x \in X\} \tag{1.10}$$

is called a (local or global) weak solution of the pessimistic bilevel optimization problem (1.6). Note that $\overline{\varphi}_p(x^0) = \liminf_{x \to x^0} \varphi_p(x)$. The function $\overline{\varphi}_p(\cdot)$ is the largest lower semicontinuous function which is not larger than $\varphi_p(\cdot)$, see Fanghänel [105]. Hence, a weak global solution of problem (1.6) exists provided that $Z \neq \emptyset$ is compact.

1.3 An Easy Bilevel Optimization Problem: Continuous Knapsack Problem in the Lower Level

To illustrate the optimistic/pessimistic approaches to the bilevel optimization problem consider

$$\text{“}\min_b\text{”}\{d^\top y + fb : b_u \le b \le b_o, \ldots, y \in \Psi(b)\}, \tag{1.11}$$

where

$$\Psi(b) := \operatorname*{Argmin}_y\{c^\top y : a^\top y \ge b, \; 0 \le y_i \le 1 \; \forall \, i = 1, \ldots, n\}$$

and $a, \; c, \; d \in \mathbb{R}^n_+$. Note that the upper level variable is called b in this problem. Assume that the indices are ordered such that

$$\frac{c_i}{a_i} \le \frac{c_{i+1}}{a_{i+1}}, \; i = 1, 2, \ldots, n-1.$$

Then, for fixed $b \in \left\{b : \sum_{i=1}^{k-1} a_i \le b \le \sum_{i=1}^{k} a_i\right\}$, the point

$$y_i = \begin{cases} 1, & i = 1, \ldots, k-1 \\ \frac{b - \sum_{j=1}^{k-1} a_j}{a_k}, & i = k \\ 0, & i = k+1, \ldots, n \end{cases} \tag{1.12}$$

is an optimal solution of the lower level problem. Its optimal function value in the lower level is

$$\varphi(b) = \sum_{i=1}^{k-1} c_i + \frac{c_k}{a_k}(b - \sum_{j=1}^{k-1} a_j),$$

which is an affine linear function of b. The function $b \mapsto \varphi(b)$ is convex. The optimal solution of the lower level problem is unique provided that $\frac{c_k}{a_k}$ is unique in the set $\left\{\frac{c_i}{a_i} : i = 1, \ldots, n\right\}$. Otherwise, the indices $i \in \left\{j : \frac{c_i}{a_i} = \frac{c_k}{a_k}\right\}$ need to be ordered such that

$$d_t \leq d_{t+1} : t, t+1 \in \left\{j : \frac{c_i}{a_i} = \frac{c_k}{a_k}\right\}$$

for the optimistic and

$$d_t \geq d_{t+1} : t, t+1 \in \left\{j : \frac{c_i}{a_i} = \frac{c_k}{a_k}\right\}$$

for the pessimistic approaches. As illustration consider the following example:

Example 1.4 The lower level problem is

$$\begin{array}{l} 10x_1 + 30x_2 + 8x_3 + 60x_4 + 4x_5 + 16x_6 + 32x_7 + 30x_8 + 120x_9 + 6x_{10} \to \min \\ 5x_1 + 3x_2 + 2x_3 + 5x_4 + x_5 + 8x_6 + 4x_7 + 3x_8 + 6x_9 + 3x_{10} \geq b \\ \forall\, i : \quad 0 \leq y_i \leq 1, \end{array}$$

and the upper level objective function is

$$\begin{aligned} F(x, f) = {} & 20x_1 + 15x_2 - 24x_3 + 20x_4 - 40x_5 \\ & + 80x_6 - 32x_7 - 60x_8 - 12x_9 - 60x_{10} + fb. \end{aligned}$$

This function is to be minimized subject to $y \in \Psi(b)$ and b is in some closed interval $[b_u, b_o]$. Note that the upper level variable is b and the lower level one is x in this example.

Using the above rules we obtain the the following sequence of the indices in the optimistic approach:

$i =$	10	1	6	5	3	7	8	2	4	9
$\frac{c_i}{a_i} =$	2	2	2	4	4	8	10	10	12	20
$\frac{d_i}{a_i} =$	−20	4	10	−40	−12	−8	−20	5	4	−2

Using the pessimistic approach we get

$i =$	6	1	10	3	5	7	2	8	4	9
$\frac{c_i}{a_i} =$	2	2	2	4	4	8	10	10	12	20
$\frac{d_i}{a_i} =$	10	4	−20	−40	−12	−8	−20	5	4	−2

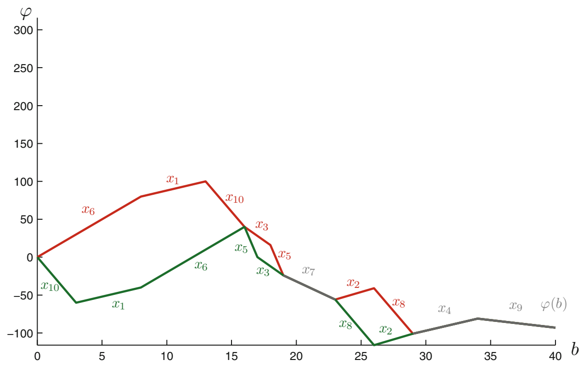

Fig. 1.3 The optimistic and pessimistic objective value functions in Example 1.4, see Winter [317]

Now, the functions $\varphi_o(b)$ and $\varphi_p(b)$ are plotted in Fig. 1.3. The upper function is the pessimistic value function, computed according to (1.5), the lower one the optimistic value function, cf. (1.3).

Both functions are continuous, but not convex. Local optima of both functions can either be found at $b \in (b_u, b_o)$ or at points

$$b \in \left\{ \sum_{i \in I} a_i : I \subseteq \{1, 2, \ldots, n\} \right\}.$$

Note that local optima can be found at vertices of the set

$$\left\{ (x, b) : b_u \le b \le b_0,\ 0 \le x_i \le 1,\ i = 1, \ldots, n,\ \sum_{i=1}^{n} a_i x_i = b \right\}. \qquad \square$$

1.4 Short History of Bilevel Optimization

The history of bilevel optimization dates back to H.v. Stackelberg who in 1934 formulated in the monograph [294] an hierarchical game of two players now called Stackelberg game. The formulation of the bilevel optimization problem goes back to Bracken and McGill [27], the notion "Bilevel Programming" has been coined probably by Candler and Norton [39], see also Vicente [305]. With the beginning of the 80s of the last century a very intensive investigation of bilevel optimization started. A number of monographs, see e.g. Bard [10], Shimizu et al. [288] and Dempe [52], edited volumes, see Dempe and Kalashnikov [57], Talbi [297] and Migdalas et al. [231] and (annotated) bibliographies, see e.g. Vicente and Calamai [306], Dempe [53] have been published in that field.

One possibility to investigate bilevel optimization problems is to transform them into one-level (or ordinary) optimization problems. This will be the topic of Chap. 3.

In the first years linear bilevel optimization problems (where all the problem functions are affine linear and the sets X and T equals the whole spaces) have been transformed using linear optimization duality or, equivalently, the Karush-Kuhn-Tucker conditions for linear optimization. Applying this approach, solution algorithms have been suggested, see e.g. Candler and Townsley [40]. The transformed problem is a special case of a mathematical program with equilibrium constraints MPEC (now sometimes called mathematical program with complementarity constraints MPCC). We can call this the KKT transformation of the bilevel optimization problem. This approach is also possible for convex parametric lower level problems satisfying some regularity assumption.

General MPCC's have been the topic of some monographs, see e.g. Luo et al. [208] and Outrata et al. [259]. Solution algorithms for MPCC's (see for instance Outrata et al. [259], Demiguel et al. [48], Leyffer et al. [201], and many others) have been suggested also for solving bilevel optimization problems.

Since MPCC's are nonconvex optimization problems, solution algorithms will hopefully compute local optimal solutions of the MPCCs. Thus, it is interesting if a local optimal solution of an the KKT transformation of a bilevel optimization problem is related to a local optimal solution of the latter problem. This has been the topic of the paper [55] by Dempe and Dutta. We will come back to this in Chap. 3.

Later on, the selection function approach to bilevel optimization has been investigated in the case when the optimal solution of the lower level problem is uniquely determined and strongly stable in the sense of Kojima [191]. Then, under some assumptions, the optimal solution of the lower level problem is a PC^1-function, see Ralph and Dempe [265] and Scholtes [283] for the definition and properties of PC^1-functions. This can then be used to determine necessary and sufficient optimality conditions for bilevel optimization (see Dempe [50]).

Using the optimal value function $\varphi(x)$ of the lower level problem (1.1), the bilevel optimization problem (1.4) can be replaced with

$$\min_{x,y}\{F(x, y) : G(x) \leq 0,\ g(x, y) \leq 0,\ f(x, y) \leq \varphi(x),\ x \in X\}.$$

This is the *optimal value transformation*. Since the optimal value function is nonsmooth even under restrictive assumptions, this is a nonsmooth, nonconvex optimization problem. Using nonsmooth analysis (see e.g. Mordukhovich [241, 242], Rockafellar and Wets [274]), optimality conditions for the optimal value transformation can be obtained (see e.g. Outrata [260], Ye and Zhu [324], Dempe et al. [56]).

Nowadays, a large number of PhD thesis have been written on bilevel optimization problems, very different types of (necessary and sufficient) optimality conditions can be found in the literature, the number of applications is huge and both exact and heuristic solution algorithms have been suggested.

1.5 Applications of Bilevel Optimization

1.5.1 Optimal Chemical Equilibria

In the monograph Dempe [52] the following application of bilevel optimization in the chemical industry is formulated:

In producing substances by chemical reactions we have often to answer the question of how to compose a mixture of chemical substances such that

- the substance we like to produce really arises as a result of the chemical reactions in the reactor and
- the amount of this substance should clearly be as large as possible or some other (poisonous or etching) substance is desired to be vacuous or at least of a small amount.

It is possible to model this problem as a bilevel optimization problem where the first aim describes the lower level problem and the second one is used to motivate the upper level objective function.

Let us start with the lower level problem. Although the chemists are technically not able to observe in situ the single chemical reactions at higher temperatures, they described the final point of the system by a convex optimization problem. In this problem, the entropy functional $f(y, p, T)$ is minimized subject to the conditions that the mass conservation principle is satisfied and masses are not negative. Thus, the obtained equilibrium state depends on the pressure p and the temperature T in the reactor as well as on the masses x of the substances which have been put into the reactor:

$$\begin{aligned} &\sum_{i=1}^{N} c_i(p, T) y_i + \sum_{i=1}^{G} y_i \ln \frac{y_i}{z} \to \min_y \\ &z = \sum_{j=1}^{G} y_j, \ Ay = \overline{A}x, \ y \geq 0, \end{aligned}$$

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; y is the vector of the masses of the substances in the resulting chemical equilibrium whereas x denotes the initial masses of substances put into the reactor; $\overline{A}$ is a submatrix of A consisting of the columns corresponding to the initial substances. The value of $c_i(p, T)$ gives the chemical potential of a substance which depends on the pressure p and the temperature T (Smith and Missen [291]). Let $y(p, T, x)$ denote the unique optimal solution of this problem. The variables p, T, x can thus be considered as parameters for the chemical reaction. The problem is now that there exists some desire about the result of the chemical reactions which should be reached as best as possible, as e.g. the goal that the mass of one substance should be as large or as small as possible in the resulting equilibrium. To reach this goal the parameters

p, T, x are to be selected such that the resulting chemical equilibrium satisfies the overall goal as best as possible (Oeder [258]):

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

1.5.2 Optimal Traffic Tolls

In more and more regions of the world, traffic on the streets is due to tolls. To model such a problem, use a directed graph $G = (V, E)$ where the nodes $v \in V := \{1, 2, \ldots, n\}$ stand for the junctions in some region and the directed edges (or arcs) $(i, j) \in E \subset V \times V$ are used to implement the streets leading from junction i to junction j. Then, the graph is used to model the map of the streets in a certain region. The streets are modeled as one-way roads here. If a street can be passed in both directions, there are opposite directed edges in the graph. The streets are assumed to have certain capacities which are modeled as a function $u : E \to \mathbb{R}$ and the cost (or time) to pass one street by a driver is given by a second function $c : E \to \mathbb{R}$. We assume here for simplicity that the costs are independent of the flow on the street. Assume further that there is a set T of pairs of nodes $(q, s) \in V \times V$ for which there is a certain demand d_{qs} of traffic running from the origin q to the destination nodes s, $(q, s) \in T$. Then, if x_e^{qs} is used to denote the part of the traffic with respect to the origin-destination pair (O-D pair in short) $(q, s) \in T$ using the street $e = (i, j) \in E$, the problem of computing the system optimum for the traffic can be modeled as a multicommodity flow problem (Ahuja et al. [1]):

$$\sum_{(q,s)\in T}\sum_{e\in E} c_e x_e^{qs} \longrightarrow \min \tag{1.13}$$

$$x_e + \sum_{(q,s)\in T} x_e^{qs} = u_e \quad \forall e \in E \tag{1.14}$$

$$\sum_{e\in O(j)} x_e^{qs} - \sum_{e\in I(j)} x_e^{qs} = \begin{cases} d_{qs}, & j = q \\ 0, & j \in V \setminus \{q, s\} \\ -d_{qs}, & j = s \end{cases} \quad \forall (q, s) \in T \tag{1.15}$$

$$x_e, x_e^{qs} \geq 0 \quad \forall (q, s) \in T, \forall e \in E. \tag{1.16}$$

Here $O(j)$ and $I(j)$ denote the set of arcs e having the node j as tail or as head, respectively, and x_e is a slack variable for arc e, x is used to abbreviate all the lower level variables (including slack variables).

Now, assume that the cost for passing a street does also depend on toll costs c_e^t which are added to the cost c_e for passing a street. Then, the objective function (1.13) is changed to

$$\sum_{(q,s)\in T} \sum_{e\in E} (c_e + c_e^t) x_e^{qs} \longrightarrow \min. \tag{1.17}$$

Let $\Psi(c^t)$ denote the set of optimal solutions of the problem of minimizing the function (1.17) subject to (1.14)–(1.16), then the problem of computing best toll costs is

$$f(c^t, x) \to \min \text{ s.t. } x \in \Psi(c^t),\ \ c^t \in C. \tag{1.18}$$

Here, C is a set of admissible toll costs and the objective function $f(c^t, x)$ can be used to express different aims, as e.g.:

1. Maximizing the revenue. In this case it makes sense to assume that, for each origin-destination pair $(q, s) \in T$ there is one (directed) path from q to s in the graph which is free of tolls (Didi-Biha et al. [88] and other references),
2. Reducing traffic in ecologically exposed areas (Dempe et al. [58]) or
3. Forcing truck drivers to use trains from one loading station to another one (Wagner [310]).

1.5.3 Optimal Operation Control of a Virtual Power Plant

Müller and Rehkopf investigated in the paper [247] the optimal control of a virtual power plant. This power plant consists of a number of decentralized micro-cogeneration units located in the residential houses of their owners and use natural gas to produce heat and electricity. This is a very efficient possibility for heat and energy supply. Moreover, the micro-cogeneration units can produce much more electricity than used in the houses and the superfluous electricity is injected into the local electricity grid. For that, the residents get a compensation helping them to cover the costs of the micro-cogeneration units. To realize this, the decentralized micro-cogeneration units are joined into a virtual power plant (VP) which collects the superfluous electricity from the decentralized suppliers and sells it on the electricity market. For the VP, which is a profit maximizing unit, it is sensible to sell the electricity to the market in time periods when the revenue on the market is high. Hence, the owner of the VP wants to ask the decentralized suppliers to inject power into the system when the national demand for electricity is large. For doing this he can apply ideas from principal-agent theory establishing an incentive system to motivate the suppliers to produce and inject power into the grid in the desired time periods. In this sense, the owners of the decentralized units are the followers (agents) and the owner of the VP is the leader (principal).

To derive a mathematical model for the VP consider the owners of the micro-cogeneration units first. It is costly to switch the units on implying that it makes sense to restrict the number of time units when the system is switched on. This and failure probability imply that a producing unit should keep working for a minimum time length and the time the system is switched off is also bounded from below after turning it off. To abbreviate these and perhaps other restrictions for the decentralized systems (which are in fact linear inequalities), we use the system $Ay \leq b$.

Under these conditions, the owner of the decentralized systems has to minimize the costs for power and heat generation depending on the costs of the used natural gas, the expenses for switching on the unit and the prices for buying and selling power. Let this function be abbreviated as $f(y)$.

Now, assume that the owners of the decentralized micro-cogeneration units sell their superfluous electricity to the VP which establishes an incentive system to control time and amount of the injected power. Let z denote the premium payed for the power supply. This value is, of course, bounded from below by some values, depending on the expenses of the decentralized units resulting from switching them on and from additional costs of natural gas. Moreover, since the costs for power and heat generation do also depend on the premium payed, the owner of the decentralized units now minimizes a function $\widetilde{f}(y, z)$ subject to the constraints $Ay \leq b$ and some (linear) conditions relating the received bonuses to the working times of the power units. Let $\Psi(z)$ denote the set of solutions of the owners of the decentralized units (production periods of the units, delivered amount of power) depending on the premium z.

Then, the upper level problem of the VP consists of maximizing the revenue from the electricity market for the power supply minus the bonuses payed to the subunits. This function is maximized subject to restrictions from the above conditions that the bonus payed is bounded by some unit costs in the lower level.

1.5.4 Spot Electricity Market with Transmission Losses

In the paper Aussel et al. [5] deregulated spot electricity markets are investigated. This problem is modeled as a generalized Nash equilibrium problem, where each player solves a bilevel optimization problem. To formulate the problem assume that a graph $G = G(V, E)$ is given where each agent (or player) is located at one of the nodes $i \in V$. The arcs E are the electricity lines. The demand D_i at each node is supposed to be known and also that the real cost for generating q_i units of electricity at node i equals $A_i q_i + B_i q_i^2$.

Now, assume that there is an independent system operator (ISO) in the electricity network who is responsible for the trade of electricity. Moreover, each agent bids his cost $b_i q_i^2 + a_i q_i$ of producing q_i units of electricity and his demand to the ISO, who distributes the electricity between the agents. The goal of the ISO is to minimize the total bid costs subject to satisfaction of the demand of the agents. Assume that $L_{ij} t_{ij}^2$ are the thermal losses along $(i, j) \in E$ which are covered equally between agents at nodes i and j if t_{ij} is the amount of electricity delivered along $(i, j) \in E$. Then, the problem of the ISO reads as

$$\begin{aligned} \sum_{i=1}^{|V|} \left(b_i q_i^2 + a_i q_i\right) &\to \min_{q,t} \\ q_i &\geq 0, \quad i \in V \\ q_i - \sum_{k:(i,k)\in E} \left(t_{ik} + 0.5 L_{ik} t_{ik}^2\right) + \sum_{k:(k,i)\in E} \left(t_{ki} - 0.5 L_{ki} t_{ki}^2\right) &\geq D_i, \quad i \in V \\ t_{ij} &\geq 0, \quad (i,j) \in E. \end{aligned}$$

Let $Q(a, b)$ denote the set of optimal solutions of this problem depending on the bid vectors announced by the producers. Then, the agents intend to maximize their profit which equals the difference between the real and the bid function for the production subject to the decision of the ISO. This leads to the following problem:

$$\begin{aligned} \left(b_i q_i^2 + a_i q_i\right) - \left(B_i q_i^2 + A_i q_i\right) &\to \max_{a_i, b_i, q, t} \\ \underline{A}_i \leq a_i &\leq \overline{A}_i \\ \underline{B}_i \leq b_i &\leq \overline{B}_i \\ (q, t) &\in Q(a, b). \end{aligned}$$

This is a bilevel optimization problem with multiple leaders where the leaders act according to a Nash equilibrium.

1.5.5 Discrimination Between Sets

In many situations as e.g. in robot control, character and speech recognition, in certain finance problems as bank failure prediction and credit evaluation, in oil drilling, in medical problems as for instance breast cancer diagnosis, methods for discriminating between different sets are used for being able to find the correct decisions implied by samples having certain characteristics (cf. DeSilets et al. [86], Hertz et al. [144], Mangasarian [215, 216], Shavlik et al. [286], Simpson [289]). In doing so, a mapping $\mathcal{T}_0$ is used representing these samples according to their characteristics as points in the input space (usually the n-dimensional Euclidean space), see Mangasarian [215]. Assume that this leads to a finite number of different points. Now, these points are classified according to the correct decisions implied by their originals. This classification can be considered as a second mapping $\mathcal{T}_1$ from the input space into the output space given by the set of all possible decisions. This second mapping introduces a partition of the input space into a certain number of disjoint subsets such that all points in one and the same subset are mapped to the same decision (via its inverse mapping). For being able to determine the correct decision implied by a new sample we have to find that partition of the input space without knowing the mapping $\mathcal{T}_1$.

Consider the typical case of discriminating between two disjoint subsets $\mathcal{A}$ and $\mathcal{B}$ of the input space $\mathbb{R}^n$ [215]. Then, for approximating this partition, piecewise

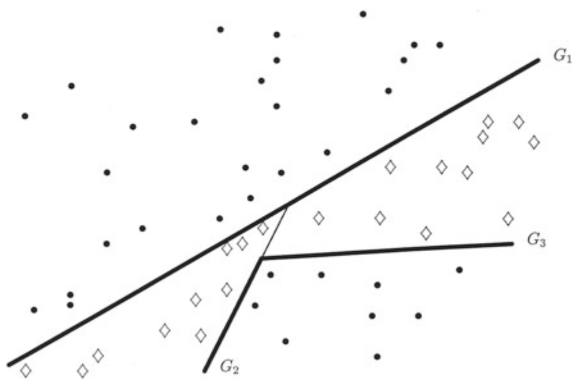

Fig. 1.4 Splitting of R^2 into three subsets each containing points of one of the sets $\mathscr{A}$ respectively $\mathscr{B}$ only

affine surfaces can be determined separating the sets $\mathscr{A}$ and $\mathscr{B}$ (cf. Fig. 1.4 where the piecewise affine surfaces are given by the bold lines). For the computation of these surfaces an algorithm is given by Mangasarian [215] which starts with the computation of one hyperplane (say G_1) separating the sets $\mathscr{A}$ and $\mathscr{B}$ as best as possible. Clearly, if both sets are separable, then a separating hyperplane is constructed. In the other case, there are some misclassified points. Now, discarding all subsets containing only points from one of the sets, the remaining subsets are partitioned in the same way again, and so on. In Fig. 1.4 this means that after constructing the hyperplane G_1 the upper-left half-space is discarded and the lower-right half-space is partitioned again (say by G_2). At last, the lower-right corner is subdivided by G_3.

This algorithm reduces this problem of discriminating between two sets to that of finding a hyperplane separating two finite sets $\mathscr{A}$ and $\mathscr{B}$ of points as best as possible. Mangasarian [216] formulated an optimization problem which selects the desired hyperplane such that the number of misclassified points is minimized. For describing that problem, let A and B be two matrices the rows of which are given by the coordinates of the s and t points in the sets $\mathscr{A}$ and $\mathscr{B}$, respectively. Then, a separating hyperplane is determined by an n-dimensional vector w and a scalar γ as $H = \{x \in \mathbb{R}^n : \langle w, x\rangle = \gamma\}$ with the property that

$$Aw > \gamma e^s, \ \ Bw < \gamma e^t$$

provided that the convex hulls of the points in the sets $\mathscr{A}$ and $\mathscr{B}$ are disjoint. Up to normalization, the above system is equivalent to

$$Aw - \gamma e^s - e^s \geq 0, \ -Bw + \gamma e^t - e^t \geq 0. \tag{1.19}$$

Then, a point in $\mathscr{A}$ belongs to the correct half-space if and only if the given inequality in the corresponding line of the last system is satisfied. Hence, using the step function a_* and the plus function a_+ which are component-wise given as

$$(a_*)_i = \begin{cases} 1 & \text{if } a_i > 0 \\ 0 & \text{if } a_i \leq 0 \end{cases}, \quad (a_+)_i = \begin{cases} a_i & \text{if } a_i > 0 \\ 0 & \text{if } a_i \leq 0 \end{cases}$$

we obtain that the system (1.19) is equivalent to the equation

$$e^{s\top}(-Aw + \gamma e^s + e^s)_* + e^{t\top}(Bw - \gamma e^t + e^t)_* = 0. \tag{1.20}$$

It is easy to see that the number of misclassified points is counted by the left-hand side of (1.20). For $a, c, d, r, u \in \mathbb{R}^l$, Mangasarian [216] characterized the step function as follows:

$$r = a_*, u = a_+ \iff \begin{cases} \begin{pmatrix} r \\ u \end{pmatrix} = \begin{pmatrix} r - u + a \\ r + u - e^l \end{pmatrix}_+ \\ \text{and } r \text{ is minimal in case of uncertainty} \end{cases}.$$

Hence,

$$c = d_+ \iff c - d \geq 0,\ c \geq 0,\ c(c-d) = 0.$$

Using both relations, we can transform the problem of minimizing the number of misclassified points or, equivalently, the minimization of the left-hand side function in (1.20) into the following optimization problem, see Mangasarian [216]

$$\begin{aligned} e^{s\top} r + e^{t\top} s &\to \min_{w,\gamma,r,u,p,v} && \\ u + Aw - \gamma e^s - e^s &\geq 0 & v - Bw + \gamma e^t - e^t &\geq 0 \\ r &\geq 0 & p &\geq 0 \\ r^\top(u + Aw - \gamma e^s - e^s) &= 0 & p^\top(v - Bw + \gamma e^t - e^t) &\geq 0 \\ -r + e^s &\geq 0 & -p + e^t &\geq 0 \\ u &\geq 0 & v &\geq 0 \\ u^\top(-r + e^s) &= 0 & v^\top(-p + e^t) &= 0. \end{aligned}$$

This problem is an optimization problem with linear complementarity constraints, a generalized bilevel optimization problem. Mangasarian has shown in [215] that the task of training neural networks can be modeled by a similar problem.

1.5.6 Support Vector Machines

Closely related to the topic of Sect. 1.5.5 are support vector machines (SVM) (Cortes and Vapnik [45], Vapnik [302]), kernel methods (Shawe-Taylor and Christianini