A Gentle and Incomplete Introduction to Bilevel Optimization

Martin Schmidt, Yasmine Beck
Trier University

June 2021

Contents

List of Algorithms

List of Figures

What Is This About?

These are lecture notes on bilevel optimization. The class of bilevel optimization problems is formally introduced and motivated using examples from different fields. Afterward, the main focus is on how to solve linear and mixed-integer linear bilevel optimization problems. To this end, we first consider various single-level reformulations of bilevel optimization problems with linear or convex follower problems, discuss geometric properties of linear bilevel problems, and study different algorithms for solving linear bilevel problems. Finally, we consider mixed-integer linear bilevel problems, discuss the main obstacles for deriving exact as well as effective solution methods, and derive a branch-and-bound method for solving these problems.

In summary, in these lecture notes, you will learn …

- to recognize bilevel optimization models in real-world applications,
- to properly model these real-world applications using the toolbox of bilevel optimization,
- about the surprising (and mostly challenging) properties of bilevel problems,
- how to reformulate bilevel problems as "ordinary" single-level problems,
- about the obstacles and pitfalls of these single-level reformulations,
- about structural properties of linear bilevel problems,
- how to solve linear bilevel problems,
- about structural properties of mixed-integer linear bilevel problems,
- how to solve mixed-integer linear bilevel problems.

Have fun!

Preface

Bilevel optimization is a wonderful sub-field of mathematical optimization with very many surprising properties of bilevel models, a lot of challenging applications, as well as a growing number of useful algorithms and elegant theoretical aspects.

As we will briefly discuss later on, it is a rather young field that computationally mainly started to develop in the 1980's. My personal interest in bi- and multilevel models started in 2014 and is still growing. That is why I need more colleagues with whom I can work on bilevel problems and more mathematicians with whom I can share and discuss ideas. To propel this, the best way is to give a lecture on that topic and that is where you, the students, and these lecture notes come into play.

I started to plan this lecture in the Covid-19 winter 2020/2021 and give the lecture for the first time in the summer term 2021 at Trier University. There are some research-oriented books on bilevel optimization out there and some surveys, which are also mainly written for a research-oriented audience—but not for students. Thus, designing this lecture and writing these lecture notes has been a rather delicate task. I very much hope that you will enjoy studying bilevel optimization problems and reading these lecture notes. However, I am very sure that tons of things can be improved, corrected, explained in more detail, etc. If you come across such an aspect, *please* let me know. Please give me your feedback so that the lecture and these notes can improve. Being in its first version also means that these lecture notes will contain (although I took care as much as I could) some mistakes. The consequence is that you should trust your own thinking more than these notes—and if this turned out to be a wise choice since you spotted a mistake, please let me know and I will carry out the required corrections as fast as I can. In this sense, these notes are a "living document". I will update it from time to time, correct mistakes, improve explanations, or simply add new sections or chapters on topics that I'm interested in.

One of the most frequent questions of students with respect to a new lecture is on what is the required knowledge for this lecture. My aim is to keep the lecture and these lecture notes as self-contained as possible. However, throughout these notes I assume that you have a solid knowledge in linear optimization (especially including duality theory) and nonlinear optimization (especially including first-order optimality conditions).

Finally, I have to thank many people that I stole content from, that I asked for input regarding some examples or historic notes, or that I paid for producing nice TikZ figures. Thanks a lot, Martine Labbé (I stole, at least, my pricing example and the re-modeling of binary variables using LP-LP bilevel feasibility from you), Ivana Ljubic (I stole, at least, the Knapsack examples and some parts of the chapter on mixed-integer bilevel from our joint papers), Thomas Kleinert (for very many bilevel discussions, many examples that I took from your PhD thesis, the historic notes, for proof-reading many section, etc.), and Fränk Plein (also for very many discussions on the topic, especially on the geometric properties of LP-LP bilevel problems). Last but not least, I have to thank Ioana Molan and Andreas Horländer for many of the TikZ pictures in these notes.

Martin Schmidt
Trier and Augustfehn, summer term 2021

1
Introduction

1.1 What is a Bilevel Problem and Why Should We Care?

Usual optimization problems are single-level problems, which can be denoted as

$$\begin{aligned} \min_{x \in \mathbb{R}^n} \quad & f(x) \\ \text{s.t.} \quad & g(x) \geq 0, \\ & h(x) = 0. \end{aligned}$$

This means that there is only one objective function f, one vector of variables x, and one set of constraints g and h. In particular, this models a situation in which a single decision maker takes all decisions, i.e., decides on the variables of the problem. Studying such a problem is very often appropriate—for instance, if a single dispatcher controls a gas transport network, if a single investment banker decides on the assets in a portfolio, or if a single logistics company decides on its supply chain.

On the other hand, there are very many situations in our day-to-day life that are different. Often, we, as a decision maker, make a decision while anticipating the (rational, i.e., optimal) reaction of another decision maker, whose decision depends on ours. Moreover, also our outcome (or in more mathematical terms, our objective function and/or feasible set) depends on the reaction of the other decision maker. Formalizing this situation leads to hierarchical or bilevel optimization problems. Before we formally define this class of problems, let us consider some informally stated examples.

Example 1.1 (Pricing). *One of the richest class of applications of bilevel optimization are pricing problems. The first decision maker, which we will also call* leader *in the following, decides on a price of a certain good (or maybe*

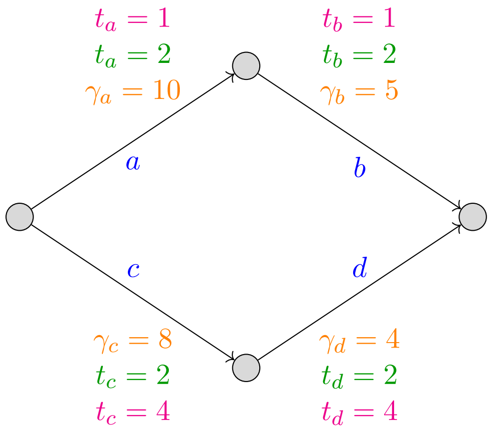

Figure 1.1: Different tolls lead to different route choices, travel costs, and toll revenues

on different prices for multiple goods) to maximize her revenue from selling these goods. The second decision maker, called follower *in what follows, then decides on purchasing the goods of the leader to generate some utility.*

Thus, the leader's decision depends on the optimal reaction of the follower and the decision of the follower, of course, depends on the (pricing) decisions of the leader.

Example 1.2 (Toll Setting)**.** *Imagine a transportation network, e.g., the German highway network, via which a set of drivers want to reach their destination, starting from their origin. Usually, the objective of these travelers is to travel from their origin to their destination at minimum costs. In this situation, costs can, e.g., be travel time, toll costs, or a combination of both.*

On the other hand, there usually is a toll setting agency, which decides on the tolls imposed on certain parts of the highway system. This toll setting agency wants to maximize the revenues based on the tolls and the travelers, afterward, minimize their traveling costs. The toll setting agency is the leader and the travelers are the followers in this setting. Again, the leader anticipates the optimal reaction of the followers, whereas the followers' decisions obviously depend on the decision of the leader. Whereas we only had one follower in the pricing example, we now have multiple followers. The former is called a single-leader single-follower *problem, whereas the latter is called a* single-leader multi-follower *problem.*

Exercise 1.1. In the toll setting example of Figure 1.1, the pink solution is worse for both players, i.e., for the follower and for the leader: The leader earns less and the follower has larger route costs compared to the "better" green solution. Does this always need to be the case—or are there settings in which a solution is better for the leader but worse for the follower (or vice versa)?

Example 1.3 (Energy Markets)**.** *Another very rich class of applications for bilevel optimization is the energy sector—especially the sub-field of energy market modeling. In many countries of the world, e.g., in Germany, electricity is traded via auctions at an energy exchange.*[1] *The rules that determine how this auction is organized is typically decided on by the state government or some regulatory authority. The aim of these rules are usually to obtain market outcomes that are optimal in terms of social welfare; see, e.g., Mas-Colell et al. (1995) for some economic background. Depending on these rules, producers and consumers trade electricity at the exchange.*

As before, the decision of the leader—here, the regulatory authority—depends on the anticipation of the followers' decisions—here, the decisions of the firms trading on the market. Moreover, the firms' decisions depend on the market regime, i.e., on the decision of the leader.

Example 1.4 (Critical Infrastructure Defense)**.** *Bilevel optimization is also of great importance for critical infrastructure defense. Imagine a set of important buildings such as airports, central stations, market squares, etc. that might be potential targets of attacks by terrorists. This infrastructure needs to be protected by, e.g., police officers. However, there are not enough officers so that every building can be protected. Terrorists then decide to attack one or some of these locations based on their expectation on which buildings are protected and which are not. Assuming some utility function*[2] *for both the police (also called defenders in this setting) and the terrorists (also called attackers in this setting), the police (as the leader) assigns officers to certain buildings in order to achieve the worst outcome for the terrorists (acting as followers).*

Example 1.5 (Interdiction Problems)**.** *In discrete bilevel optimization, maybe the most heavily studied problem is the interdiction problem. Here, the leader is a defender that interdicts certain resources of the follower so that they cannot be used anymore by the follower. Many of these problems are defined on graphs. For instance, the follower might want to find a shortest path in a graph from an origin to a destination. The leader, acting as the interdictor, can interdict some of the arcs in the graph so that they cannot be part of a feasible path of the follower anymore. The number of interdicted arcs is further constrained by an interdiction budget of the leader.*

mention drug smuggling?!

This broad range of examples from security, counter-terrorism, drug smuggling, energy markets, revenue management, and transport makes it obvious that it is important to formally model, analyze, and solve the respective mathematical models.

Maybe add example from data science/ML?

[1] See, e.g., https://www.eex.com/en/ for the European Energy Exchange in Leipzig, Germany.

[2] . . . although this might sound cynical.

1.2 A Bit More Formal, Please

Since we are now convinced that it makes sense to study hierarchical optimization problems because they often appear in practice, let us formally define them now.

Definition 1.6 (Bilevel Optimization Problem)**.** *A bilevel optimization problem reads*

$$\begin{aligned} \min_{x \in X, y} \quad & F(x, y) && \text{(1.1a)} \\ \text{s.t.} \quad & G(x, y) \geq 0, && \text{(1.1b)} \\ & y \in S(x), && \text{(1.1c)} \end{aligned}$$

where $S(x)$ is the set of optimal solutions of the x-parameterized problem

$$\begin{aligned} \min_{y \in Y} \quad & f(x, y) && \text{(1.2a)} \\ \text{s.t.} \quad & g(x, y) \geq 0. && \text{(1.2b)} \end{aligned}$$

Problem (1.1) is the so-called upper-level (or the leader's) problem and Problem (1.2) is the so-called lower-level (or the follower's) problem, which is parameterized by the leader's decision x. Moreover, the variables $x \in \mathbb{R}^{n_x}$ are the upper-level variables (or leader's decisions) and $y \in \mathbb{R}^{n_y}$ are lower-level variables (or follower's decisions). The objective functions are given by $F, f : \mathbb{R}^{n_x} \times \mathbb{R}^{n_y} \to \mathbb{R}$ and the constraint functions by $G : \mathbb{R}^{n_x} \times \mathbb{R}^{n_y} \to \mathbb{R}^m$ as well as $g : \mathbb{R}^{n_x} \times \mathbb{R}^{n_y} \to \mathbb{R}^\ell$. The sets $X \subseteq \mathbb{R}^{n_x}$ and $Y \subseteq \mathbb{R}^{n_y}$ are typically used to denote integrality constraints. For instance, $Y = \mathbb{Z}^{n_y}$ makes the lower-level problem an integer program. In what follows, we call upper-level constraints $G_i(x, y) \geq 0$, $i \in \{1, \dots, m\}$, *coupling constraints* if they explicitly depend on the lower-level variable vector y. Moreover, all upper-level variables that appear in the lower-level constraints are called *linking variables.*

We use the nomenclature that the bilevel problem (1.1) is called an "UL-LL problem" where UL and LL can be LP, QP, MILP, MIQP, etc. if the upper-/lower-level problem is a linear, a quadratic, a mixed-integer linear, a mixed-integer quadratic, etc. program in both the variables of the leader and the follower. If the concrete specification of both levels is not required, we also use a shorter nomenclature and say, e.g., that the problem is a bilevel LP, if both levels are LPs.

Instead of using the point-to-set mapping S one can also use the so-called *optimal value function*

$$\varphi(x) := \min_{y \in Y} \{f(x, y) \colon g(x, y) \geq 0\} \tag{1.3}$$

and re-write Problem (1.1) as

$$
\begin{aligned}
\min_{x \in X, y \in Y} \quad & F(x,y) && \text{(1.4a)}\\
\text{s.t.} \quad & G(x,y) \geq 0, \ g(x,y) \geq 0, && \text{(1.4b)}\\
& f(x,y) \leq \varphi(x), && \text{(1.4c)}
\end{aligned}
$$

to which we will refer to as the optimal-value-function or value-function reformulation.

Exercise 1.2. Consider the linear bilevel problem

$$
\begin{aligned}
\min_{x,y} \quad & x + y \\
\text{s.t.} \quad & -x - 2y \geq -10, \\
& 2x - y \geq 0, \\
& -x + 2y \geq 0, \\
& x \geq 0, \\
& y \in \arg\min_{\bar{y}} \{\bar{y} \colon x + \bar{y} \geq 3\}.
\end{aligned}
\tag{1.5}
$$

(i) Plot the linear inequalities in a coordinate system.

(ii) Determine the bilevel feasible set of Problem (1.5).

(iii) Reformulate Problem (1.5) into a single-level problem using the optimal value function.

(iv) Determine the optimal solution of the value-function reformulation in which the lower-level constraint $x + y \geq 3$ is omitted.

Definition 1.7 (Shared Constraint Set)**.** *The set*

$$\Omega := \{(x,y) \in X \times Y \colon G(x,y) \geq 0, \ g(x,y) \geq 0\}$$

is called the shared constraint set*. Its projection onto the x-space is denoted by*

$$\Omega_x := \{x \colon \exists y \text{ with } (x,y) \in \Omega\}.$$

Definition 1.8 (Bilevel Feasible Set; Inducible Region)**.** *The set*

$$\mathcal{F} := \{(x,y) \colon (x,y) \in \Omega, \ y \in S(x)\}$$

is called the bilevel feasible set *or* inducible region.

Definition 1.9 (High-Point Relaxation)**.** *The problem of minimizing the upper-level objective function over the shared constraint set, i.e.,*

$$
\begin{aligned}
\min_{x,y} \quad & F(x,y) \\
s.t. \quad & (x,y) \in \Omega,
\end{aligned}
$$

is called the high-point relaxation (HPR) *of Problem 1.1.*

Note that the high-point relaxation is identical to the original bilevel problem (1.1) except for the constraint $y \in S(x)$, i.e., except for the lower-level optimality. Thus, it is obviously a relaxation of (1.1).

Next, we recap some of the previously discussed examples and formally state the corresponding bilevel problems.

Example 1.10 (The Pricing Example Revisited)**.** *A first bilevel pricing problem with linear constraints, linear upper-level objective and bilinear lower-level objective has been proposed by Bialas and Karwan (1984). The following problem considered in Labbé et al. (1998) provides a general framework for such pricing problems:*

$$\max_{x, y=(y_1, y_2)} \quad x^\top y_1 \tag{1.6a}$$

$$\text{s.t.} \quad Ax \leq a, \tag{1.6b}$$

$$y \in \arg\min_{\bar{y}} \left\{ (x + d_1)^\top \bar{y}_1 + d_2^\top \bar{y}_2 \colon D_1 \bar{y}_1 + D_2 \bar{y}_2 \geq b \right\}. \tag{1.6c}$$

The vector y of lower-level variables is partitioned into two sub-vectors y_1 and y_2, called plans, that specify the levels of some activities such as purchasing goods or services. The upper-level player influences the activities from plan y_1 through the price vector x that is additionally imposed onto y_1. By doing so, the goal of the leader is to maximize her revenue given by $x^\top y_1$. The price vector x is subject to linear constraints that may, among others, impose lower and upper bounds on the prices. The vectors d_1 and d_2 represent linear disutilities faced by the lower-level player when executing the activity plans y_1 as well as y_2. Note that d_2 may also encompass the price for executing the activities not influenced by the upper-level player. These activities may, e.g., be substitutes offered by competitors for which prices are known and fixed. The lower-level player determines his activity plans y_1 and y_2 to minimize the sum of total disutility and the price paid for plan y_1 subject to linear constraints. Remark that if the model allows negative prices then it implicitly permits subsidies, which may be appropriate, e.g., in the context of a central agency determining taxes. In order to avoid the situation in which the leader would maximize her profit by setting prices to infinity for these activities y_1 that are essential, one may assume that the set $\{y_2 \colon D_2 y_2 \geq b\}$ is non-empty. Indeed, in this case, there exists a feasible point for the lower level that does not use any activity influenced by the upper level.

Example 1.11 (Bilevel Knapsack Interdiction)**.** *In the following, we consider the bilevel knapsack interdiction problem that has been investigated by Caprara et al. (2016). In this setting, the leader and the follower own a private knapsack which is filled with items from a common set of items $[n] := \{1, \dots, n\}$. For each item $i \in [n]$, we denote p_i as the corresponding profit and with v_i and w_i, we associate the item's weights for the leader and* *the follower, respectively. The leader's aim is to minimize the follower's maximum profit by prohibiting the usage of certain items by the follower. For this purpose, the leader first selects a subset of items respecting her so-called interdiction budget B. Then, the follower can choose from the remaining items maximizing her profit considering the knapsack capacity C. The bilevel knapsack interdiction problem is formally stated as follows*

$$\min_{x} \quad p^\top y \tag{1.7a}$$
$$\text{s.t.} \quad v^\top x \leq B, \tag{1.7b}$$
$$x \in \{0,1\}^n, \tag{1.7c}$$
$$y \in \arg\max_{y'} \left\{ p^\top y' : y' \in Y(x) \right\}, \tag{1.7d}$$

with $B, C \in \mathbb{R}$, $p, v, w \in \mathbb{R}^n$ and $Y(x) = \{y \in \{0,1\}^n : w^\top y \leq C,\, y_i \leq 1 - x_i,\, i \in [n]\}$ denotes the set of feasible decisions of the follower, which is parameterized by the leader's decision x. In Problem (1.7), both players consider the same objective function that is optimized in opposite directions.

Example 1.12 (An Academic Example; see Kleinert (2021))**.** *We now consider the bilevel problem*

$$\min_{x,y} \quad F(x,y) = x + 6y \tag{1.8a}$$
$$\text{s.t.} \quad -x + 5y \leq 12.5, \tag{1.8b}$$
$$x \geq 0, \tag{1.8c}$$
$$y \in S(x), \tag{1.8d}$$

where the lower-level optimal solutions $S(x)$ are given by the linear problem

$$\min_{y} \quad f(x,y) = -y \tag{1.9a}$$
$$\text{s.t.} \quad 2x - y \geq 0, \tag{1.9b}$$
$$-x - y \geq -6, \tag{1.9c}$$
$$-x + 6y \geq -3, \tag{1.9d}$$
$$x + 3y \geq 3. \tag{1.9e}$$

Both levels are linear optimization problems and all variables are continuous. Thus, we consider an LP-LP bilevel problem, which is the easiest class of bilevel models. The problem of this example is illustrated in Figure 1.2. The figure reveals several interesting and important obstacles of bilevel programming:

(a) The feasible region of the follower problem corresponds to the gray area. Thus, the follower problem—and therefore the bilevel problem—is infeasible for certain decisions of the leader, e.g., $x = 0$.

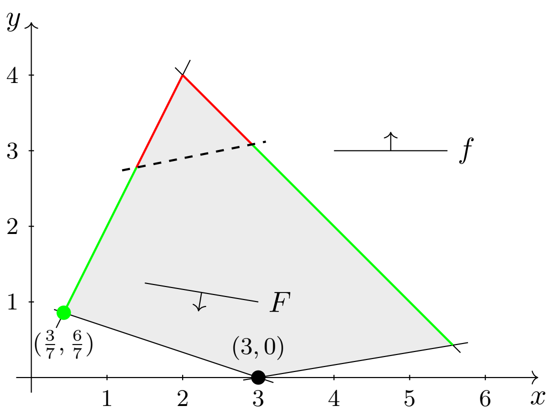

Figure 1.2: The shared constrained set (gray area), the nonconvex set of optimal follower solutions (green and red lines) lifted to the x-y-space, and the nonconvex and disconnected bilevel feasible set (green lines) of the bilevel problem (1.8) and (1.9).

(b) The set $\{(x, y) : x \in \Omega_x, y \in S(x)\}$ *denotes the optimal follower solutions lifted to the* x-y-*space, and is given by the green and red facets. Obviously, this set is nonconvex.*

(c) The single leader constraint indicated by the dashed line renders certain optimal responses of the follower infeasible. Thus, the bilevel feasible region $\mathcal{F}$ *corresponds to the green facets. Consequently, the feasible set of Problem* (1.8) *is not only nonconvex but also disconnected.*

(d) The optimal solution of Problem (1.8) *is* $(3/7, 6/7)$ *with objective function value* $39/7$. *In contrast, ignoring the follower's objective, i.e., solving the high-point relaxation, yields the optimal solution* $(3, 0)$ *with objective function value* 3. *Note that the latter point is not bilevel feasible.*

This example shows that bilevel optimization problems are inherently difficult to solve since even their easiest instantiation are nonconvex optimization problems.

Before we really start to analyze bilevel problems in detail, let us mention some other general literature, where you can find additional (and also much more detailed) information. First, we refer to the survey articles by Colson et al. (2005, 2007) and Kleinert, Labbé, Ljubić, et al. (2021) and the books by Bard (1998), Dempe (2002), and Dempe, Kalashnikov, et al. (2015). Other very early survey articles include Anandalingam and Friesz (1992), Ben-Ayed (1993), Kolstad (1985), and Vicente and Calamai (1994) as well as Wen and Hsu (1991) regarding the field of linear bilevel optimization. Last but not least, Dempe (2020) contains, to the best of our knowledge, the largest annotated list of references in the field of bilevel optimization.

Remark 1.13. *For modeling practically relevant applications, one is, of course, not restricted to use "only" two levels like we do here in bilevel optimization. If the optimization problem has constraints that again contain an optimization problem that has constraints, which again contain an optimization problem ... and so on, the problem is called a* multilevel optimization problem*. You might imagine that three or four levels do not make the problem easier to analyze and solve—especially since we already saw that even bilevel problems are extremely challenging in their easiest instantiation. Nevertheless, multilevel optimization is often required to model real-world situations; see, e.g., Ambrosius et al. (2020), Grimm, Kleinert, et al. (2019), Grimm, Martin, et al. (2016), Grimm, Schewe, et al. (2019), Kleinert and Schmidt (2019b), and Schewe et al. (2020).*

Exercise 1.3 (See Dempe, Kalashnikov, et al. (2015))**.** A research team wants to optimize the production of chemical substances. The aim is to produce $m \in \mathbb{N}$ substances as a result of chemical reactions in a reactor. The reactor is operated at a certain temperature T and a certain pressure p. There are $n \in \mathbb{N}$ different reactants, which can be added to the reactor to produce the desired substances. If there are no net changes in the amount of reactants and products, the reaction is said to be in a chemical equilibrium. In this state, the function

$$f(y, p, T) = \sum_{i=1}^{m} c_i(p, T) y_i + \sum_{i=1}^{k} y_i \ln \frac{y_i}{\sum_{j=1}^{m} y_j}$$

takes its minimal value. Here, y_i denotes the mass of substance $i \in \{1, \dots, m\}$ and $k \leq m$ is the number of gaseous substances. In particular, the chemical equilibrium is *uniquely* determined by temperature, pressure, and the composition of the reactants. Further, the following chemical laws apply:

- *Law of Constant Composition*: A chemical compound always contains the same elements in the same mass proportion. Here, the mass proportions in the resulting substances are given by a matrix A of appropriate dimension. Each row of A corresponds to a chemical element and each column gives the amount of the different elements in the substances. In the same manner, the matrix B gives the mass proportions corresponding to the reactants.
- *Principle of Mass Conservation*: In chemical reactions, mass is neither created nor destroyed. Thus, the sum of the masses of the reactants must equal the sum of the masses of the products, i.e., $Ay = Bx$. Here, x denotes the vector containing the masses of the reactants.

It is assumed that the mass of all substances is nonnegative. Furthermore, there are physical restrictions regarding the pressure and the temperature in the reactor as well as the amount of available reactants, i.e., $(p, T, x) \in X$ for a properly chosen set X. The goal of the research team is to minimize the linear expression $d^\top y$. The vector $d \in \mathbb{R}^m$ captures the aim to compose a mixture of reactants such that the amount of the desired substances is as large as possible, while the amount of (toxic) by-products is as small as possible.

(i) Formulate a bilevel problem that models the optimal chemical equilibrium.

(ii) Who acts as the leader and "who" acts as the follower?

(iii) Are there *linking variables* and/or *coupling constraints*?

Exercise 1.4. Prof. Jones is a collector of rare artifacts. He recently returned from an adventure in South America from which he also brought some valuable treasures. Alice and Bob, a famous robber couple, thus plan their next raid on Prof. Jones. Alice is the mastermind of the duo and came across some inside knowledge about the number $n \in \mathbb{N}$, the values $v \in \mathbb{R}^n$, and the weights $w \in \mathbb{R}^n$ of the artifacts in Prof. Jones' possession as well as the security measures at his mansion. The property is heavily protected but Bob, the henchman of the duo, can access the building by climbing through a bathroom window on the second floor. For climbing, Bob needs to have his hands free. Therefore, he will carry the stolen items in a backpack. Alice's task is to buy a backpack of appropriate size $b \in \mathbb{R}$ that Bob will use in the raid. The backpack should not be too big and not too small, i.e., $b_l \leq b \leq b_u$ with $0 \leq b_l \leq b_u \in \mathbb{R}$. The costs $c \in \mathbb{R}$ for the backpack are assumed to be proportional to its size. Bob cannot split any items, he can either take item $i \in \{1, \dots, n\}$ or leave it, i.e., $x_i = 1$ or $x_i = 0$. Moreover, Bob can only take a subset of items such that the capacity of the backpack is not exceeded, i.e., $w^\top x \leq b$. The aim of the robber duo is to maximize their profits, which is the difference between the value of the stolen items $v^\top x$ and the costs for buying the backpack cb.

(i) Formulate a bilevel problem that models the optimal raid strategy.

(ii) Who acts as the leader and who acts as the follower?

(iii) Are there *linking variables* and/or *coupling constraints*?

(iv) Unfortunately, the raid on Prof. Jones went horribly wrong. Bob got caught by the police and he had to spend four years in prison. After his release, Bob still holds a grudge against Alice since she got away with impunity. Therefore, they no longer work together but do solo robberies instead. During his time in prison, fellow criminals helped Bob to work on his robbery skills, which now gives him an advantage over Alice. This means that he will always be first to arrive at a new potential robbery target. Unlike Alice, Bob is not interested in the value of his stolen items. His aim is to leave the items for Alice to steal that yield the worst possible outcome for her. Alice and Bob own a backpack of size b_A and b_B, respectively, which they use for their raids.

What needs to be adapted in the previous modeling to account for the change of circumstances?

Exercise 1.5 (See Dempe, Kalashnikov, et al. (2015)). Consider a simplified energy network that is given by a directed graph $G = (V, E)$. At each node $v \in V$, exactly one producer and exactly one consumer is located. The edges $e \in E$ denote the electricity lines in the network. The demands d_v as well as the real costs $A_v q_v + B_v q_v^2$ for producing $q_v \geq 0$ units of electricity are known for all $v \in V$. The flow f_e along the electricity lines $e \in E$ is nonnegative and each line has a maximal transmission capacity $c_e \geq 0$. Thus, $0 \leq f_e \leq c_e$ has to hold for all electricity lines $e \in E$. It is assumed that there are no transmission losses when electricity is transported in the energy network. The network is regulated by an *Independent System Operator* (ISO) who is responsible for the trade and transport of electricity. Each producer bids the costs $a_v q_v + b_v q_v^2$ for producing q_v units of electricity in an auction organized by the ISO. The ISO ensures that the demand of each consumer is satisfied, i.e.,

$$q_v - \sum_{e \in \delta^{\text{out}}(v)} f_e + \sum_{e \in \delta^{\text{in}}(v)} f_e \geq d_v$$

holds for all $v \in V$. Here, $\delta^{\text{in}}(v)$ denotes the incoming and $\delta^{\text{out}}(v)$ denotes the outgoing electricity lines at node $v \in V$. The ISO distributes the units of electricity to be produced among the producers such that the demand is satisfied at the lowest costs. The producers aim to maximize their profits for the produced quantities of electricity, which is the difference between the supply bids and the real costs. Further, the bids need to meet certain requirements, i.e., $\underline{A}_v \leq a_v \leq \bar{A}_v$ as well as $\underline{B}_v \leq b_v \leq \bar{B}_v$ have to hold for all $v \in V$.

(i) Formulate a bilevel problem that models this simplified energy market.

(ii) Who acts as the leader and who acts as the follower? Do you notice anything special?

(iii) Are there *linking variables* and/or *coupling constraints*?

Exercise 1.6. Consider the linear bilevel problem

$$\begin{aligned}
\min_{x,y} \quad & -x - 2y \\
\text{s.t.} \quad & x + 4y \geq 12, \\
& x \geq 0, \\
& y \in \arg\min_{\bar{y}} \left\{ \bar{y} \colon x + \bar{y} \geq 5,\ -x - \bar{y} \geq -10,\ -x + 4\bar{y} \geq 0 \right\}.
\end{aligned} \tag{1.10}$$

(i) Plot the linear inequalities in a coordinate system.

(ii) What is the shared constraint set Ω? Determine its projection onto the x-space Ω_x.

(iii) Determine the solution of the high-point relaxation (HPR) of Problem (1.10).

(iv) Suppose that the leader sticks with the solution found in (iii). What would be the optimal response of the follower?

(v) Determine the bilevel feasible set of Problem (1.10). Can you say something about its properties?

(vi) Determine the optimal solution of Problem (1.10).

Exercise 1.7. Consider the linear bilevel problem

$$\begin{aligned}
\min_{x,y} \quad & -x + y \\
\text{s.t.} \quad & 0 \leq x \leq 8, \\
& y \in \arg\min_{\bar{y}} \left\{ d\bar{y} \colon x - 3\bar{y} \geq -12,\ -x - \bar{y} \geq -8 \right\}.
\end{aligned} \tag{1.11}$$

For which choices of $d \in \mathbb{R}$ is Problem (1.11) infeasible, unbounded, or solvable? What can you say about the high-point relaxation (HPR) regarding infeasibility, unboundedness, or solvability?

Exercise 1.8. Consider the linear bilevel problem

$$\begin{aligned} \min_{x,y} \quad & x - 2y \\ \text{s.t.} \quad & x + 2y \geq 12, \\ & -x + 2y \geq -2, \\ & y \in \arg\min_{\bar{y}} \{\bar{y} \colon x - \bar{y} \geq -3,\ -2x - \bar{y} \geq -24,\ \bar{y} \geq 0\}. \end{aligned} \tag{1.12}$$

(i) Plot the linear inequalities in a coordinate system.

(ii) What is the shared constraint set Ω? Determine its projection onto the x-space Ω_x.

(iii) Reformulate Problem (1.12) into a single-level problem using the optimal value function.

(iv) Determine the solution of the high-point relaxation (HPR) of Problem (1.12).

(v) What is the set of optimal follower solutions projected onto the x-y-space?

(vi) Determine the bilevel feasible set. Can you say something about the solvability of Problem (1.12)?

(vii) If we move the coupling constraints to the lower level, what changes?

1.3 A Brief History of Bilevel Optimization

Bilevel optimization dates back to the seminal publications on leader-follower games by von Stackelberg (1934, 1952). The formulation introduced in the last section was first used in Bracken and McGill (1973) in the context of a military application regarding the cost-minimal mix of weapons. Another very early discussion of multilevel, or, in particular, two-level problems can be found in Candler and Norton (1977). Candler and Norton (1977) recognized already in the early days of bilevel optimization that such problems are very challenging to solve. More precisely, the authors noticed that even in the "simplest case" of continuous variables and linear objective functions and constraints, the feasible set of bilevel problems may be nonconvex and disconnected. In fact, formal complexity results, which were derived much later, state that even this "easiest" class of linear bilevel problems is already strongly NP-hard. Candler and Norton (1977) also proposed an enumerative algorithm for linear bilevel problems similar to the simplex method, but they had "no doubt others could develop more efficient algorithms". After Bialas and Karwan (1978) proposed the so-called kth-best algorithm—another enumerative and simplex-inspired method—Fortuny-Amat and McCarl (1981) introduced a game-changing approach for convex-quadratic bilevel problems in 1981. They replaced the follower problem by its necessary and sufficient Karush–Kuhn–Tucker (KKT) conditions to derive an equivalent single-level problem that can be further reformulated and tackled by standard mixed-integer solvers. We will discuss this approach in detail in Chapter 4. Bard and Moore (1990), Bard (1988), Edmunds and Bard (1991), and Hansen et al. (1992) picked up the idea later and this approach is still standard for solving bilevel problems with convex follower problems today. Alternative approaches, e.g., penalty methods or descent approaches, have been proposed by Anandalingam and White (1990) as well as by Savard and Gauvin (1994). In the 1990s, the largest instances of linear bilevel problems that have been solved consisted of 250 leader variables, 150 follower variables, and 150 follower constraints; see Hansen et al. (1992). Although cutting planes were derived in the following years, see Audet, Haddad, et al. (2007) and Audet, Savard, et al. (2007), computational linear and convex bilevel optimization did not attract much attention in the 2000s and not many computational results have been reported. Moore and Bard (1990) developed a branch-and-bound approach for bilevel problems with mixed-integer follower problems and also reported some first numerical results already in 1990. However, only very little computational progress has been reported until DeNegre and Ralphs (2009) introduced a branch-and-cut approach for purely integer bilevel problems in 2009. In our opinion, this work can be considered a tipping point for computational bilevel optimization and many computationally oriented works for various classes of bilevel problems appeared in the last ten years.

Exercise 1.9. Read the early publications by von Stackelberg (1934, 1952), Bracken and McGill (1973), as well as Candler and Norton (1977). If you have time and find this interesting, read Fortuny-Amat and McCarl (1981) as well; we will come back to their paper in this lecture anyway.

2

Mathematical Background

In this section, we briefly review the basics of linear and nonlinear optimization that we will later need during our studies of bilevel optimization problems. For both types of problems we mainly discuss the corresponding duality theorems as well as the classic necessary (and sometimes also sufficient) first-order optimality conditions. All theoretical results in this chapter are given without proofs but we refer to seminal textbooks in which these can be found.

2.1 Linear Optimization

We consider linear optimization problems of the form

$$\begin{aligned} \min_{x \in \mathbb{R}^n} \quad & c^\top x && \text{(2.1a)} \\ \text{s.t.} \quad & Ax = b, && \text{(2.1b)} \\ & x \geq 0, && \text{(2.1c)} \end{aligned}$$

where $c \in \mathbb{R}^n$, $b \in \mathbb{R}^m$, and $A \in \mathbb{R}^{m \times n}$. This is the so-called standard form of a linear optimization problem (LP). It can be shown that every linear optimization problem can be written in this way by introducing suitable variable splittings and/or slack variables.

As usual, we call a vector $x \in \mathbb{R}^n$ *feasible* if it satisfies the constraints, i.e., if $Ax = b$ and $x \geq 0$ holds. Moreover, we call the problem *bounded* if there exists a constant $C \in \mathbb{R}$ with

$$c^\top x \geq C \quad \text{for all feasible } x.$$

For linear optimization problems, we have the following nice existence result.

Theorem 2.1. *The linear optimization problem* (2.1) *is either infeasible, unbounded, or solvable.*

Exercise 2.1. Just as a refresher: Prove Theorem 2.1.

The *dual problem* of the linear optimization problem (2.1) is the linear problem

$$\max_{\lambda \in \mathbb{R}^m} \quad b^\top \lambda \tag{2.2a}$$
$$\text{s.t.} \quad A^\top \lambda \leq c. \tag{2.2b}$$

Here and in what follows, we will use Latin letters for primal variables and Greek letters for dual variables.

In bilevel optimization, we often make use of optimality conditions to replace optimization problems with these conditions. For linear optimization problems, these conditions are usually given by the strong duality theorem. However, we first state the weak duality theorem.

Theorem 2.2. *Let $x \in \mathbb{R}^n$ be a feasible point of the primal problem* (2.1) *and let $\lambda \in \mathbb{R}^m$ be a feasible point of the dual problem* (2.2). *Then,*

$$b^\top \lambda \leq c^\top x \tag{2.3}$$

holds.

Exercise 2.2. Another refresher: Prove Theorem 2.2.

Next, we state the strong duality theorem.

Theorem 2.3. *Consider the pair* (2.1) *and* (2.2) *of primal and dual LPs. Then, the following statements are equivalent:*

(a) The problems (2.1) *and* (2.2) *both are feasible.*

(b) The problems (2.1) *and* (2.2) *both have optimal solutions $x^* \in \mathbb{R}^n$ and $\lambda^* \in \mathbb{R}^m$ and*

$$c^\top x^* = b^\top \lambda^* \tag{2.4}$$

holds.

(c) The problems (2.1) *and* (2.2) *both have a finite optimal objective value.*

Finally, we also state the complementarity slackness theorem.

Theorem 2.4. *Consider the pair* (2.1) *and* (2.2) *of primal and dual LPs. Moreover, let $\bar{x} \in \mathbb{R}^n$ be feasible for* (2.1) *and let $\bar{\lambda} \in \mathbb{R}^m$ be feasible for* (2.2). *Then, the following statements are equivalent:*

(a) $\bar{x} \in \mathbb{R}^n$ is optimal for (2.1) *and $\bar{\lambda} \in \mathbb{R}^m$ is optimal for* (2.2).

(b) It holds

$$(c - A^\top \bar{\lambda})^\top \bar{x} = 0.$$

(c) For all components $\bar{x}_j$ of the primal solution it holds

$$\bar{x}_j > 0 \implies A_{\cdot j}^\top \bar{\lambda} = c_j,$$

i.e., if the primal variable has slack ($\bar{x}_j > 0$), the corresponding jth dual inequality is active.

Here and in what follows, $M_{\cdot j}$ represents the jth column of the matrix M and $M_{i\cdot}$ represents its ith row.

For later reference, we also state the next two corollaries, which we will often use to tackle bilevel optimization problems with linear lower-level problems.

Corollary 2.5. *The primal optimization problem (2.1) is equivalent to the system*

$$\begin{aligned} Ax &= b, && (2.5a)\\ x &\geq 0, && (2.5b)\\ A^\top \lambda &\leq c, && (2.5c)\\ b^\top \lambda &\geq c^\top x. && (2.5d) \end{aligned}$$

Here, "equivalent" means the following: Whenever x is an optimal solution of the LP (2.1), then there exists a dual vector λ so that (x, λ) satisfy (2.5) and whenever there exists (x, λ) that satisfy (2.5), then x is an optimal solution of (2.1).

Proof. The claim follows directly from the strong duality theorem. □

Corollary 2.6. *The primal optimization problem (2.1) is equivalent to the system*

$$\begin{aligned} Ax &= b, && (2.6a)\\ x &\geq 0, && (2.6b)\\ c - A^\top \lambda &\geq 0, && (2.6c)\\ x_i(c - A^\top \lambda)_i &= 0, \quad i \in \{1, \dots, n\}. && (2.6d) \end{aligned}$$

Here, "equivalent" means the following: Whenever x is an optimal solution of the LP (2.1), then there exists a dual vector λ so that (x, λ) satisfy (2.6) and whenever there exists (x, λ) that satisfy (2.6), then x is an optimal solution of (2.1).

Proof. The claim follows directly from the complementarity slackness theorem. □

Exercise 2.3. Consider the linear bilevel problem

$$\begin{aligned} \min_{x,y} \quad & c^\top x + d^\top y \\ \text{s.t.} \quad & Ax + By \geq a, \\ & Cx + Dy = b, \\ & x \geq 0, \end{aligned} \tag{P}$$

with $x, c \in \mathbb{R}^{n_x}$, $y, d \in \mathbb{R}^{n_y}$, $A \in \mathbb{R}^{m \times n_x}$, $B \in \mathbb{R}^{m \times n_y}$, $a \in \mathbb{R}^m$, $C \in \mathbb{R}^{\ell \times n_x}$, $D \in \mathbb{R}^{\ell \times n_y}$, and $b \in \mathbb{R}^\ell$.

(i) State the dual problem of (P) and use λ and μ for the respective dual variables.

(ii) Let $(\bar{x}, \bar{y})$ be feasible for Problem (P) and let $(\bar{\lambda}, \bar{\mu})$ be feasible for the dual problem of (P). Which inequality follows from weak duality?

2.2 Nonlinear Optimization

In this section, we consider the situation in which some of the constraints or the objective function can be nonlinear. The general form of such a nonlinear optimization problem (NLP) reads

$$\min_{x \in \mathbb{R}^n} \quad f(x) \tag{2.7a}$$

$$\text{s.t.} \quad g_i(x) \geq 0, \quad i \in I = \{1, \dots, m\}, \tag{2.7b}$$

$$h_j(x) = 0, \quad j \in J = \{1, \dots, p\}. \tag{2.7c}$$

We assume that the objective function $f : \mathbb{R}^n \to \mathbb{R}$ as well as the constraint functions $g_i : \mathbb{R}^n \to \mathbb{R}$, $i \in I$, and $h_j : \mathbb{R}^n \to \mathbb{R}$, $j \in J$, are continuously differentiable. The feasible set is denoted by $\mathcal{F}$.

Definition 2.7 (Local Minimizer)**.** *A point $x^* \in \mathbb{R}^n$ is called a* local minimizer *of Problem (2.7) if x^* is feasible and if an $\varepsilon > 0$ exists such that $f(x) \geq f(x^*)$ for all $x \in \mathcal{F} \cap B_\varepsilon(x^*)$.*

Here and in what follows we denote by

$$B_\varepsilon(x^*) := \{x \in \mathbb{R}^n \colon \|x - x^*\| < \varepsilon\}$$

the open ε-ball at x^* and $\|x\| = \sqrt{x^\top x}$ denotes the Euclidean norm in $\mathbb{R}^n$.

Definition 2.8 (Strict Local Minimizer)**.** *A point $x^* \in \mathbb{R}^n$ is called a* strict local minimizer *of Problem (2.7) if x^* is feasible and if an $\varepsilon > 0$ exists such that $f(x) > f(x^*)$ for all $x \in (\mathcal{F} \cap B_\varepsilon(x^*)) \setminus \{x^*\}$.*

Besides local minimizers we will also consider global minimizers.

Definition 2.9 ((Strict) Global Minimizers)**.** *A point $x^* \in \mathbb{R}^n$ is called a* global minimizer *of Problem (2.7) if x^* is feasible and if $f(x) \geq f(x^*)$ holds for all $x \in \mathcal{F}$. The point is called a* strict global minimizer *if $f(x) > f(x^*)$ holds for all $x \in \mathcal{F} \setminus \{x^*\}$.*

Naturally, the question arises under which assumptions a (global) minimum of a nonlinear optimization problem exists. The answer is given by the very classic theorem of Weierstraß.

Theorem 2.10 (Theorem of Weierstraß). *Suppose that the set $\mathcal{F}$ is nonempty and compact and that the function $f : \mathcal{F} \to \mathbb{R}$ is continuous. Then, f has at least one global minimizer and at least one global maximizer.*

Our goal now is to state the first-order optimality conditions of Problem (2.7), i.e., the Karush–Kuhn–Tucker (KKT) conditions. To this end, we need some more notation.

Definition 2.11 (Active Inequality Constraints). *Let $x \in \mathcal{F}$ be a feasible point of Problem (2.7). Then, the set*

$$I(x) := \{i \in I : g_i(x) = 0\}$$

is called the set of active inequality constraints *at the point x.*

Definition 2.12 (Abadie Constraint Qualification). *We say that a feasible point $x \in \mathcal{F}$ of Problem (2.7) satisfies the* Abadie constraint qualification (ACQ) *if $T_X(x) = T_{lin}(x)$ holds.*

In the last definition, we used the two cones $T_X(x)$, i.e., the tangential cone of X at x, and $T_{\text{lin}}(x)$, i.e., the linearized tangential cone of X at x. We will not discuss the details here. They can be found in every textbook on nonlinear optimization and are also a core topic of my lecture on "Nonlinear Optimization".

Definition 2.13 (Lagrangian Function). *The function*

$$\mathcal{L}(x, \lambda, \mu) := f(x) - \sum_{i=1}^{m} \lambda_i g_i(x) - \sum_{j=1}^{p} \mu_j h_j(x)$$

is called Lagrangian function *of Problem (2.7).*

Using the Lagrangian function we can now define the Karush–Kuhn–Tucker (KKT) conditions.

Definition 2.14 (KKT Conditions, KKT Point, Lagrangian Multipliers). *We consider Problem (2.7) with continuously differentiable functions f, g, and h.*[1]

(a) The conditions

$$\nabla_x \mathcal{L}(x, \lambda, \mu) = 0, \tag{2.8a}$$

$$h(x) = 0, \tag{2.8b}$$

$$\lambda \geq 0, \; g(x) \geq 0, \; \lambda^\top g(x) = 0 \tag{2.8c}$$

[1] A quantity without the index usually stands for the vector; e.g., $h(x) = (h_j(x))_{j=1,\dots,p}$.

are called Karush–Kuhn–Tucker *(or* KKT*)* conditions *of Problem* (2.7)*. Here and in what follows,*

$$\nabla_x \mathcal{L}(x, \lambda, \mu) = \nabla f(x) - \sum_{i=1}^{m} \lambda_i \nabla g_i(x) - \sum_{j=1}^{p} \mu_j \nabla h_j(x)$$

is the gradient of the Lagrangian function with respect to the variables x.

(b) Every vector $((x^*)^\top, (\lambda^*)^\top, (\mu^*)^\top)^\top \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p$ that satisfies the KKT conditions is called a KKT point *of Problem* (2.7)*. The components of λ^* and μ^* are called* Lagrangian multipliers*.*

With these definitions, we can now state the famous KKT theorem under the ACQ.

Theorem 2.15 (KKT Theorem under the Abadie CQ)**.** *Let $x^* \in \mathbb{R}^n$ be a local minimizer of Problem* (2.7)*. Moreover, suppose that the Abadie CQ holds at x^*. Then, there exist Lagrangian multipliers $\lambda^* \in \mathbb{R}^m$ and $\mu^* \in \mathbb{R}^p$ so that $((x^*)^\top, (\lambda^*)^\top, (\mu^*)^\top)^\top$ is a KKT point of Problem* (2.7)*.*[2]

The KKT theorem also holds under other constraint qualifications that are stronger than the ACQ, where "stronger" means that the other constraint qualification implies the ACQ. One very prominent example is the LICQ.

Definition 2.16 (Linear Independence Constraint Qualification)**.** *Let $x \in \mathbb{R}^n$ be a feasible point of Problem* (2.7) *and let $I(x)$ be the set of active inequality constraints at x. We say that the* linear independence constraint qualification (LICQ)*is satisfied in x if the gradients*

$$\begin{aligned} &\nabla g_i(x) \quad \textit{for all} \quad i \in I(x), \\ &\nabla h_j(x) \quad \textit{for all} \quad j = 1, \ldots, p \end{aligned}$$

are linearly independent.

Thus, the following theorem holds because the LICQ implies the ACQ.

Theorem 2.17 (KKT Theorem under the LICQ)**.** *Let $x^* \in \mathbb{R}^n$ be a local minimizer of Problem* (2.7) *that satisfies the LICQ. Then, there exist Lagrangian multipliers $\lambda^* \in \mathbb{R}^m$ and $\mu^* \in \mathbb{R}^p$ so that (x^*, λ^*, μ^*) is a KKT point of Problem* (2.7)*.*

[2]For the ease of presentation, we omit the many transpositions in $((x^*)^\top, (\lambda^*)^\top, (\mu^*)^\top)^\top$ from now on and simply write (x^*, λ^*, μ^*) instead.

2.3 Convex Optimization

We now consider convex optimization problems.The most important property of convex optimization problems is that local and global optima coincide.

Theorem 2.18. *Let $f : \mathcal{F} \to \mathbb{R}$ be a convex function defined on a convex set $\mathcal{F} \subseteq \mathbb{R}^n$. Then, the following statements are true.*

(a) Every local minimizer of f on $\mathcal{F}$ is also a global minimizer of f on $\mathcal{F}$.

(b) If f is strictly convex, then f has at most one local minimizer on $\mathcal{F}$ and this local minimizer (if it exists) then also is the unique strict global minimizer of f on $\mathcal{F}$.

(c) Let $\mathcal{F}$ be open, f be continuously differentiable on $\mathcal{F}$, and suppose that $x^* \in \mathcal{F}$ is a stationary point of f.[3] Then, x^* is a global minimizer of f on $\mathcal{F}$.

Exercise 2.4. Prove Theorem 2.18.

Exercise 2.5. Regarding claim (b) of Theorem 2.18: Can you construct an optimization problem with f being strictly convex and $\mathcal{F}$ being convex for which no minimizer exists.

In the remainder of this section, we now study the meaning of the KKT conditions for convex optimization problems. To this end, we consider the problem

$$\min_{x \in \mathbb{R}^n} \quad f(x) \tag{2.9a}$$

$$\text{s.t.} \quad g_i(x) \geq 0, \quad i = 1, \dots, m, \tag{2.9b}$$

$$b_j^\top x = \beta_j, \quad j = 1, \dots, p, \tag{2.9c}$$

where $f : \mathbb{R}^n \to \mathbb{R}$ and $g_i : \mathbb{R}^n \to \mathbb{R}$, $i = 1, \dots, m$, are continuously differentiable, $b_j \in \mathbb{R}^n$, $j = 1, \dots, p$, are vectors and $\beta_j \in \mathbb{R}$, $j = 1, \dots, p$, are scalars. Moreover, f is supposed to be convex and the g_i, $i = 1, \dots, m$, are supposed to be concave. Thus, we consider a convex objective function over a convex feasible set.

Definition 2.19 (Slater's Constraint Qualification)**.** *We say that the convex problem (2.9) satisfies the* constraint qualification of Slater *if there exists a vector $\hat{x} \in \mathbb{R}^n$ so that*

$$\begin{aligned} g_i(\hat{x}) > 0 \quad & \textit{for all} \quad i = 1, \dots, m, \\ b_j^\top \hat{x} = \beta_j \quad & \textit{for all} \quad j = 1, \dots, p \end{aligned}$$

holds. This means that $\hat{x}$ is strictly feasible w.r.t. the inequality constraints and feasible w.r.t. the equality constraints.

[3] A point x is called a *stationary point* of f if $\nabla f(x) = 0$ holds.

Note that a convex problem that satisfies Slater's CQ possess a non-empty interior of the feasible set defined by the inequality constraints.

Theorem 2.20 (KKT Theorem for Convex Problems under Slater's CQ)**.** *Let $x^* \in \mathbb{R}^n$ be a local (and thus global) minimizer of the convex problem (2.9). Moreover, suppose that Slater's CQ is satisfied. Then, there exist Lagrangian multipliers $\lambda^* \in \mathbb{R}^m$ and $\mu^* \in \mathbb{R}^p$ so that (x^*, λ^*, μ^*) satisfies the KKT conditions*

$$\begin{aligned} \nabla f(x^*) - \sum_{i=1}^{m} \lambda_i^* \nabla g_i(x^*) - \sum_{j=1}^{p} \mu_j^* b_j = 0&, \\ b_j^\top x^* = \beta_j&, \quad j = 1, \dots, p, \\ g_i(x^*) \geq 0&, \quad i = 1, \dots, m, \\ \lambda_i^* g_i(x^*) = 0&, \quad i = 1, \dots, m, \\ \lambda_i^* \geq 0&, \quad i = 1, \dots, m, \end{aligned}$$

of Problem (2.9).

Up to now, we have shown that the KKT conditions are also necessary first-order optimality conditions for convex problems under Slater's CQ. For general nonlinear problems, the KKT conditions (under a reasonable CQ) are not sufficient conditions. However, for convex problems, the KKT conditions are also sufficient conditions.

Theorem 2.21. *Let $(x^*, \lambda^*, \mu^*) \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p$ be a KKT point of the convex problem (2.9). Then, x^* is a local (and thus global) minimizer of Problem (2.9).*

Exercise 2.6. Consider the quadratic problem

$$\begin{aligned} \min_{x \in \mathbb{R}^n} \quad & \frac{1}{2} x^\top Q x + c^\top x \\ \text{s.t.} \quad & Ax \geq b, \\ & Cx = d \end{aligned} \tag{2.10}$$

with $c \in \mathbb{R}^n$, $Q \in \mathbb{R}^{n \times n}$ being symmetric and positive semi-definite, $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$, $C \in \mathbb{R}^{\ell \times n}$, and $d \in \mathbb{R}^\ell$.

(i) Is Problem (2.10) a convex optimization problem?

(ii) Derive the KKT conditions of Problem (2.10).

Hint: You may use the following theorem without proof.

> **Theorem 2.22.** *Let $f : \mathcal{F} \to \mathbb{R}$ be twice continuously differentiable on the open set $\mathcal{F} \subseteq \mathbb{R}^n$. Then, the function f is convex if and only if the Hessian matrix $\nabla^2 f(x)$ is positive semi-definite for all $x \in \mathcal{F}$.*

2.4 Mathematical Programs with Complementarity Constraints

We now consider problems of the form

$$\begin{aligned}
\min_{x \in \mathbb{R}^n} \quad & f(x) && \text{(2.11a)}\\
\text{s.t.} \quad & g_i(x) \geq 0, \quad i \in I = \{1, \dots, m\}, && \text{(2.11b)}\\
& h_j(x) = 0, \quad j \in J = \{1, \dots, p\}, && \text{(2.11c)}\\
& \varphi_\ell(x) \geq 0, \quad \ell \in \{1, \dots, r\}, && \text{(2.11d)}\\
& \psi_\ell(x) \geq 0, \quad \ell \in \{1, \dots, r\}, && \text{(2.11e)}\\
& \varphi_\ell(x)\psi_\ell(x) = 0, \quad \ell \in \{1, \dots, r\}. && \text{(2.11f)}
\end{aligned}$$

This problem is called a *mathematical program with complementarity constraints (MPCC)* and it can be seen as the nonlinear optimization problem (2.7), extended by the two functions $\varphi, \psi : \mathbb{R}^n \to \mathbb{R}^r$ and the three last sets of constraints (2.11d)–(2.11f). We also assume that the functions φ, ψ are continuously differentiable so that Problem (2.11) looks like a usual NLP. However, the three last sets of constraints add a significant difficulty to the problem.

The reason is that the LICQ does not hold at any feasible point of (2.11).

Theorem 2.23. *Let x be feasible for Problem (2.11). Then, the LICQ does not hold at x.*

Exercise 2.7. Prove Theorem 2.23.

Moreover, the same can be shown for the Abadie CQ, which was the weakest CQ under which we have stated the KKT theorem.

Theorem 2.24. *Let x be feasible for Problem (2.11). Then, the ACQ does not hold at x.*

Exercise 2.8. Prove Theorem 2.24.

Example 2.25. *We already considered a setting in which the feasible set has the structure of the feasible set of Problem (2.11) in Corollary 2.6. The set of inequalities and equations in this corollary reads*

$$\begin{aligned}
Ax &= b,\\
x &\geq 0,\\
c - A^\top \lambda &\geq 0,\\
x_i(c - A^\top \lambda)_i &= 0, \quad i \in \{1, \dots, n\}.
\end{aligned}$$

With $h(x) = Ax - b$, $\varphi(x) = x$, and $\psi(x, \lambda) = c - A^\top \lambda$, this exactly matches the setting in Problem (2.11). We will come back to this situation in Chapter 4.

Example 2.26. *The KKT conditions* (2.8) *in Definition 2.14 also fit into the framework of the MPCC* (2.11)*.*

The interested reader is referred to the textbook by Luo et al. (1996) for more details on this class of problems.

3

Solution Concepts: Pessimistic vs. Optimistic Problems

Example 3.1. *Imagine Alice (university professor in psychology) and Bob (car mechanic), having a perfect marriage until now, i.e., they are both happy with their marriage. Thus, they both want to achieve the best for themselves and for the other one. Of course, one of them wears the pants and this is Alice*[1]*; see Figure 3.1. In bilevel notation, Alice is the leader and Bob is the follower. Let's further model reality: Alice is managing the total budget that they both can spent. Moreover, let us assume that the objective function of Alice is to maximize her own utility, which is based (i) on the budget she can spend and (ii) on how much Bob leaves her alone. She is, however, limited on what of the entire budget she can spent on her own since there is a minimum budget—at that moment, we as the modeler, still believe in this marriage—that she has to leave to Bob. So let us suppose that Alice makes her decision, meaning that Bob receives €120 for next Saturday. Bob has two opportunities. Either he can solely go to a pub and drink as many beers and shots that he can take within his budget of €120 or he can invite some friends, buy some beer, and order some Pizza for all of them. We further assume that both possibilities lead to the same objective function value for Bob.*

(i) *Let's assume Bob still adores Alice. Since Bob knows the objective of Alice, he takes his budget, goes solely to a pub, and drinks as much as he can as long as he stays within his budget. He arrives back home horribly drunk and sleeps in the guest room. By doing all this, Bob is maximizing his own utility*[2] *and, at the same time, ensures that Alice obtains her best-possible utility as well.*

(ii) *Let's assume Bob finally managed to notice that Alice—although still*

[1] Our assumptions are getting closer to reality.

[2] We are not going to question this further.

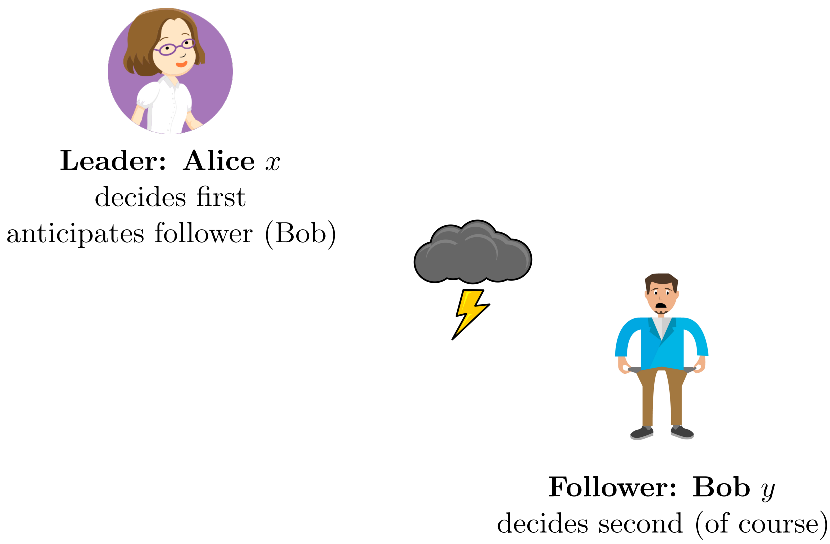

Figure 3.1: The blessing and the curse of marriage

> *being married with him—enjoys a serious flirtation with her new PhD student at the university. Working on a diabolic plan and thus still staying with Alice, Bob again receives his €120. Now, he acts differently. He invites all of his friends, ignores Pizza for being able to spend more money on beer, and tries to be as loud as he/they can.*
>
> *His own utility function is the same as in the last case. He could have drunk more without his friends but €120 spent on beer for him and his friends means the same for him. But(!): Alice is really pissed now because she cannot relax as much as in the other solution of Bob.*

We see, Bob—being the follower—might be more powerful as we thought at the beginning.

Let's put this into math. In Section 1.2, we defined the bilevel problem as

$$\min_{x \in X, y} \quad F(x, y) \tag{3.1a}$$
$$\text{s.t.} \quad G(x, y) \geq 0, \tag{3.1b}$$
$$y \in S(x), \tag{3.1c}$$

where $S(x)$ is the set of optimal solutions of the x-parameterized lower-level problem, which we define as

$$\min_{y \in Y} \quad f(x, y) \tag{3.2a}$$
$$\text{s.t.} \quad g(x, y) \geq 0. \tag{3.2b}$$

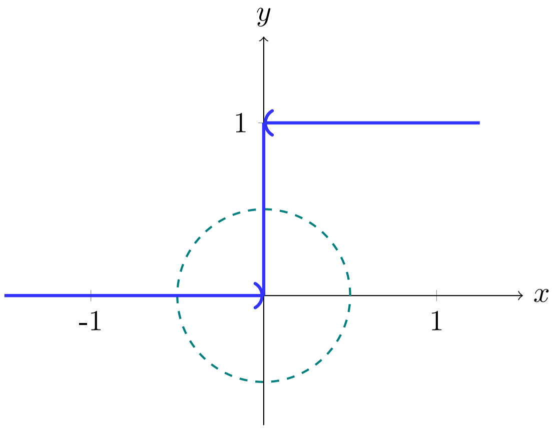

Figure 3.2: The optimal solution y of the lower-level problem in dependence of the upper-level decision x in Example 3.2

One might think that this definition is the same as the one that we obtain if we replace the upper level (3.1) with

$$\min_{x \in X} \quad F(x, y) \tag{3.3a}$$
$$\text{s.t.} \quad G(x, y) \geq 0, \tag{3.3b}$$
$$y \in S(x). \tag{3.3c}$$

Indeed, one really needs to take a closer look to see the difference: We omitted the y below the "min" of the upper level and this really makes a difference. But why?

Example 3.2 (See Example 1.2 in Dempe, Kalashnikov, et al. (2015))**.** *Consider the bilevel problem*

$$\min_{x} \quad F(x, y) = x^2 + y \quad \textit{s.t.} \quad y \in S(x)$$

with

$$S(x) = \arg\min_{y} \{-xy \colon 0 \leq y \leq 1\}.$$

The best response of the follower is illustrated in Figure 3.2. Formally, it is given by

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

This means that the mapping $x \mapsto F(x, S(x))$ looks like it is illustrated in Figure 3.3. This is not a function and its minimum is unclear since it depends on the response $y \in S(x)$ of the follower if the leader chooses $x = 0$. For

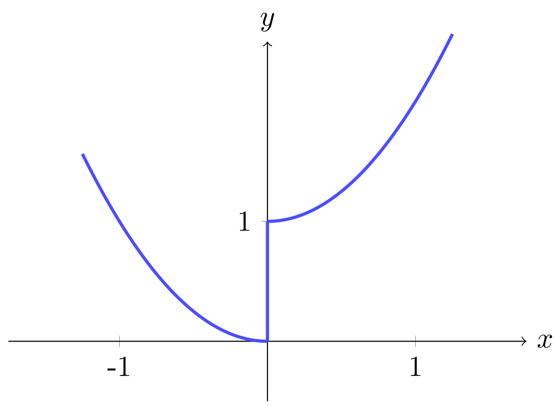

Figure 3.3: Graph of the mapping $x \mapsto F(x, S(x))$ in Example 3.2

the follower, all responses $y \in S(0) = [0, 1]$ are optimal so that the optimal lower-level solution is not unique. If the follower chooses $y = 0$, the optimal leader's decision is $x = 0$, leading to an objective function value of the leader of 0. However, if the follower chooses $y = 1$, the objective function value of the leader is 1, which is worse than 0 from the point of view of the leader.

This means the following. If the follower does not possess a unique solution, i.e., the set $S(x)$ is not a singleton, then the follower can be "leader-friendly" (which corresponds to $y = 0$ in our example) or the follower can choose a different solution, e.g., $y = 1$, which is worse for the leader (obtaining an objective function value of 1 in this case).

This especially means that Problem (3.3) is ill-posed. To resolve this issue, one usually distinguishes between two different solution concepts in bilevel optimization: the *optimistic solution* and the *pessimistic solution*.

Definition 3.3 (Optimistic bilevel problem)**.** *Problem* (3.1) *with the lower-level problem* (3.2) *is called the* optimistic bilevel problem.

In this case, the leader controls those y that are part of the rational reaction set $S(x)$ of the follower. This is indicated by having the y also below the "min" of the objective function of the leader.

Besides this optimistic variant of the problem there also exists the so-called pessimistic variant, which again is known in at least two versions depending on whether the bilevel problem has coupling constraints or not.

Definition 3.4 (Pessimistic bilevel problem without coupling constraints)**.** *The Problem*

$$\min_{x \in X} \max_{y \in S(x)} \quad F(x, y) \tag{3.4a}$$

$$\text{s.t.} \quad G(x) \geq 0, \tag{3.4b}$$

where $S(x)$ is the set of optimal solutions of the x-parameterized lower-level problem, which is defined as in (3.2), is called the pessimistic bilevel problem.

Definition 3.5 (Pessimistic bilevel problem with coupling constraints)**.** *The Problem*

$$\min_{x \in X} \quad F(x) \tag{3.5a}$$
$$\text{s.t.} \quad G(x, y) \geq 0 \quad \text{for all } y \in S(x), \tag{3.5b}$$

where $S(x)$ is the set of optimal solutions of the x-parameterized lower-level problem, which is defined as in (3.2), is called the pessimistic bilevel problem.

Exercise 3.1. Prove that the assumption that the leader's objective function does not depend on the follower's variables in Problem (3.5) is without loss of generality.

Note that the decision on the solution concept (optimistic vs. pessimistic) is very important and can even change whether a solution exists or not. For instance, the optimal solution $(x, y) = (0, 0)$ with objective function value 0 is attained in Example 3.2 if one considers the optimistic bilevel problem. However, for all other choices of $y \in S(0) = [0, 1]$, the bilevel problem is not solvable since the infimum 0 of the upper-level's objective function is not attained anymore. This, in particular, also applies to the pessimistic bilevel problem in this example.

Remark 3.6. *If the lower-level solution is unique for all $x \in \Omega_x$, both the pessimistic and the optimistic variants of the bilevel problem coincide.*

Besides the concepts of pessimistic and optimistic bilevel problems, local and global solutions can be defined as it is the case for general optimization problems as well; see Section 2.2 for the definitions for standard single-level optimization problems.

For their definition in the bilevel context we first define the graph of the solution set mapping $S(\cdot)$.

Definition 3.7 (Graph of the solution set mapping)**.** *The set*

$$\operatorname{gph} S := \{(x, y) \colon y \in S(x)\}$$

is called the graph of the solution set mapping $S(\cdot)$.

Definition 3.8 (Local and global optimal solution)**.** *A feasible point (x^*, y^*) of the bilevel problem (1.1) is a* local optimal solution *if there exists an $\varepsilon > 0$ such that*

$$F(x, y) \geq F(x^*, y^*)$$

holds for all $(x, y) \in \operatorname{gph} S \cap \Omega$ with

$$\|(x, y) - (x^*, y^*)\| < \varepsilon.$$

A local optimal solution is called a global optimal solution *if $\varepsilon > 0$ can be chosen arbitrarily large.*

Exercise 3.2. Consider the linear bilevel problem

$$
\begin{aligned}
\min_{x} \quad & x + 2y \\
\text{s.t.} \quad & x \geq 0, \\
& y \in \arg\min_{\bar{y}} \{x \colon 3x + 2\bar{y} \leq 20,\, x + 2\bar{y} \leq 12,\, x - 2\bar{y} \leq 4,\, 0 \leq \bar{y} \leq 5\}.
\end{aligned}
\tag{3.6}
$$

(i) Plot the linear inequalities in a coordinate system.

(ii) Determine the bilevel feasible set of Problem (3.6).

(iii) Determine the optimal solution of Problem (3.6) considering the optimistic bilevel problem.

(iv) Suppose that the leader sticks with the solution found in (iii). What would be the optimal response of the pessimistic follower?

(v) Determine the optimal solution of Problem (3.6) considering the pessimistic bilevel problem.

Exercise 3.3. Consider the LP-LP bilevel problem

$$
\begin{aligned}
\min_{x,y} \quad & c_x^\top x + c_y^\top y \\
\text{s.t.} \quad & Ax + By \geq a, \\
& y \in \arg\min_{\bar{y}} \left\{d^\top \bar{y} \colon Cx + D\bar{y} \geq b\right\}
\end{aligned}
\tag{3.7}
$$

with $x, c_x \in \mathbb{R}^{n_x}$, $y, c_y, d \in \mathbb{R}^{n_y}$, $A \in \mathbb{R}^{m \times n_x}$, $B \in \mathbb{R}^{m \times n_y}$ and $a \in \mathbb{R}^m$ as well as $C \in \mathbb{R}^{\ell \times n_x}$, $D \in \mathbb{R}^{\ell \times n_y}$, and $b \in \mathbb{R}^\ell$. Use the value-function reformulation corresponding to Problem (3.7) to show that for the optimistic version of the problem, we can assume without loss of generality that all upper-level variables are linking variables; see Bolusani and Ralphs (2020).

This is the formulation used on the exercise sheet. One could, however, be more general as long as the lower-level objective does not depend on the non-linking variables

maybe add existence theorems for bilevel problems?!

4

Single-Level Reformulations

Starting with the seminal paper by Fortuny-Amat and McCarl (1981), the solution method for bilevel problems used most frequently in practice is to reformulate the bilevel model as an "ordinary", i.e., single-level, problem. This single-level reformulation can then be solved[1] with state-of-the-art general-purpose solvers for the resulting classes of problems.

In this chapter, we will consider three different single-level reformulations:

(a) one being based on the optimal value function of the lower-level problem,

(b) one using the KKT conditions of the lower-level problem,

(c) and one based on a strong-duality theorem for the lower-level problem.

While the first one can be applied for every bilevel problem, the two latter ones require that the lower-level problem possesses some compact optimality certificate, which is not only necessary but also sufficient. We will take care about the details in the following sections.

4.1 A Single-Level Reformulation using the Optimal Value Function

Let us start again with the general optimistic bilevel problem

$$\min_{x \in X, y} \quad F(x, y) \tag{4.1a}$$

$$\text{s.t.} \quad G(x, y) \geq 0, \tag{4.1b}$$

$$y \in S(x), \tag{4.1c}$$

[1]At least after some further transformations. We will discuss the further pitfalls later on.

where $S(x)$ is the set of optimal solutions of the x-parameterized problem

$$\min_{y \in Y} \quad f(x,y) \tag{4.2a}$$

$$\text{s.t.} \quad g(x,y) \geq 0. \tag{4.2b}$$

We already considered this problem in Definition 1.6. By using the optimal value function

$$\varphi(x) := \min_{y \in Y} \{f(x,y) \colon g(x,y) \geq 0\},$$

we can equivalently re-write the problem as

$$\min_{x \in X, y \in Y} \quad F(x,y) \tag{4.3a}$$

$$\text{s.t.} \quad G(x,y) \geq 0, \ g(x,y) \geq 0, \tag{4.3b}$$

$$f(x,y) \leq \varphi(x). \tag{4.3c}$$

This looks like a usual single-level problem. However, the problem is the optimal value function $\varphi : \mathbb{R}^{n_x} \to \mathbb{R}$. Of course, we can evaluate this function but this evaluation corresponds to solving the lower-level problem (4.2) for the given x, which is the argument of this function. Thus, the evaluation is rather expensive. Moreover, in almost all cases, the optimal value function is not known in algebraic, i.e., in closed, form. Finally, it is usually nonsmooth (even under strong assumptions), which can be easily seen, e.g., in Example 1.12.

Although this all sounds as if the single-level reformulation using the optimal value function is not useful, it can indeed be very useful for problems that lead to a special structure of the optimal value function. We will come back to this aspect later on.

4.2 The KKT Reformulation for LP-LP Bilevel Problems

The most classic approach to obtain a single-level reformulation is to exploit optimality conditions for the lower-level problem. Since these optimality conditions need to be necessary and sufficient, the application of these conditions is usually only possible for convex lower-level problems that satisfy a reasonable constraint qualification—which typically is Slater's constraint qualification in the convex case.

To avoid over-complicating the presentation, we first present the classic KKT reformulation using the example of an LP-LP bilevel problem of the form

$$\min_{x,y} \quad c_x^\top x + c_y^\top y \tag{4.4a}$$

$$\text{s.t.} \quad Ax + By \geq a, \tag{4.4b}$$

$$y \in \arg\min_{\bar{y}} \left\{ d^\top \bar{y} \colon Cx + D\bar{y} \geq b \right\} \tag{4.4c}$$

with $c_x \in \mathbb{R}^{n_x}$, $c_y, d \in \mathbb{R}^{n_y}$, $A \in \mathbb{R}^{m \times n_x}$, $B \in \mathbb{R}^{m \times n_y}$, and $a \in \mathbb{R}^m$ as well as $C \in \mathbb{R}^{\ell \times n_x}$, $D \in \mathbb{R}^{\ell \times n_y}$, and $b \in \mathbb{R}^\ell$. Note that we already omitted a linear term depending on the upper-level variables x in the lower-level objective function since this term would not have any influence on the optimal solutions of the lower level as it is constant from the point of view of the follower.

The lower-level problem in (4.4c) can be seen as the x-parameterized linear problem

$$\min_y \quad d^\top y \quad \text{s.t.} \quad Dy \geq b - Cx. \tag{4.5}$$

Its Lagrangian function is given by

$$\mathcal{L}(y, \lambda) = d^\top y - \lambda^\top (Cx + Dy - b)$$

and the KKT conditions are given by dual feasibility

$$D^\top \lambda = d, \quad \lambda \geq 0,$$

primal feasibility

$$Cx + Dy \geq b,$$

and the KKT complementarity conditions

$$\lambda_i (C_{i\cdot} x + D_{i\cdot} y - b_i) = 0 \quad \text{for all } i = 1, \ldots, \ell.$$

As before, $C_{i\cdot}$ denotes the ith row of C. Since the lower-level feasible region is polyhedral, the Abadie constraint qualification holds and the KKT conditions are both necessary and sufficient. Thus, the LP-LP bilevel problem (4.4) can be reformulated as

$$\begin{aligned}
\min_{x,y,\lambda} \quad & c_x^\top x + c_y^\top y && \text{(4.6a)}\\
\text{s.t.} \quad & Ax + By \geq a, \ Cx + Dy \geq b, && \text{(4.6b)}\\
& D^\top \lambda = d, \ \lambda \geq 0, && \text{(4.6c)}\\
& \lambda_i (C_{i\cdot} x + D_{i\cdot} y - b_i) = 0 \quad \text{for all } i = 1, \ldots, \ell. && \text{(4.6d)}
\end{aligned}$$

Note that we now optimize over an extended space of variables since we additionally have to include the lower-level dual variables λ. By optimizing over x, y, and λ simultaneously in Problem (4.6), any global solution of (4.6) corresponds to an optimistic bilevel solution.

Problem (4.6) is linear except for the KKT complementarity conditions in (4.6d) that turn the problem into a nonconvex and nonlinear optimization problem (NLP). More precisely, Problem (4.6) is a mathematical program with complementarity constraints (MPCC); see, e.g., Luo et al. (1996) and Section 2.4. Thus, and unfortunately, standard NLP algorithms usually cannot be applied for such problems since classic constraint qualifications like the Mangasarian–Fromowitz or the linear independence constraint qualification are violated at every feasible point; see, e.g., Ye and Zhu (1995) or Section 2.4 again.

Exercise 4.1. Derive the KKT reformulation of the LP-QP bilevel problem that is given by

$$\begin{aligned} \min_{x,y} \quad & c_x^\top x + c_y^\top y \\ \text{s.t.} \quad & Ax + By \geq a, \\ & y \in \arg\min_{\bar{y}} \left\{ \frac{1}{2} y^\top Q y + d^\top y \colon Cx + Dy \geq b \right\} \end{aligned} \tag{4.7}$$

with x, $c_x \in \mathbb{R}^{n_x}$, y, c_y, $d \in \mathbb{R}^{n_y}$, $A \in \mathbb{R}^{m \times n_x}$, $B \in \mathbb{R}^{m \times n_y}$, $a \in \mathbb{R}^m$, $C \in \mathbb{R}^{\ell \times n_x}$, $D \in \mathbb{R}^{\ell \times n_y}$, $b \in \mathbb{R}^\ell$, and $Q \in \mathbb{R}^{n_y \times n_y}$ being symmetric and positive semi-definite.

4.3 The KKT Reformulation for Parametric Convex Lower-Level Problems

We have now stated the KKT reformulation (4.6) of the LP-LP bilevel problem (4.4) but we have not yet discussed the relationship between their solutions if, in a more general setting, a parametric convex lower-level problem is considered. This topic is a bit more delicate as one might think at a first glance since it turns out that the equivalence of these problems depends on whether global or local solutions are considered and on the satisfaction of constraint qualifications.

To shed some more light on these aspects we consider the bilevel problem

$$\min_{x \in X, y} \quad F(x,y) \tag{4.8a}$$
$$\text{s.t.} \quad y \in S(x), \tag{4.8b}$$

where $S(x)$ is the set of optimal solutions of the x-parameterized convex problem

$$\min_{y \in Y} \quad f(x,y) \tag{4.9a}$$
$$\text{s.t.} \quad g(x,y) \geq 0. \tag{4.9b}$$

Thus, we assume that $y \mapsto f(x,y)$ is a convex function and $y \mapsto g(x,y)$ is a concave function for all $x \in X$, i.e., for all feasible leader's decisions. Moreover, we assume that Y is a convex set that, e.g., contains simple bound constraints on the variables of the follower. This means that the lower-level problem is indeed an x-parametric convex problem. Please note further at this point that we simplified the upper-level problem a bit since no coupling constraints are present anymore.

In what follows, we also need Slater's constraint qualification for the lower level. For the ease of presentation, we now assume for what follows that all constraints g_i, $i = 1, \ldots, \ell$, are nonlinear.

Definition 4.1 (Slater's constraint qualification for the lower level)**.** *For a given upper-level feasible point $x \in X$ of the bilevel problem (4.8) we say that Slater's constraint qualification holds for the lower-level problem (4.9) if there exists a point $\hat{y}(x)$ with $g_i(x, \hat{y}(x)) > 0$ for all $i = 1, \ldots, \ell$.*

Under Slater's constraint qualification (for all possible x decided on by the leader) we then know that we can re-write the bilevel problem using the KKT conditions of the lower-level problem, i.e., we obtain the single-level reformulation

$$\min_{x,y,\lambda} \quad F(x,y) \tag{4.10a}$$
$$\text{s.t.} \quad x \in X, \tag{4.10b}$$
$$\nabla_y \mathcal{L}(x,y,\lambda) = 0, \tag{4.10c}$$
$$g(x,y) \geq 0, \tag{4.10d}$$
$$\lambda \geq 0, \tag{4.10e}$$
$$\lambda^\top g(x,y) = 0. \tag{4.10f}$$

Here,

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

is the gradient of the lower level's Lagrangian function w.r.t. y.

Let us now first shed some light on the relation between the global solutions of the bilevel problem and the global solutions of its KKT reformulation.

Theorem 4.2 (See Theorem 2.1 in Dempe and Dutta (2012))**.** *Let (x^*, y^*) be a global optimal solution of the bilevel problem (4.8) and assume that the lower-level problem is a convex optimization problem that satisfies Slater's constraint qualification for x^*. Then, the point (x^*, y^*, λ^*) is a global optimal solution of the single-level reformulation (4.10) for every*

$$\lambda^* \in \Lambda(x^*, y^*) := \left\{ \lambda \geq 0 \colon \nabla_y \mathcal{L}(x^*, y^*, \lambda) = 0, \ \lambda^\top g(x^*, y^*) = 0 \right\}.$$

Proof. Since the x^*-parameterized lower-level problem is convex and since this parametric convex problem satisfies Slater's constraint qualification for the given x^*, the KKT theorem for convex problems (Theorem 2.20) implies that $\lambda^* \in \Lambda(x^*, y^*)$ holds if and only if $(x^*, y^*) \in \operatorname{gph} S$. $\square$

The opposite direction is also true under some assumptions.

Theorem 4.3 (See Theorem 2.3 in Dempe and Dutta (2012))**.** *Let (x^*, y^*, λ^*) be a global optimal solution of Problem (4.10) and let the lower-level problem (4.9) be convex. Moreover, suppose that Slater's constraint qualification is satisfied for the lower-level problem for every $x \in X$. Then, (x^*, y^*) is a global optimal solution of the bilevel problem (4.8).*

Proof. Suppose that (x^*, y^*, λ^*) is a global optimal solution of Problem (4.10). Thus, $\Lambda(x^*, y^*) \neq \emptyset$ holds. Since the objective function F of (4.10) does not depend on $\lambda \in \Lambda(x^*, y^*)$, each point (x^*, y^*, λ) with $\lambda \in \Lambda(x^*, y^*)$ is a global optimal solution as well.

Assume now that (x^*, y^*) is not a global optimal solution of the bilevel problem (4.8). Then, there exists a point (x, y) with $x \in X$ and $y \in S(x)$ such that

$$F(x, y) < F(x^*, y^*)$$

holds. Since $y \in S(x)$ and Slater's constraint qualification holds at x, the respective KKT conditions are valid and thus there exists a vector $\lambda \in \mathbb{R}^\ell$ of Lagrangian multipliers such that

$$\begin{aligned} \nabla_y f(x, y) - \sum_{i=1}^{\ell} \lambda_i \nabla_y g_i(x, y) &= 0, \\ \lambda^\top g(x, y) &= 0, \\ \lambda &\geq 0, \\ g(x, y) &\geq 0 \end{aligned}$$

holds. Consequently, (x, y, λ) is a feasible point for the KKT reformulation (4.10) that has a better objective function value as (x^*, y^*, λ^*). This is a contradiction to the global optimality of (x^*, y^*, λ^*) and the claim follows. □

The last two theorems tell us that the original bilevel optimization problem and its single-level KKT reformulation are equivalent under the assumption that Slater's constraint qualification holds for all possible x decided on by the leader and if global optimal solutions are considered. The theorems so far do not give any insight on the relationship between the local minima of these two problems.

Before we consider these local optima, let us first study whether the assumptions in the last two theorems regarding Slater's constraint qualification are really necessary.

Example 4.4 (See Example 2.2 in Dempe and Dutta (2012))**.** *Let us consider the x-parameterized convex lower-level problem*

$$\min_{y_1, y_2} \quad y_1 \quad s.t. \quad y_1^2 - y_2 \leq x, \ y_1^2 + y_2 \leq 0; \tag{4.11}$$

see Figure 4.1 for an illustration. If $x = 0$, the only feasible point of this lower-level problem is $y = (y_1, y_2) = (0, 0)$ and, thus, Slater's constraint qualification is violated. If we consider $x \geq 0$ (this will be our upper-level constraint later on), the lower-level's optimal solutions are given by

$$y(x) = \begin{cases} (0, 0), & \text{if } x = 0, \\ \left(-\sqrt{x/2}, -x/2\right), & \text{if } x > 0. \end{cases}$$

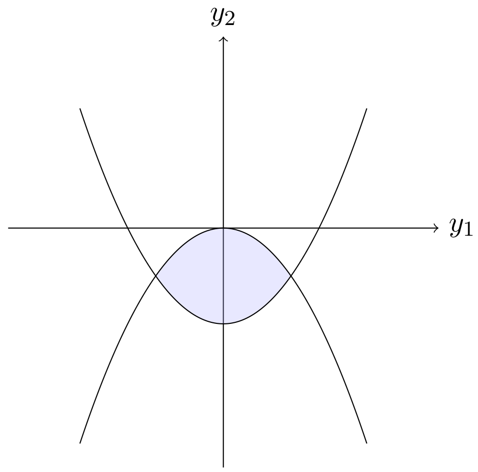

Figure 4.1: The feasible region of the x-parameterized convex lower-level problem (4.11) of Example 4.4 with $x = 1$.

Some further calculations reveal that the corresponding Lagrangian multiplier is given by

$$\lambda_1(x) = \lambda_2(x) = \frac{1}{4\sqrt{x/2}}$$

for $x > 0$. If $x = 0$, the problem does not satisfy Slater's constraint qualification so that the KKT conditions are not satisfied. Hence, no properly defined Lagrangian multipliers exist in this case.

Consider now the bilevel problem

$$\min_{x,y} \quad x \quad s.t. \quad x \geq 0,\ y \in S(x), \tag{4.12}$$

where $S(x)$ is again the solution set mapping of the lower-level problem (4.11). Obviously, the unique global optimal solution of this bilevel problem is $x = 0$, $y = (0, 0)$ with objective function value 0. Moreover, there exist no local optimal solutions.

Lastly, we consider the corresponding MPCC. The Lagrangian of the lower-level problem reads

$$\mathcal{L}(x, y, \lambda) = y_1 - \lambda_1(x - y_1^2 + y_2) - \lambda_2(-y_1^2 - y_2)$$

and its gradient w.r.t. y is given by

$$\nabla_y \mathcal{L}(x, y, \lambda) = \begin{pmatrix} 1 + 2\lambda_1 y_1 + 2\lambda_2 y_1 \\ -\lambda_1 + \lambda_2 \end{pmatrix}.$$

Hence, the MPCC is given by

$$\begin{aligned}\min_{x,y_1,y_2,\lambda_1,\lambda_2} \quad & x \\ s.t. \quad & x \geq 0, \\ & y_1^2 - y_2 \leq x, \ y_1^2 + y_2 \leq 0, \\ & \lambda_1 \geq 0, \ \lambda_2 \geq 0, \\ & \lambda_1(x - y_1^2 + y_2) = 0, \ \lambda_2(-y_1^2 - y_2) = 0, \\ & 1 + 2\lambda_1 y_1 + 2\lambda_2 y_1 = 0, \ -\lambda_1 + \lambda_2 = 0.\end{aligned}$$

The point $(x, y(x), \lambda(x))$ is, by construction, feasible for the MPCC for $x > 0$ and the corresponding objective function value of the bilevel problem converges to 0 for $x \to 0$. However, the problem does not possess an optimal solution since for $x = 0$, the uniquely determined lower-level's solution is $y = (0, 0)$ but no feasible multipliers exist in this case.

The take-home message is the following: A global optimal solution of the bilevel problem does not need to correspond to a global optimal solution of its KKT reformulation if the lower-level problem does not satisfy Slater's constraint qualification for the given upper-level part of the bilevel problem's solution.

We have seen that the assumption of Theorem 4.2 cannot be neglected. Next, we also show that the assumptions of Theorem 4.3 are essential as well.

Example 4.5 (See Example 2.4 in Dempe and Dutta (2012))**.** *We consider the bilevel problem*

$$\min_{x,y} \quad (x-1)^2 + y^2 \quad s.t. \quad x \in \mathbb{R}, \ y \in S(x),$$

where $S(x)$ denotes the solution set mapping of the x-parameterized convex lower-level problem

$$\min_{y} \quad x^2 y \quad s.t. \quad y^2 \leq 0.$$

It is easy to see that $y = 0$ is the only feasible solution and thus the uniquely determined global optimal solution of the lower-level problem (independent of the leader's decision x). In particular, this means that there exists no x for which Slater's constraint qualification holds for the lower-level problem. Since $y = 0$ always is the optimal follower's decision, the uniquely determined global optimal solution of the bilevel problem is $(x, y) = (1, 0)$.

Let us now consider the corresponding KKT reformulation:

$$\begin{aligned}\min_{x,y,\lambda}\quad & (x-1)^2+y^2\\ s.t.\quad & x\in\mathbb{R},\\ & y^2\le 0,\\ & \lambda\ge 0,\\ & \lambda y^2=0,\\ & x^2+2\lambda y=0.\end{aligned}$$

It is easy to see that all feasible solutions of this MPCC are of the form $(0,0,\lambda)$ *with* $\lambda \ge 0$. *Since the objective function does not depend on* λ, *all these points are also global optimal solutions of the MPCC. However, none of them correspond to the optimal solution* $(1,0)$ *of the bilevel problem.*

Now that we have clarified the relationship between the global optimal values of the bilevel problem and its KKT reformulation we will also consider the relationship of the local optimal values. It turns out that local minima of the KKT reformulation do not need to be local optima of the bilevel problem.

Example 4.6. *We start by studying the lower-level problem*

$$\begin{aligned}\min_{y_1,y_2}\quad & y_1^2+(y_2+1)^2\\ s.t.\quad & (y_1-x_1)^2+(y_2-1-x_1)^2\le 1,\\ & (y_1+x_2)^2+(y_2-1-x_2)^2\le 1.\end{aligned}$$

As usual, let $S(x)$ *be its solution set. The problem is illustrated in Figure 4.2. The feasible set is the intersection of the* x*-parameterized discs and the centers of the discs are illustrated using the blue dotted lines. The right arc belongs to* x_1 *and the left arc belongs to* x_2. *The solution of the lower-level problem is then the point in the intersection of these discs which is closest to the point* $(0,-1)$. *The upper-level problem is given by*

$$\min_{x,y_1,y_2}\quad -y_2 \quad s.t.\quad y_1y_2=0,\ x\ge 0,\ y\in S(x).$$

All points $y \in S(x)$ *with* $x \ge 0$ *and* $y_2 \ge 0$ *have a non-positive objective function value and the points with a strictly negative objective function value are those with* $y_2 > 0$, *which then have to satisfy* $y_1 = 0$ *due to upper-level feasibility. Let us consider now the point* $x^* = (0,0)$ *and* $y^* = (0,0)$. *The point* (x^*, y^*) *is feasible for the bilevel problem. However, it is not a local minimum. To show this, we consider the points* $(0, y_2)$ *with* $y_2 > 0$ *that have a smaller objective function value. If* $(0, y_2)$ *with* $y_2 > 0$ *is feasible for the bilevel problem, this implies that* $x_1 = x_2$. *Otherwise the point in the intersection of the discs would be left or right from the* y_2*-axis and, thus,* $y_1 \ne 0$ *would*

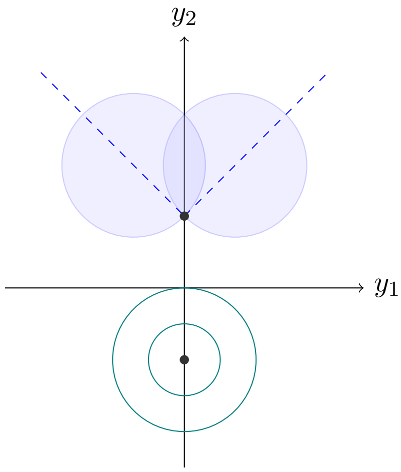

Figure 4.2: Illustration of the lower-level problem of Example 4.6

need to hold. Thus, we can construct a bilevel feasible sequence (x^k, y^k) *with* $y^k = (0, y_2^k)$, $y_2^k > 0$ *for all* k *and* (x_1^k, x_2^k) *satisfying* $x_1^k = x_2^k$ *for all* k. *This sequence converges to* (x^*, y^*) *for* $y_2^k \searrow 0$ *but the leader's objective function values are strictly negative for all* k. *Hence,* (x^*, y^*) *is* not *a local minimum of the bilevel problem.*

Let us now come to the KKT reformulation of the problem, which is given by

$$
\begin{aligned}
\min_{x,y,\lambda} \quad & -y_2 \\
\text{s.t.} \quad & y_1 y_2 = 0, \ x \geq 0, \\
& (y_1 - x_1)^2 + (y_2 - 1 - x_1)^2 \leq 1, \\
& (y_1 + x_2)^2 + (y_2 - 1 - x_2)^2 \leq 1, \\
& \lambda_1 \geq 0, \ \lambda_2 \geq 0, \\
& \nabla_{y_1} \mathcal{L} = 2y_1 + 2\lambda_1(y_1 - x_1) + 2\lambda_2(y_1 + x_2) = 0, \\
& \nabla_{y_2} \mathcal{L} = 2(y_2 + 1) + 2\lambda_1(y_2 - 1 - x_1) + 2\lambda_2(y_2 - 1 - x_2) = 0, \\
& \lambda_1(1 - (y_1 - x_1)^2 - (y_2 - 1 - x_1)^2) = 0, \\
& \lambda_2(1 - (y_1 + x_2)^2 - (y_2 - 1 - x_2)^2) = 0.
\end{aligned}
$$

Again, we are considering points with $y_2 > 0$. *Thus,* $y_1 = 0$ *still needs to hold and* $x_1 = x_2$ *is still valid, too. The partial derivative of the Lagrangian w.r.t.* y_1 *then leads to* $\lambda_1 = \lambda_2$. *Moreover, for the point* (x^*, y^*) *with* $x^* = (0, 0)$ *and* $y^* = (0, 0)$ *we obtain* $\lambda_1^* + \lambda_2^* = 1$ *from the second partial derivative of the Lagrangian. This motivates the definition of the same sequence* (x^k, y^k, λ^k) *as above but extended with* $\lambda_1^k = \lambda_2^k$ *for all* k. *Moreover, we know that the limit of the multipliers is* $(\lambda_1^*, \lambda_2^*) = (1/2, 1/2)$. *As before, we see that the point* (x^*, y^*, λ^*) *is* not *a local minimum of the KKT reformulation. However,* *all other points* $(x^*, y^*, \tilde{\lambda})$ *with* $\tilde{\lambda} \in \Lambda(x^*, y^*)$ *are local minima of the KKT reformulation.*

Finally, since (x^*, y^*) *is not a local minimum of the bilevel problem, the KKT reformulation has local minima that do not correspond to local minima of the bilevel problem.*

4.4 A Mixed-Integer Linear Reformulation of the KKT Reformulation

The only reason for the nonconvexity of Problem (4.6) are the bilinear products of the lower-level dual variables λ_i and the upper-level primal variables x in the term

$$\lambda_i C_{i\cdot} x$$

and the bilinear products of the lower-level dual variables λ_i and the lower-level primal variables y in the term

$$\lambda_i D_{i\cdot} y.$$

We can linearize these terms if we exploit the combinatorial structure of the KKT complementarity conditions in (4.6d). The key idea here is to consider the complementarity conditions $\lambda_i(C_{i\cdot}x + D_{i\cdot}y - b_i) = 0$, $i = 1, \dots, \ell$, as disjunctions stating that either

$$\lambda_i = 0 \quad \text{or} \quad C_{i\cdot}x + D_{i\cdot}y = b_i$$

needs to hold. These two cases can be modeled using binary variables

$$z_i \in \{0, 1\}, \quad i = 1, \dots, \ell,$$

in the following mixed-integer linear way:

$$\lambda_i \le M z_i, \quad C_{i\cdot}x + D_{i\cdot}y - b_i \le M(1 - z_i).$$

Here, M is a sufficiently large constant. Consequently, $z_i = 1$ models the case that the primal inequality is active, whereas $z_i = 0$ models the inactive case in which the dual variable is zero. By construction, we thus obtain the following result.

Theorem 4.7. *Suppose that* M *is a sufficiently large constant. Then, Problem* (4.6) *is equivalent to the mixed-integer linear optimization problem*

$$\begin{aligned}
\min_{x,y,\lambda,z} \quad & c_x^\top x + c_y^\top y && (4.13a)\\
\text{s.t.} \quad & Ax + By \ge a, \ Cx + Dy \ge b, && (4.13b)\\
& D^\top \lambda = d, \ \lambda \ge 0, && (4.13c)\\
& \lambda_i \le M z_i \quad \textit{for all } i = 1, \dots, \ell, && (4.13d)\\
& C_{i\cdot}x + D_{i\cdot}y - b_i \le M(1 - z_i) \quad \textit{for all } i = 1, \dots, \ell, && (4.13e)\\
& z_i \in \{0, 1\} \quad \textit{for all } i = 1, \dots, \ell. && (4.13f)
\end{aligned}$$

Here, "equivalence" means that for every globally optimal solution (x, y, λ) of Problem (4.6) there exists z so that (x, y, λ, z) is a globally optimal solution of Problem (4.13) and that for every globally optimal solution (x, y, λ, z) of Problem (4.13), (x, y, λ) is a globally optimal solution of Problem (4.6).

The resulting MILP reformulation can then be solved by general-purpose MILP solvers such as Gurobi, CPLEX, or SCIP.[2] Unfortunately, this reformulation has a severe disadvantage because one needs to determine a big-M constant that is both valid for the primal constraint as well as for the dual variable. The primal validity is usually ensured by the assumption that the high-point relaxation is bounded, which is typically justified in practical applications. However, the dual feasible set is unbounded for bounded primal feasible sets; see Clark (1961) and Williams (1970). Thus, it is rather problematic to bound the dual variables of the follower. In practice, often "standard" values such as magic constants like 10^6 are used without any theoretical justification or heuristics are applied to compute a big-M value. For instance, in Pineda, Bylling, et al. (2018), big-M values are determined from local solutions of the MPCC (4.6). In Pineda and Morales (2019) it is shown by an illustrative counter-example that such heuristics may deliver invalid values. Moreover, validating the correctness of a given big-M is shown to be NP-hard in general in Kleinert, Labbé, Plein, et al. (2020).

Exercise 4.2. Derive the mixed-integer linear reformulation of the KKT reformulation of the LP-QP bilevel problem

$$\begin{aligned} \min_{x,y} \quad & c_x^\top x + c_y^\top y \\ \text{s.t.} \quad & Ax + By \geq a, \\ & y \in \arg\min_{\bar{y}} \left\{ \frac{1}{2} y^\top Q y + d^\top y \colon Cx + Dy \geq b \right\} \end{aligned} \tag{4.14}$$

with x, $c_x \in \mathbb{R}^{n_x}$, y, c_y, $d \in \mathbb{R}^{n_y}$, $A \in \mathbb{R}^{m \times n_x}$, $B \in \mathbb{R}^{m \times n_y}$, $a \in \mathbb{R}^m$, $C \in \mathbb{R}^{\ell \times n_x}$, $D \in \mathbb{R}^{\ell \times n_y}$, $b \in \mathbb{R}^\ell$, and $Q \in \mathbb{R}^{n_y \times n_y}$ being symmetric and positive semi-definite. What does qualitatively change with respect to the LP-LP case? (*Hint*: What about the dual polyhedron of the lower level?)

Exercise 4.3. Read the paper "Solving Linear Bilevel Problems Using Big-Ms: Not All That Glitters Is Gold" (Pineda and Morales 2019).

Exercise 4.4. Read the paper "There's No Free Lunch: On the Hardness of Choosing a Correct Big-M in Bilevel Optimization" (Kleinert, Labbé, Plein, et al. 2020).

4.5 The Strong-Duality Based Reformulation

Besides the approach based on the lower-level's KKT conditions one can also use a strong-duality theorem for the lower-level problem if such a theorem is

[2] Gurobi and CPLEX are commercial software packages that can be used for free in a purely academic context whereas SCIP is an open-source solver.

at hand. In the linear case considered up to now, this is the case—so let's go. The dual problem to (4.5) is given by

$$\max_{\lambda} \quad (b - Cx)^\top \lambda \quad \text{s.t.} \quad D^\top \lambda = d, \ \lambda \geq 0. \tag{4.15}$$

Note that also this dual problem is an x-parameterized linear problem but now the objective function (and not the constraints) depends on x. For a given decision x of the leader, weak duality of linear optimization states that

$$d^\top y \geq (b - Cx)^\top \lambda$$

holds for every primal and dual feasible pair y and λ. Thus, by strong duality, we know that every such feasible pair is a pair of optimal solutions if

$$d^\top y \leq (b - Cx)^\top \lambda$$

holds. Consequently, we can reformulate the bilevel problem as

$$\begin{aligned} \min_{x,y,\lambda} \quad & c_x^\top x + c_y^\top y && \text{(4.16a)}\\ \text{s.t.} \quad & Ax + By \geq a, \ Cx + Dy \geq b, && \text{(4.16b)}\\ & D^\top \lambda = d, \ \lambda \geq 0, && \text{(4.16c)}\\ & d^\top y \leq (b - Cx)^\top \lambda. && \text{(4.16d)} \end{aligned}$$

Here, the ℓ KKT complementarity constraints in (4.6d) are replaced with the scalar inequality in (4.16d). Note that the general nonconvexity of LP-LP bilevel problems is reflected in this single-level reformulation due to the bilinear products of the primal upper-level variables x and the dual lower-level variables λ.

Remark 4.8. *Obviously, the KKT reformulation* (4.6) *and the strong-duality based reformulation* (4.16) *are equivalent since*

$$\begin{aligned} & \lambda_i (C_{i\cdot} x + D_{i\cdot} y - b_i) = 0 \quad \textit{for all } i = 1, \dots, \ell \\ \iff \quad & \lambda^\top (Cx + Dy - b) = 0 \\ \iff \quad & \lambda^\top D y = \lambda^\top (b - Cx) \\ \iff \quad & d^\top y = \lambda^\top (b - Cx) \end{aligned}$$

holds, where we used $\lambda \geq 0$ *and* $Cx + Dy - b \geq 0$ *for the first equivalence as well as* $D^\top \lambda = d$ *for the last one.*

Note that the strong-duality inequality in (4.16d) does not allow to exploit disjunctive arguments (such as in the case of the KKT reformulation), which further would allow to linearize the nonlinearities with additional binary variables. Thus, in the case of (4.16), one has to tackle the nonlinearity and nonconvexity directly.

Exercise 4.5. Derive the strong-duality based reformulation of the LP-QP bilevel problem

$$\begin{aligned}
\min_{x,y} \quad & c_x^\top x + c_y^\top y \\
\text{s.t.} \quad & Ax + By \geq a, \\
& y \in \arg\min_{\bar{y}} \left\{ \frac{1}{2} y^\top Q y + d^\top y \colon Cx + Dy \geq b \right\}
\end{aligned} \tag{4.17}$$

with x, $c_x \in \mathbb{R}^{n_x}$, y, c_y, $d \in \mathbb{R}^{n_y}$, $A \in \mathbb{R}^{m \times n_x}$, $B \in \mathbb{R}^{m \times n_y}$, $a \in \mathbb{R}^m$, $C \in \mathbb{R}^{\ell \times n_x}$, $D \in \mathbb{R}^{\ell \times n_y}$, $b \in \mathbb{R}^\ell$, and $Q \in \mathbb{R}^{n_y \times n_y}$ being symmetric and positive semi-definite.

4.6 Excursus: How to Really Solve a Mixed-Integer Linear Problem?

We have seen that we can re-state the LP-LP bilevel problem as a mixed-integer linear optimization problem if we are able to find finite but sufficiently large big-M constants. Let's assume this is possible. How do I then solve the resulting problem?

It's easy! Download a mixed-integer optimization solver, for instance, Gurobi at

https://www.gurobi.com.

Get a free academic license and start to code. But how? Here, we discuss how to do this using the programming language Python.[3] Let's exemplarily see how this works.

```
#!/usr/bin/python3

# This example is a modified model taken from
# https://bit.ly/3nrwl2Z

import gurobipy as gp
from gurobipy import GRB

# Create a new model
model = gp.Model("my-milp")

# Create variables
x = model.addVar(vtype=GRB.BINARY, name="x")
y = model.addVar(vtype=GRB.BINARY, name="y")
z = model.addVar(vtype=GRB.BINARY, name="z")
```

[3] If you do not have it installed on your computer go to https://www.python.org and download it. It's for free.

```
# Set objective
model.setObjective(x + y + 2 * z, GRB.MAXIMIZE)

# Add constraint: x + 2 y + 3 z <= 4
model.addConstr(x + 2 * y + 3 * z <= 4, "c0")

# Add constraint: x + y >= 1
model.addConstr(x + y >= 1, "c1")

# Optimize model
model.optimize()

for v in model.getVars():
    print("variable " + v.varName + ": " + str(v.x))

print("Objective value: " + str(model.objVal))
```

This is the output that you should get:

```
Academic license - for non-commercial use only
Optimize a model with 2 rows, 3 columns and 5 nonzeros
Variable types: 0 continuous, 3 integer (3 binary)
Coefficient statistics:
  Matrix range     [1e+00, 3e+00]
  Objective range  [1e+00, 2e+00]
  Bounds range     [1e+00, 1e+00]
  RHS range        [1e+00, 4e+00]
Found heuristic solution: objective 2.0000000
Presolve removed 2 rows and 3 columns
Presolve time: 0.00s
Presolve: All rows and columns removed

Explored 0 nodes (0 simplex iterations) in 0.00 seconds
Thread count was 1 (of 8 available processors)

Solution count 2: 3 2

Optimal solution found (tolerance 1.00e-04)
Best objective 3.000000000000e+00,
   best bound 3.000000000000e+00, gap 0.0000%
variable x: 1.0
variable y: 0.0
variable z: 1.0
Objective value: 3.0
```

Exercise 4.6. Consider the linear bilevel problem

$$\begin{aligned} \min_{x,y} \quad & x - y \\ \text{s.t.} \quad & x - 2y \geq -8, \\ & y \in \arg\min_{\bar{y}} \{\bar{y} \colon x + \bar{y} \geq 7,\ -x + 5\bar{y} \geq 2,\ -x + 2\bar{y} \geq -4,\ -x - \bar{y} \geq -13\}. \end{aligned} \tag{4.18}$$

(i) Derive the KKT reformulation of Problem (4.18).

(ii) Derive the mixed-integer linear reformulation of the KKT reformulation.

(iii) Determine appropriate big-M constants for the reformulation in (ii).

(iv) Use Gurobi to solve

  (a) the mixed-integer linear reformulation of the KKT reformulation of Problem (4.18) using the big-M constants determined in (iii),

  (b) the linearized KKT reformulation of Problem (4.18) exploiting special ordered sets of type 1 (SOS1).

(v) Derive the strong-duality based reformulation of Problem (4.18).

(vi) Use Gurobi to solve the strong-duality based reformulation.

(vii) Elaborate on possible advantages and/or disadvantages of the considered approaches. Which approach should be preferred and why?

5

Linear Bilevel Problems

We start this chapter on linear bilevel problems again with another example that illustrates the surprising properties of bilevel problems that have a linear upper- and lower-level problem.

5.1 Surprising Properties—Revisited

Example 5.1 (See Kleinert, Labbé, Ljubić, et al. (2021))**.** *We consider the problem*

$$\min_{x,y} \left\{ y \colon y \in \arg\min_{\bar{y}} \{-\bar{y} \colon (x, \bar{y}) \in \mathcal{P}\} \right\},$$

with the lower-level's feasible region given by

$$\mathcal{P} = \{(x, y) \colon y \geq 0,\ y \leq 1 + x,\ y \leq 3 - x,\ 0 \leq x \leq 1\}$$

is depicted in Figure 5.1 (left). The feasible points of the high-point relaxation coincide with the lower-level feasible region $\mathcal{P}$ since there are no upper-level

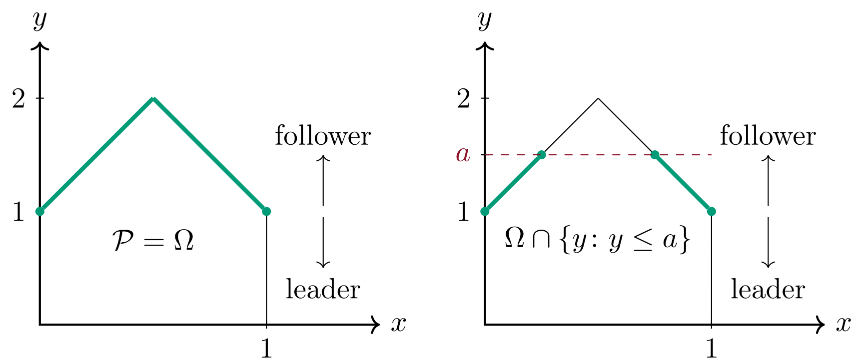

Figure 5.1: Illustration of the LP-LP bilevel problem of Example 5.1.

constraints in this example. The horizontal segment linking the origin and the point $(1,0)$ *constitutes the set of solutions of the high-point relaxation, i.e., those points in* Ω *that minimize the upper-level objective function. Since the corresponding upper-level objective function is* 0 *on this segment, this leads to a lower bound of* 0 *for the entire bilevel LP. The bilevel feasible region* $\mathcal{F}$ *is given by the union of the two segments in green. As we have seen before,* $\mathcal{F}$ *is nonconvex although both levels are linear optimization problems. The problem has the two optimal solutions* $(0,1)$ *and* $(1,1)$ *with value* 1.

Now, if we add the constraint $y \leq a$ *with* $1 < a < 2$ *to the upper level, the bilevel feasible region is reduced to two disjoint green segments as depicted in Figure 5.1 (right). Nonetheless, these segments constitute faces of the high-point relaxation. Note, however, that the set of optimal solutions of the bilevel problem remains unchanged. A worse situation happens if the right-hand side of the constraint added to the upper level is set to* $a \in (0,1)$. *Then, the bilevel feasible region is empty, i.e., the bilevel LP has no feasible point, although the high-point relaxation is feasible. This last example is also useful to illustrate the effect of moving coupling constraints, i.e., upper-level constraints involving variables of the lower level, between the two levels. If, e.g., the constraint* $y \leq 1/2$ *is added to the lower level, then the problem becomes feasible and all points* $(x, 1/2)$ *with* $0 \leq x \leq 1$ *are bilevel optimal. The two facts that (i) coupling constraints of a bilevel LP may lead to a disconnected bilevel feasible region and that (ii) they cannot be moved to the lower level without changing the set of optimal solutions have been discussed by Audet, Haddad, et al. (2006) and Mersha and Dempe (2006).*

Example 5.2 (See Kleinert, Manns, et al. (2021))**.** *Let us now consider the linear bilevel problem*

$$\begin{aligned}
\min_{x,y \in \mathbb{R}} \quad & x \\
\text{s.t.} \quad & y \geq 0.5x + 1, \ \ x \geq 0, \\
& y \in \operatorname*{arg\,min}_{\bar{y} \in \mathbb{R}} \left\{ \bar{y} \colon \bar{y} \geq 2x - 2, \ \ \bar{y} \geq 0.5 \right\},
\end{aligned}$$

with optimal solution $(2,2)$*; see Figure 5.2 (left). When strengthening the bound* $\bar{y} \geq 0.5$ *in the lower-level problem using the constraints* $y \geq 0.5x + 1$ *of the upper-level problem, one finds that the minimum value of* $0.5x + 1$ *is* 1 *due to* $x \geq 0$, *which increases the bound of* $\bar{y}$ *to* $\bar{y} \geq 1$. *This yields the problem*

$$\begin{aligned}
\min_{x,y \in \mathbb{R}} \quad & x \\
\text{s.t.} \quad & y \geq 0.5x + 1, \ \ x \geq 0, \\
& y \in \operatorname*{arg\,min}_{\bar{y} \in \mathbb{R}} \{ \bar{y} \colon \bar{y} \geq 2x - 2, \ \ \bar{y} \geq 1 \},
\end{aligned}$$

having the optimal solution $(0,1) \neq (2,2)$*; see Figure 5.2 (right). See also the thesis by Manns (2020) for further examples.*

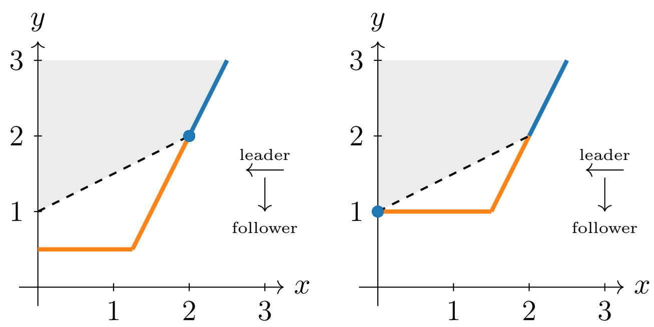

Figure 5.2: Feasible set and optimal solution of the bilevel problem in Example 5.2 without (left) and with (right) bound tightening applied. Figure taken from Kleinert, Manns, et al. (2021).

What have we seen now? We first solved an LP-LP bilevel problem and then added another constraint to the lower-level problem. This new (or tightened) constraint is not active in the original solution of the bilevel problem but changes the optimal solution if added. This cannot happen in "ordinary", i.e., single-level, optimization problems.

In single-level optimization, this means that the so-called IIC (independence of irrelevant constraints) property holds. In Macal and Hurter (1997), the IIC property for linear bilevel problems is defined as follows.

Definition 5.3 (Independence of irrelevant constraints)**.** *Let $\mathcal{S}$ be the set of optimal solutions of a linear bilevel problem P. Further, let $\tilde{P} := P(u, v, w)$ be the modified problem, in which the inequality $u^\top x + v^\top y \geq w$ is added to the follower problem of P and let $\tilde{\mathcal{S}}$ be its set of optimal solutions. P is called* independent of irrelevant constraints*, if for any $(u, v, w) \in \mathbb{R}^{n_x+n_y+1}$ with $u^\top x^* + v^\top y^* \geq w$ it holds*

$$(x^*, y^*) \in \tilde{\mathcal{S}}$$

for every $(x^*, y^*) \in \mathcal{S}$.

Further, it is shown in Macal and Hurter (1997) that only bilevel problems possess the IIC property for which the solution of the high-point relaxation is also a solution to the bilevel problem. Consequently, most practical bilevel problems (for which the objectives of the leader and the follower are not aligned) lack the IIC property.

Exercise 5.1. Consider the linear bilevel problem

$$\begin{aligned} \min_{x,y} \quad & x + 3y \\ \text{s.t.} \quad & 1 \leq x \leq 5, \\ & y \in \arg\min_{\bar{y}} \left\{ -\bar{y} \colon x + \bar{y} \geq 3,\, 0 \leq \bar{y} \leq 3 \right\}. \end{aligned} \tag{P}$$

(i) Determine the solution of the HPR of Problem (P), e.g., using Gurobi.

(ii) Check if Problem (P) possesses the IIC property using Theorem 5.5.

(iii) Without explicitly determining an optimal solution of the bilevel problem (P), what can you say about the relation between an optimal solution of Problem (P) and the solution of its HPR?

(iv) Consider the augmented problem ($\tilde{\mathrm{P}}$), in which the inequality $-2x - y \geq -6$ is added to the follower problem of (P). Is ($\tilde{\mathrm{P}}$) independent of irrelevant constraints? What about the relation between an optimal solution of the bilevel problem ($\tilde{\mathrm{P}}$) and the solution of its HPR?

(v) What can you say about the relation between the set of optimal solutions $\mathcal{S}$ of Problem (P) and the set of optimal solutions $\tilde{\mathcal{S}}$ of ($\tilde{\mathrm{P}}$)?

Hint: You can use the following result without proof.

Assumption 5.4. *Let $(\hat{x}, \hat{y})$ be a solution of the high-point relaxation (HPR) of the bilevel problem*

$$\begin{aligned} \min_{x \in \mathbb{R}^n, y} \quad & F(x, y) \\ \text{s.t.} \quad & G(x, y) \geq 0, \\ & y \in \arg\min_{\bar{y} \in \mathbb{R}^m} \{ f(x, \bar{y}) \colon g(x, \bar{y}) \geq 0 \} \end{aligned} \tag{5.1}$$

and let $\mathcal{A}(\hat{x}, \hat{y})$ be the set of active lower-level constraints at $(\hat{x}, \hat{y})$, i.e., $g_i(\hat{x}, \hat{y}) = 0$ if and only if $i \in \mathcal{A}(\hat{x}, \hat{y})$. Then, there exist $\lambda_i \geq 0$ for all $i \in \mathcal{A}(\hat{x}, \hat{y})$ with

$$\nabla_{y_j} f(\hat{x}, \hat{y}) - \sum_{i \in \mathcal{A}(\hat{x}, \hat{y})} \lambda_i \nabla_{y_j} g_i(\hat{x}, \hat{y}) = 0 \quad \text{for all } j = 1, \ldots, m.$$

Theorem 5.5 (See Theorem 1 in Macal and Hurter (1997))**.** *Problem (5.1) is independent of irrelevant constraints (IIC) if and only if Assumption 5.4 is satisfied.*

5.2 Geometric Properties of LP-LP Bilevel Problems

For the remainder of this chapter, we consider LP-LP bilevel problems of the form

$$\min_{x,y} \quad c_x^\top x + c_y^\top y \tag{5.2a}$$
$$\text{s.t.} \quad Ax \geq a, \tag{5.2b}$$
$$y \in \arg\min_{\bar{y}} \left\{ d^\top \bar{y} \colon Cx + D\bar{y} \geq b \right\} \tag{5.2c}$$

with $c_x \in \mathbb{R}^{n_x}$, $c_y, d \in \mathbb{R}^{n_y}$, $A \in \mathbb{R}^{m \times n_x}$, and $a \in \mathbb{R}^m$ as well as $C \in \mathbb{R}^{\ell \times n_x}$, $D \in \mathbb{R}^{\ell \times n_y}$, and $b \in \mathbb{R}^\ell$. Note that this problem does not contain coupling constraints to avoid the further difficulties that arise due to disconnected bilevel feasible sets.

Our goal now is to understand the geometric properties of LP-LP bilevel problems. The main source of the remainder of this section is the book by Bard (1998).

Theorem 5.6. *Consider the LP-LP bilevel problem (5.2). Suppose that $S(x)$ is a singleton for all $x \in \Omega_x$ and that Ω is non-empty and bounded. The bilevel-feasible set of Problem (5.2) can then be written equivalently as the intersection of the shared constraint set with the feasible points of a piecewise linear equality constraint. In particular, the bilevel-feasible set is a union of faces of the shared constraint set.*

This claim should not be a surprise at this point since we saw this fact already in Example 1.12 and 5.1.

Proof. We start by first re-writing the bilevel-feasible set

$$\mathcal{F} := \{(x, y) \colon (x, y) \in \Omega,\ y \in S(x)\}$$

explicitly as

$$\mathcal{F} := \left\{(x, y) \colon (x, y) \in \Omega,\ d^\top y = \min_{\bar{y}}\{d^\top \bar{y} \colon Cx + D\bar{y} \geq b\}\right\}$$

and use the optimal value function

$$\varphi(x) = \min_{y} \left\{d^\top y \colon Dy \geq b - Cx\right\}$$

again. Since $S(x)$ is a singleton for all $x \in \Omega_x$, the optimal value function $\varphi(x)$ is a well-defined function. By using the strong-duality theorem (Theorem 2.3), we can also express the optimal value function by means of the dual LP as

$$\varphi(x) = \max_{\lambda} \left\{(b - Cx)^\top \lambda \colon D^\top \lambda = d,\ \lambda \geq 0\right\}.$$

From the classic theory of linear optimization we know that the optimal solution is attained in one of the vertices of the feasible set, which, for the dual LP, does not depend on the leader's decision x anymore. Let $\lambda^1, \ldots, \lambda^s$ be the set of all the dual polyhedron's vertices, i.e., the set of vertices of the polyhedron defined by

$$D^\top \lambda = d, \quad \lambda \geq 0. \tag{5.3}$$

Thus, we can further equivalently re-write the optimal value function as

$$\varphi(x) = \max \left\{(b - Cx)^\top \lambda \colon \lambda \in \{\lambda^1, \ldots, \lambda^s\}\right\}. \tag{5.4}$$

This shows that $\varphi(x)$ is a piecewise linear function and re-writing the bilevel-feasible set as

$$\mathcal{F} = \left\{(x, y) \in \Omega \colon d^\top y - \varphi(x) = 0\right\} \tag{5.5}$$

shows the claim that the bilevel-feasible set can be written as the intersection of the shared constraint set with a piecewise linear equality constraint.

Consider now again the definition of the optimal value function using the vertices of the dual polyhedron of the lower-level problem in (5.4). Suppose that for a given x the corresponding solution is the vertex λ^k. By using dual feasibility (5.3), we obtain

$$0 = d^\top y - \varphi(x) = (D^\top \lambda^k)^\top y - (\lambda^k)^\top (b - Cx) = (\lambda^k)^\top (Cx + Dy - b).$$

Thus, for those λ_i^k, $i \in \{1, \dots, \ell\}$, with $\lambda_i^k > 0$ we get $(Cx + Dy - b)_i = 0$. Hence, the bilevel-feasible set is a union of faces of the shared constraint set. $\square$

In other words, the latter result states the following.

Corollary 5.7. *Suppose that the assumptions of Theorem (5.6) hold. Then, the LP-LP bilevel problem (5.2) is equivalent to minimizing the upper-level's objective function over the intersection of the shared constraint set with a piecewise linear equality constraint.*

The last results make clear what we have also seen before in Example 1.12 and 5.1.

Corollary 5.8. *Suppose that the assumptions of Theorem (5.6) hold. Then, a solution of the LP-LP bilevel problem (5.2) is always attained at a vertex of the bilevel-feasible set.*

Theorem 5.9. *Suppose that the assumptions of Theorem (5.6) hold. Then, a solution (x^*, y^*) of the LP-LP bilevel problem (5.2) is always attained at a vertex of the shared constraint set Ω.*

Proof. Let $(x^1, y^1), \dots, (x^r, y^r)$ be the distinct vertices of the shared constraint set Ω. Since Ω is a convex polyhedron, any point in Ω can be written as a convex combination of these vertices, i.e.,

$$(x^*, y^*) = \sum_{i=1}^{r} \alpha_i (x^i, y^i)$$

with

$$\sum_{i=1}^{r} \alpha_i = 1 \quad \text{and} \quad \alpha_i \geq 0 \quad \text{for all } i = 1, \dots, r.$$

From the proof of Theorem 5.6 it follows that the optimal value function φ is convex and continuous. Since the bilevel solution (x^*, y^*) is, of course, bilevel feasible, the representation in (5.5) leads to

$$\begin{aligned} 0 &= d^\top y^* - \varphi(x^*) \\ &= d^\top \left(\sum_{i=1}^{r} \alpha_i y^i\right) - \varphi\left(\sum_{i=1}^{r} \alpha_i x^i\right) \\ &\geq \sum_{i=1}^{r} \alpha_i d^\top y^i - \sum_{i=1}^{r} \alpha_i \varphi(x^i) \\ &= \sum_{i=1}^{r} \alpha_i \left(d^\top y^i - \varphi(x^i)\right). \end{aligned} \tag{5.6}$$

By the definition of the optimal value function we also have

$$\varphi(x^i) = \min_y \left\{d^\top y \colon Cx^i + Dy \geq b\right\} \leq d^\top y^i.$$

This implies $d^\top y^i - \varphi(x^i) \geq 0$. Consequently, for all $i \in \{1, \dots, r\}$ with $\alpha_i > 0$ it holds $d^\top y^i = \varphi(x^i)$ since we otherwise get a contradiction in (5.6). Hence, for those i with $\alpha_i > 0$ we obtain $(x^i, y^i) \in \mathcal{F}$. From the last corollary we know that (x^*, y^*) is a vertex of the bilevel-feasible set. Suppose now that there are two indices i and j with $\alpha_i > 0$ and $\alpha_j > 0$. Thus, $(x^i, y^i) \in \mathcal{F}$ and $(x^j, y^j) \in \mathcal{F}$ holds and we can write (x^*, y^*) as a proper convex combination of two bilevel feasible points, which is a contradiction to the last corollary. Thus, (x^*, y^*) is a vertex of the shared constraint set. □

By combining the last theorem with the last corollary, we have seen that the set of vertices of the bilevel-feasible set $\mathcal{F}$ is a subset of the vertices of the shared constraint set Ω. This also shows that $\mathcal{F}$ consists of faces of Ω and that every extreme point of $\mathcal{F}$ is an extreme point of Ω.[1]

5.3 Complexity Results

The first paper proving the hardness of the LP-LP bilevel problem was Jeroslow (1985), in which general multilevel models have been considered. As a direct consequence of the results in this paper one obtains the NP-hardness of LP-LP bilevel problems. The problem is also strongly NP-hard, which is shown in Hansen et al. (1992) by a reduction from KERNEL. Two years later it has been shown in Vicente, Savard, et al. (1994) that even checking whether a given point is a local minimum of a bilevel problem is NP-hard.

[1] A subset E of the convex set $S \subseteq \mathbb{R}^n$ is called an *extremal set* of S if $z \in E$ with $z = \lambda x + (1-\lambda)y$ and $0 < \lambda < 1$, $x, y \in S$ implies that $x, y \in E$ holds. An *extreme point* of S is an extremal set that is a singleton.

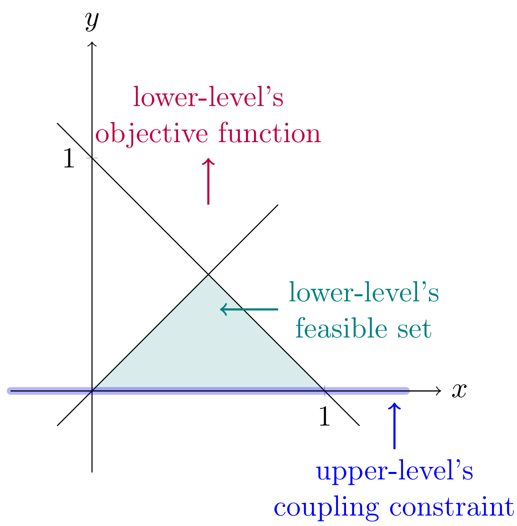

Figure 5.3: Modeling a binary variable with an LP-LP bilevel feasibility problem.

Although we are not going to prove these hardness results here, we try to get some intuition on the hardness of LP-LP bilevel problems. We have already seen that LP-LP bilevel problems are nonconvex optimization problems, which are hard to solve to global optimality in general. Moreover, we know that mixed-integer (or mixed-binary) optimization is hard as well. In the paper by Audet, Hansen, et al. (1997) it is noted that a binary constraint, say $x \in \{0, 1\}$, appearing in a single-level optimization problem can be modeled by an additional variable y and the upper-level constraints $y = 0$ and

$$y = \arg\max_{\bar{y}} \{\bar{y} \colon \bar{y} \le x, \ \bar{y} \le 1 - x\};$$

see Figure 5.3 for an illustration. As a consequence, linear optimization problems with binary variables are a special case of bilevel LPs.

6

Algorithms for Linear Bilevel Problems

6.1 The Kth-Best Algorithm

One of the first proposed algorithms to solve LP-LP bilevel problems is the simplex-inspired Kth-best algorithm; see Bialas and Karwan (1984). We again consider the LP-LP bilevel problem

$$\min_{x,y} \quad c_x^\top x + c_y^\top y \tag{6.1a}$$

$$\text{s.t.} \quad Ax \geq a, \tag{6.1b}$$

$$y \in \arg\min_{\bar{y}} \left\{ d^\top \bar{y} \colon Cx + D\bar{y} \geq b \right\} \tag{6.1c}$$

as in Section 5.2. Moreover, we assume that the bilevel-feasible set is nonempty and bounded and that $S(x)$ is a singleton for all $x \in \Omega_x$.

The idea is mainly based on Theorem 5.9, which states that the bilevel-optimal solution is attained at one of the vertices of the shared constraint set Ω. Thus, similar to the simplex method for linear problems, we can carry out a search over the vertices of Ω to find a solution. To this end, we consider the high-point relaxation

$$\min_{x,y} \quad c_x^\top x + c_y^\top y \tag{6.2a}$$

$$\text{s.t.} \quad Ax \geq a, \tag{6.2b}$$

$$Cx + Dy \geq b \tag{6.2c}$$

of Problem (6.1). Let us denote with

$$(x^1, y^1), (x^2, y^2), \ldots, (x^r, y^r) \tag{6.3}$$

the ordered set of vertices of Ω, i.e., of basic feasible solutions of the high-point relaxation. The ordering is chosen so that

$$c_x^\top x^i + c_y^\top y^i \leq c_x^\top x^{i+1} + c_y^\top y^{i+1}$$

holds for $i = 1, \dots, r-1$.

Hence, the problem of solving the LP-LP bilevel problem can be posed as finding the minimum-index vertex that is feasible for the bilevel problem, i.e., we want to find the index

$$K^* = \min \left\{ i \in \{1, \dots, r\} \colon (x^i, y^i) \in \mathcal{F} \right\}.$$

This means that we want to find the first vertex in the ordered list in (6.3) whose y-component is an optimal solution of the follower's problem. It is then clear that (x^{K^*}, y^{K^*}) is a global optimal solution of the LP-LP bilevel problem (6.1).

The method is formally given in Algorithm 1.

Algorithm 1 The Kth-Best Algorithm

1: Set $i \leftarrow 1$. Solve Problem (6.2) to obtain the optimal solution (x^1, y^1). Set $W \leftarrow \{(x^1, y^1)\}$ and $T \leftarrow \emptyset$.

2: Test if $y^i \in S(x^i)$ holds, i.e., if y^i is the optimal follower's response to the leader's decision x^i. To this end, we solve the x^i-parameterized follower's problem

$$\begin{aligned} \min_y \quad & d^\top y \\ \text{s.t.} \quad & Dy \geq b - Cx^i. \end{aligned}$$

Let us denote the optimal solution by $\tilde{y}$.

3: **if** $\tilde{y} = y^i$ **then**

4: Set $K^* \leftarrow i$ and return the LP-LP bilevel solution (x^i, y^i).

5: **end if**

6: Let W^i denote the adjacent extreme points of (x^i, y^i) such that $(x, y) \in W^i$ implies

$$c_x^\top x + c_y^\top y \geq c_x^\top x^i + c_y^\top y^i.$$

Set $T \leftarrow T \cup \{(x^i, y^i)\}$ and $W \leftarrow (W \cup W^i) \setminus T$.

7: Set $i \leftarrow i + 1$ and choose (x^i, y^i) with

$$c_x^\top x^i + c_y^\top y^i = \min_{x,y} \left\{ c_x^\top x + c_y^\top y \colon (x, y) \in W \right\}.$$

Go to Step 2.

Remark 6.1. *(a) Note that the uniqueness of the follower's problem is required in Step 2 and 3, where we check if the current vertex is bilevel feasible.*

(b) A crucial and costly part of the algorithm (that we do not discuss here in detail) is the computation of all adjacent extreme points in Step 6. For more details; see Bard (1998).

Exercise 6.1. Reconsider the linear bilevel problem (P).

(i) Solve Problem (P) graphically.

(ii) Determine all the vertices of the shared constraint set of Problem (P) and check which vertices are adjacent.

(iii) Solve Problem (P) by hand using the Kth-best algorithm (Algorithm 1).

6.2 Branch-and-Bound

In Section 4.2 we have seen that the general LP-LP bilevel problem

$$\begin{aligned} \min_{x,y} \quad & c_x^\top x + c_y^\top y \\ \text{s.t.} \quad & Ax + By \geq a, \\ & y \in \arg\min_{\bar{y}} \left\{ d^\top \bar{y} \colon Cx + D\bar{y} \geq b \right\} \end{aligned}$$

can be equivalently re-written via the KKT reformulation as the MPCC

$$\begin{aligned} \min_{x,y,\lambda} \quad & c_x^\top x + c_y^\top y \\ \text{s.t.} \quad & Ax + By \geq a, \ Cx + Dy \geq b, \\ & D^\top \lambda = d, \ \lambda \geq 0, \\ & \lambda_i (C_{i\cdot} x + D_{i\cdot} y - b_i) = 0 \quad \text{for all } i = 1, \dots, \ell. \end{aligned}$$

Moreover, we have seen in Section 4.4 that the latter problem can be re-stated as the mixed-integer linear optimization problem

$$\begin{aligned} \min_{x,y,\lambda,z} \quad & c_x^\top x + c_y^\top y && \text{(6.4a)} \\ \text{s.t.} \quad & Ax + By \geq a, \ Cx + Dy \geq b, && \text{(6.4b)} \\ & D^\top \lambda = d, \ \lambda \geq 0, && \text{(6.4c)} \\ & \lambda_i \leq M z_i \quad \text{for all } i = 1, \dots, \ell, && \text{(6.4d)} \\ & C_{i\cdot} x + D_{i\cdot} y - b_i \leq M(1 - z_i) \quad \text{for all } i = 1, \dots, \ell, && \text{(6.4e)} \\ & z_i \in \{0, 1\} \quad \text{for all } i = 1, \dots, \ell. && \text{(6.4f)} \end{aligned}$$

We can solve this problem without further ado by putting it into a state-of-the-art mixed-integer solver such as Gurobi (Gurobi 2021) or CPLEX (IBM 2021).

These solvers will tackle this problem using the classic branch-and-bound method; see Land and Doig (1960) for the original paper. This means, they will branch on the auxiliary binary variables that model whether the ith lower-level constraint is binding or whether the ith dual variable vanishes. Alternatively, we can branch on the KKT complementarity constraints

$$\lambda_i(C_{i\cdot}x + D_{i\cdot}y - b_i) = 0 \quad \text{for all } i = 1, \dots, \ell,$$

which is, mathematically speaking, the same. However, the latter complementarity-constraint based branching does not require choosing sufficiently large big-M values, which is often a drawback of the mixed-integer linear approach (6.4) for the KKT reformulation.

The idea of branch-and-bound for LP-LP bilevel problems is simple:

(a) Start with solving the problem

$$\begin{aligned} \min_{x,y,\lambda} \quad & c_x^\top x + c_y^\top y && (6.5a)\\ \text{s.t.} \quad & Ax + By \geq a, \ Cx + Dy \geq b, && (6.5b)\\ & D^\top \lambda = d, \ \lambda \geq 0. && (6.5c) \end{aligned}$$

This is the high-point relaxation extended with the dual variables λ and the lower level's dual polyhedron given by

$$D^\top \lambda = d, \quad \lambda \geq 0.$$

(b) Usually, there will be an $i \in \{1, \dots, \ell\}$ so that the ith KKT complementarity condition is not satisfied, i.e.,

$$\lambda_i(C_{i\cdot}x + D_{i\cdot}y - b_i) > 0$$

holds. Take such an i and construct two new sub-problems: one in which the constraint

$$\lambda_i = 0$$

is added and one in which the constraint

$$C_{i\cdot}x + D_{i\cdot}y = b_i$$

is added.

(c) Then, we choose one of the unsolved sub-problems and proceed in the same way.

Every node in the branch-and-bound tree is thus defined by the root-node problem (6.5) as well as the index sets $D \subseteq \{1, \dots, \ell\}$ and $P \subseteq \{1, \dots, \ell\}$ that contain those indices for which the dual constraint $\lambda_i = 0$ or the primal constraint $C_{i\cdot}x + D_{i\cdot}y = b_i$ is added to Problem (6.5), respectively. Thus, we denote a node by its corresponding index-set pair (P, D), which again corresponds to the problem

$$
\begin{aligned}
\min_{x,y,\lambda} \quad & c_x^\top x + c_y^\top y && \text{(6.6a)}\\
\text{s.t.} \quad & Ax + By \geq a, \ Cx + Dy \geq b, && \text{(6.6b)}\\
& D^\top \lambda = d, \ \lambda \geq 0, && \text{(6.6c)}\\
& C_{i\cdot} x + D_{i\cdot} y = b_i \quad \text{for all } i \in P, && \text{(6.6d)}\\
& \lambda_i = 0 \quad \text{for all } i \in D. && \text{(6.6e)}
\end{aligned}
$$

This leads to the branch-and-bound method formally stated in Algorithm 2.

Algorithm 2 Branch-and-Bound for LP-LP Bilevel Problems

1: $u \leftarrow +\infty$ and $Q \leftarrow \{(\emptyset, \emptyset)\}$.
2: **while** $Q \neq \emptyset$ **do**
3: Choose any $(P, D) \in Q$ and set $Q \leftarrow Q \setminus \{(P, D)\}$.
4: Solve Problem (6.6) for P and D.
5: **if** Problem (6.6) for P and D is infeasible **then**
6: Go to Step 2.
7: **end if**
8: Let $(\bar{x}, \bar{y}, \bar{\lambda})$ denote the solution of Problem (6.6) for P and D.
9: **if** $c_x^\top \bar{x} + c_y^\top \bar{y} \geq u$ **then**
10: Go to Step 2.
11: **end if**
12: **if** $(\bar{x}, \bar{y}, \bar{\lambda})$ satisfies the complementarity conditions $\lambda_i (C_{i\cdot} x + D_{i\cdot} y - b_i) = 0$ for all $i \in \{1, \dots, \ell\}$ **then**
13: Set $(x^*, y^*, \lambda^*) \leftarrow (\bar{x}, \bar{y}, \bar{\lambda})$ as well as $u \leftarrow c_x^\top x^* + c_y^\top y^*$ and go to Step 2.
14: **end if**
15: Choose any $i \in \{1, \dots, \ell\}$ with $\lambda_i (C_{i\cdot} x + D_{i\cdot} y - b_i) > 0$ and set $Q \leftarrow Q \cup \{(P \cup \{i\}, D), (P, D \cup \{i\})\}$.
16: **end while**
17: **if** $u < +\infty$ **then**
18: Return the optimal solution (x^*, y^*, λ^*).
19: **else**
20: Return the statement "The given LP-LP bilevel problem is infeasible."
21: **end if**

Let us now analyze this branch-and-bound method. First, we formally introduce the notion of a relaxation.

Definition 6.2 (Relaxation)**.** *Consider the optimization problem* $\min\{f(x) \colon x \in \mathcal{F}\}$*. The optimization problem* $\min\{g(x) \colon x \in \mathcal{F}'\}$ *is* *called a* relaxation *of the other problem if* $\mathcal{F} \subseteq \mathcal{F}'$ *and if* $g(x) \leq f(x)$ *holds for all* $x \in \mathcal{F}$.

The easiest way to obtain a relaxation is to simply delete constraints from a given set of constraints. This is exactly what we did to derive the high-point relaxation, which means that the wording is reasonable.

Moreover, we see that Problem (6.6) for given P and D, i.e.,

$$\begin{aligned} \min_{x,y,\lambda} \quad & c_x^\top x + c_y^\top y \\ \text{s.t.} \quad & Ax + By \geq a, \; Cx + Dy \geq b, \\ & D^\top \lambda = d, \; \lambda \geq 0, \\ & C_{i \cdot} x + D_{i \cdot} y = b_i \quad \text{for all } i \in P, \\ & \lambda_i = 0 \quad \text{for all } i \in D, \end{aligned}$$

is a relaxation of the problem

$$\begin{aligned} \min_{x,y,\lambda} \quad & c_x^\top x + c_y^\top y && \text{(6.7a)} \\ \text{s.t.} \quad & Ax + By \geq a, \; Cx + Dy \geq b, && \text{(6.7b)} \\ & D^\top \lambda = d, \; \lambda \geq 0, && \text{(6.7c)} \\ & \lambda_i (C_{i \cdot} x + D_{i \cdot} y - b_i) = 0 \quad \text{for all } i \in \{1, \dots, \ell\}, && \text{(6.7d)} \\ & C_{i \cdot} x + D_{i \cdot} y = b_i \quad \text{for all } i \in P, && \text{(6.7e)} \\ & \lambda_i = 0 \quad \text{for all } i \in D. && \text{(6.7f)} \end{aligned}$$

Note that the latter is the KKT reformulation of the LP-LP bilevel problem, which we extended by the equality constraints corresponding to the sets P and D.

To prove the correctness of the branch-and-bound method in Algorithm 2, we have to show that

(a) the bounding step in Step 9 as well as the pruning[1] of infeasible nodes in Step 5 are correct and that

(b) the branching step in Step 15 is correct.

This is done in the following two lemmas.

Lemma 6.3 (Bounding Lemma). *Let* $P, D \subseteq \{1, \dots, \ell\}$ *be given. Moreover, denote the optimal objective function value of the relaxation* (6.6) *by* z^{rel} *and the optimal objective function value of Problem* (6.7) *by* z *(if they exist; otherwise they are set to* ∞*). Then, it holds*

$$z^{rel} \leq z.$$

Furthermore, the infeasibility of the relaxation (6.6) *implies the infeasibility of Problem* (6.7).

[1]"Pruning nodes" is also often called "fathoming nodes".

Proof. Both statements immediately follow from the definition of a relaxation (Definition 6.2). □

Lemma 6.4 (Branching Lemma)**.** *Let $P, D \subseteq \{1, \dots, \ell\}$ be given. Moreover, let the point (x, y, λ) be feasible for Problem (6.7) for given sets P and D. Let $i \in \{1, \dots, \ell\}$. Then, the point (x, y, λ) is either feasible for Problem (6.7) for the sets $(P \cup \{i\}, D)$ or for Problem (6.7) for the sets $(P, D \cup \{i\})$.*

Theorem 6.5 (Correctness Theorem)**.** *Suppose that the root-node relaxation (6.5) of the KKT reformulation (4.6) is bounded. Then, Algorithm 2 terminates after a finite number of visited nodes with a global optimal solution of (4.6) or with the correct indication of infeasibility.*

Proof. The only thing that is left to prove is that the algorithm terminates after a finite number of visited nodes. This, however, follows immediately since we only have a finite number of KKT complementarity conditions to branch on. □

To sum up, we have seen that we can use a tailored branch-and-bound method to solve LP-LP bilevel problems. In particular, we have seen that it does not require the choice of any big-M values. Moreover, it is comparably easy to implement, which is not necessarily true for the Kth-best algorithm; see Remark 6.1.

Remark 6.6. *It is rather easy to realize a branch-and-bound method for linear bilevel problems in modern mixed-integer linear solvers such as* Gurobi *or* CPLEX *by using so-called* special ordered sets of type 1 *(SOS1).*

A set of non-negative variables $x_1, \dots, x_n$ is called a special ordered set of type 1 if there exists exactly one index $i \in \{1, \dots, n\}$ with $x_i > 0$ and $x_j = 0$ for all $j \neq i$. We denote this property of the set of variables $x_1, \dots, x_n$ in the following via

$$\mathit{SOS1}(x_1, \dots, x_n).$$

This property of a subset of variables of a mixed-integer linear optimization problem can also be communicated to a general-purpose solver such as those mentioned above.

If we now introduce the non-negative auxiliary variables

$$s_i = (C_{i\cdot}x + D_{i\cdot}y - b_i) \quad \text{for} \quad i = 1, \dots, \ell$$

we can state the complementarity conditions

$$(C_{i\cdot}x + D_{i\cdot}y - b_i) = 0 \quad \text{or} \quad \lambda_i = 0 \quad \text{for} \quad i = 1, \dots, \ell$$

equivalently as

$$\mathit{SOS1}(s_i, \lambda_i) \quad \text{for} \quad i = 1, \dots, \ell.$$

By doing so, the mixed-integer linear solver takes care of the branching on these SOS1 conditions.

Exercise 6.2. Consider the linear bilevel problem

$$\begin{aligned} \min_{x,y_1,y_2} \quad & 7x - 2y_1 + 3y_2 \\ \text{s.t.} \quad & x \geq 0, \\ & y \in S(x), \end{aligned} \tag{6.8}$$

where $S(x)$ is the set of optimal solutions of the x-parameterized problem

$$\begin{aligned} \min_{y=(y_1,y_2)} \quad & 2y_1 - y_2 \\ \text{s.t.} \quad & 4x + y_1 + y_2 \leq 3, \\ & 2x - 2y_1 + 5y_2 \leq 5, \\ & 3x - y_1 - 2y_2 \leq 1, \\ & y_1, y_2 \geq 0. \end{aligned} \tag{6.9}$$

(i) Find out what the so-called breadth-first search and depth-first search methods are. What is the difference?

(ii) Determine the KKT reformulation of the bilevel problem (6.8) and (6.9).

(iii) Apply the branch-and-bound method (Algorithm 2) to solve the bilevel problem (6.8) and (6.9) by hand. Use Gurobi to solve the arising subproblems and visualize your progress in a search tree. Apply the branch-and-bound method twice using

  (a) the depth-first search strategy,

  (b) the breadth-first search strategy

  to select the index-set pair (P, D) in Step 3 of the algorithm.

(iv) Compare your results. Which strategy is faster to obtain a feasible solution? Which strategy is faster to find the optimal solution and to verify optimality?

(v) Verify the solution found in (iii) by using Gurobi to solve the linearized KKT reformulation of the bilevel problem (6.8) and (6.9) by exploiting SOS1-type constraints.

6.3 A Penalty Alternating Direction Method for LP-LP Bilevel Problems

We have seen in the previous chapters that LP-LP bilevel problem are, in general, hard optimization problems—both in theory and practice. In such a situation it can be reasonable to also consider heuristic solution methods, i.e., methods that do not provably compute a global minimum in finite time. Instead, such heuristics are usually fast methods without optimality guarantee for the output. However, one wants to compute at least a feasible point that is, hopefully, of good quality. In this section, we derive such a heuristic for LP-LP bilevel problems.

In contrast to the KKT reformulation that we used to set up the branch-and-bound method in the last section, we now start with the reformulation based on the strong-duality theorem. This means, we consider Problem (4.16), i.e.,

$$\begin{aligned}
\min_{x,y,\lambda} \quad & c_x^\top x + c_y^\top y \\
\text{s.t.} \quad & Ax + By \geq a, \ Cx + Dy \geq b, \\
& D^\top \lambda = d, \ \lambda \geq 0, \\
& d^\top y \leq (b - Cx)^\top \lambda.
\end{aligned}$$

As we have discussed in Section 4.5, the "only" nasty aspect of the latter reformulation is the strong-duality inequality

$$d^\top y \leq (b - Cx)^\top \lambda. \tag{6.10}$$

This constraint leads to a nonconvex feasible set due to the bilinear product

$$x^\top C^\top \lambda,$$

which is nonconvex since both upper-level primal variables x and lower-level dual variables λ are variables of the reformulation. The key idea now is to split the reformulated problem (4.16) into two problems that are much easier to solve since they are split along the just discussed bilinearity.

However, before we do so, we briefly dive into the field of (penalty) alternating direction methods (ADMs).

6.3.1 Penalty Alternating Direction Methods

We now briefly review the alternating direction method (ADM) and an extension of this method; the penalty ADM (PADM). In the next section, we then discuss how these methods can be used to compute a stationary point of the classic strong-duality based single-level reformulation of the linear bilevel problem.

We start with discussing general ADMs. To this end, we consider an optimization problem in the specific form

$$\min_{x,y} \quad f(x,y) \tag{6.11a}$$

$$\text{s.t.} \quad g(x,y) = 0, \quad h(x,y) \geq 0, \tag{6.11b}$$

$$x \in X, \quad y \in Y. \tag{6.11c}$$

Here, $x \in \mathbb{R}^n$ and $y \in \mathbb{R}^m$ are variable vectors. The feasible set of this problem is abbreviated by

$$\mathcal{F} := \{(x,y) \in X \times Y : g(x,y) = 0, \, h(x,y) \geq 0\} \subseteq X \times Y.$$

For discussing the theoretical properties of ADMs, we need the following assumption.

Algorithm 3 A Standard Alternating Direction Method

1: Choose initial values $(x^0, y^0) \in X \times Y$.
2: **for** $i = 0, 1, \dots$ **do**
3: Compute

$$x^{i+1} \in \arg\min_x \left\{ f(x, y^i) \colon g(x, y^i) = 0,\ h(x, y^i) \geq 0,\ x \in X \right\}.$$

4: Compute

$$y^{i+1} \in \arg\min_y \left\{ f(x^{i+1}, y) \colon g(x^{i+1}, y) = 0,\ h(x^{i+1}, y) \geq 0,\ y \in Y \right\}.$$

5: **end for**

Assumption 6.7. *The objective function $f : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ and the constraint functions $g : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^k$ and $h : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^\ell$ are continuous and the sets X and Y are non-empty and compact.*

A standard ADM proceeds as follows. Given an iterate (x^i, y^i), we first solve Problem (6.11) for y fixed to y^i. Thus, we obtain a new x-iterate x^{i+1}. We now fix x to this new iterate x^{i+1}, solve Problem (6.11) again, and obtain y^{i+1}. Repeating these two steps yields the method that is listed in Algorithm 3.

Under certain mild assumptions one can show that the ADM of Algorithm 3 converges to a partial minimum.

Definition 6.8 (Partial Minimum)**.** *A feasible point $(x^*, y^*) \in \mathcal{F}$ of Problem (6.11) is called a* partial minimum *if*

$$\begin{aligned} f(x^*, y^*) &\leq f(x, y^*) \quad \text{for all } (x, y^*) \in \mathcal{F}, \\ f(x^*, y^*) &\leq f(x^*, y) \quad \text{for all } (x^*, y) \in \mathcal{F} \end{aligned}$$

holds.

The following general convergence result is taken from Gorski et al. (2007).

Theorem 6.9. *Let $\left\{(x^i, y^i)\right\}_{i=0}^\infty$ be a sequence with $(x^{i+1}, y^{i+1}) \in \Theta(x^i, y^i)$, where*

$$\begin{aligned} \Theta(x^i, y^i) := \Big\{ (x^*, y^*) : \ & f(x^*, y^i) \leq f(x, y^i) \text{ for all } x \in X, \\ & f(x^*, y^*) \leq f(x^*, y) \text{ for all } y \in Y \Big\}. \end{aligned}$$

Suppose that Assumption 6.7 holds and that the solution of the first optimization problem is always unique. Then, every convergent subsequence of $\left\{(x^i, y^i)\right\}_{i=0}^\infty$ converges to a partial minimum. For two limit points z, z' of such subsequences it holds that $f(z) = f(z')$.

For what follows, we also note that stronger convergence results can be obtained if stronger assumptions on f and $\mathcal{F}$ are made. For later reference, we state these results as a corollary; see Geißler et al. (2017, 2018), Gorski et al. (2007), and Wendell and Hurter (1976) for the proofs and more detailed discussions.

Corollary 6.10. *Suppose that the assumptions of Theorem 6.9 are satisfied. Then, the following holds:*

(a) If f is continuously differentiable, then every convergent subsequence of $\left\{(x^i, y^i)\right\}_{i=0}^{\infty}$ converges to a stationary point of Problem (6.11).

(b) If f is continuously differentiable and if f and $\mathcal{F}$ are convex, then every convergent subsequence of $\left\{(x^i, y^i)\right\}_{i=0}^{\infty}$ converges to a global minimum of Problem (6.11).

Let us comment on the main rationale of the alternating direction method discussed so far. The considered Problem (6.11) can be seen as a quasi block-separable problem, where the blocks are given by the variables x and y as well as their respective feasible sets X and Y. We add the notion "quasi" here since there still are the constraints g and h that couple the feasible sets of the two blocks. The main idea of an ADM is to alternatingly solve in the directions of the blocks separately until the method stagnates.

In practice it can often be observed that an even stronger decoupling of Problem (6.11) is favorable (Boyd et al. 2011; Geißler et al. 2015, 2017, 2018). Thus, we now go one step further and relax the coupling constraints g and h. To this end, we introduce the weighted ℓ_1 penalty function

$$\phi_1(x, y; \mu, \rho) := f(x, y) + \sum_{t=1}^{k} \mu_t |g_t(x, y)| + \sum_{t=1}^{\ell} \rho_t [h_t(x, y)]^-.$$

Here, $[\alpha]^- := \max\{0, -\alpha\}$ holds and μ and ρ are vectors of penalty parameters of size k and ℓ, respectively. The penalty ADM consists of an inner and an outer loop. In the inner loop we apply a standard ADM like in Algorithm 3 to the penalty problem

$$\min_{x,y} \quad \phi_1(x, y; \mu, \rho) \quad \text{s.t.} \quad x \in X, \ y \in Y. \tag{6.12}$$

If this inner loop iteration terminates with a partial minimum of Problem (6.12), we check whether the coupling constraints are satisfied. If they are, we terminate. If not, we increase the penalty parameters and proceed with computing a partial minimum of the new penalty problem in the next inner loop. This method is formally stated in Algorithm 4. For later reference, we also state the convergence results for the PADM (Algorithm 4), which have been derived in Geißler et al. (2017). There, all details and proofs can be found.

Algorithm 4 The ℓ_1 Penalty Alternating Direction Method

```
Choose initial values (x^{0,0}, y^{0,0}) ∈ X × Y and initial penalty parameters μ^0, ρ^0 ≥ 0.
for j = 0, 1, ... do
    Set i ← 0.
    while (x^{j,i}, y^{j,i}) is not a partial minimum of (6.12) with μ = μ^j and ρ = ρ^j do
        Compute x^{j,i+1} ∈ arg min_x {φ_1(x, y^{j,i}; μ^j, ρ^j): x ∈ X}.
        Compute y^{j,i+1} ∈ arg min_y {φ_1(x^{j,i+1}, y; μ^j, ρ^j): y ∈ Y}.
        Set i ← i + 1.
    end while
    if g(x^{j,i}, y^{j,i}) = 0 and h(x^{j,i}, y^{j,i}) ≥ 0 then
        Return (x^{j,i}, y^{j,i}).
    else
        Choose new penalty parameters μ^{j+1} ≥ μ^j and ρ^{j+1} ≥ ρ^j.
    end if
end for
```

Theorem 6.11. *Suppose that Assumption 6.7 holds and that $\mu_t^j \nearrow \infty$ for all $t = 1, \dots, k$ and $\rho_t^j \nearrow \infty$ for all $t = 1, \dots, \ell$. Moreover, let $\{(x^j, y^j)\}_{j=0}^{\infty}$ be a sequence of partial minima of (6.12) (for $\mu = \mu^j$ and $\rho = \rho^j$) generated by Algorithm 4 with $(x^j, y^j) \to (x^*, y^*)$. Then, there exist weights $\bar{\mu}, \bar{\rho} \geq 0$ such that (x^*, y^*) is a partial minimizer of the weighted ℓ_1 feasibility measure*

$$\chi_{\bar{\mu},\bar{\rho}}(x, y) := \sum_{t=1}^{k} \bar{\mu}_t |g_t(x, y)| + \sum_{t=1}^{\ell} \bar{\rho}_t [h_t(x, y)]^- .$$

If, in addition, (x^*, y^*) is feasible for the original problem (6.11), the following holds:

(a) If f is continuous, then (x^*, y^*) is a partial minimum of Problem (6.11).

(b) If f is continuously differentiable, then (x^*, y^*) is a stationary point of Problem (6.11).

(c) If f is continuously differentiable and if f and $\mathcal{F}$ are convex, then (x^*, y^*) is a global optimum of Problem (6.11).

6.3.2 Applying the PADM to Linear Bilevel Problems

Now, we apply the PADM to the single-level reformulation (4.16) of the original LP-LP bilevel problem. As already discussed, the problematic constraint is the strong-duality inequality (6.10). This is due to two reasons. On the one hand, it is the only nonlinear constraint and thus the reason for the nonconvexity of the problem. On the other hand, it is the only constraint that couples the variable blocks (x, y) and λ. Thus, we relax this constraint and obtain the penalty problem reformulation

$$\begin{aligned}
\min_{x,y,\lambda} \quad & c_x^\top x + c_x^\top y + \rho \left[b^\top \lambda - x^\top C^\top \lambda - d^\top y \right]^- && (6.13a)\\
\text{s.t.} \quad & Ax + By \geq a, && (6.13b)\\
& Cx + Dy \geq b, && (6.13c)\\
& D^\top \lambda = d, && (6.13d)\\
& \lambda \geq 0. && (6.13e)
\end{aligned}$$

Moreover, we smoothen the penalty term by exploiting weak duality of the lower level that is equivalent to that

$$d^\top y - b^\top \lambda + x^\top C^\top \lambda \geq 0$$

holds for every feasible point of Problem (6.13). Thus,

$$\left[b^\top \lambda - x^\top C^\top \lambda - d^\top y \right]^- = \max\{0, d^\top y - b^\top \lambda + x^\top C^\top \lambda\} = d^\top y - b^\top \lambda + x^\top C^\top \lambda$$

holds and we obtain the equivalent penalty problem

$$\min_{x,y,\lambda} \quad c_x^\top x + c_y^\top y + \rho \left(d^\top y - b^\top \lambda + x^\top C^\top \lambda \right) \quad \text{s.t.} \quad (6.13b)\text{–}(6.13e), \quad (6.14)$$

which is a smooth but still nonconvex optimization problem. To be more specific, Problem (6.14) is a nonconvex quadratic optimization problem. A closer look also reveals that Problem (6.14) is exactly of the form in (6.12) if the first block of variables is (x, y) and if the second block of variables is λ. Thus, the splitting of the feasible set is obtained by identifying[2]

$$\begin{aligned}
x \in X \quad & \longleftrightarrow \quad \text{Constraints (6.13b), (6.13c)}, && (6.15a)\\
y \in Y \quad & \longleftrightarrow \quad \text{Constraints (6.13d), (6.13e)}. && (6.15b)
\end{aligned}$$

Note further that this splitting corresponds to a primal-dual splitting of the single-level reformulation (4.16). Given this splitting, the first sub-problem that needs to be solved if we apply Algorithm 4 to Problem (6.14) reads

$$\begin{aligned}
\min_{x,y} \quad & c_x^\top x + c_y^\top y + \rho \left((C^\top \bar{\lambda})^\top x + d^\top y \right) && (6.16a)\\
\text{s.t.} \quad & Ax + By \geq a, \quad Cx + Dy \geq b, && (6.16b)
\end{aligned}$$

[2]Note that in (6.15), x and y denote the variable blocks of the general problem formulation (6.11). All other occurrences of x and y in this section stand for the respective upper- and lower-level variables of the considered bilevel problem.

where $\bar{\lambda}$ is a given constant vector and where we already omitted the constant objective function term $b^\top \bar{\lambda}$. This problem has the same feasible set as the classic high-point relaxation of the original LP-LP bilevel problem. However, the objective function coefficients are modified in dependence of the penalty parameter ρ, the current dual estimate $\bar{\lambda}$, and the lower-level objective function coefficients d.

The second sub-problem is equivalent to

$$\max_{\lambda} \quad (b - C\bar{x})^\top \lambda \tag{6.17a}$$

$$\text{s.t.} \quad D^\top \lambda = d, \tag{6.17b}$$

$$\lambda \geq 0, \tag{6.17c}$$

which is exactly the dual lower-level problem. Here, three interesting aspects can be observed. First, the second sub-problem only depends on the upper level's primal variables $\bar{x}$—not on the lower level's primal variables $\bar{y}$. The reason is that we again omit constant terms in the objective function, i.e., $c_x^\top \bar{x} + c_y^\top \bar{y} + \rho d^\top \bar{y}$. Second, Sub-problem (6.17) does not depend on the penalty parameter ρ anymore since it only scales the remaining objective function. Third, the second sub-problem may be unbounded for a given estimate for $\bar{x}$. This means that the primal lower-level problem is infeasible for the upper-level decision $\bar{x}$. We do not go into the details on how to resolve this here. Some more details can be found in Kleinert and Schmidt (2019a).

By using Corollary 6.10 and Theorem 6.11 we now state two theoretical results for the PADM (Algorithm 4) applied to Problem (4.16).

Theorem 6.12. *Consider the inner loop of Algorithm 4 applied to Problem (6.14) for a fixed penalty parameter $\rho > 0$ and let $\{(x^{j,i}, y^{j,i}, \lambda^{j,i})\}_{i=0}^{\infty}$ be the generated sequence of iterates. Moreover, let one of the two sub-problems (6.16) or (6.17) always have a unique solution. Then, every convergent subsequence of $\{(x^{j,i}, y^{j,i}, \lambda^{j,i})\}_{i=0}^{\infty}$ converges to a stationary point of the penalty problem (6.14).*

Proof. The feasible set of Problem (6.14) is a polyhedron and thus convex and the objective function is continuously differentiable but nonconvex. Thus, Case (i) of Corollary 6.10 applies. □

Regarding the obtained stationary points, two possible situations may appear. First, the stationary point may have a strong-duality error

$$\chi^{\text{sd}}(x, y, \lambda) := \left| d^\top y - b^\top \lambda + x^\top C^\top \lambda \right|$$

of zero, which means that the stationary point is also bilevel feasible. Second, the stationary point has a non-zero strong-duality error $\chi^{\text{sd}}(x, y, \lambda) > 0$ and is thus not bilevel feasible. The latter case then motivates to proceed with a larger penalization of the strong-duality term.

Using now Theorem 6.11 implies the following main convergence result.

Theorem 6.13. *Suppose that $\rho^j \nearrow \infty$ holds and let $\{(x^j, y^j, \lambda^j)\}_{j=0}^{\infty}$ be a sequence of stationary points of* (6.14) *(for $\rho = \rho^j$) generated by Algorithm 4 applied to Problem* (4.16) *with $(x^j, y^j, \lambda^j) \to (x^*, y^*, \lambda^*)$ for $j \to \infty$. Then, (x^*, y^*, λ^*) is a stationary point of the strong duality error χ^{sd}. If, in addition, (x^*, y^*, λ^*) is bilevel feasible, i.e., the strong duality error χ^{sd} is zero, then (x^*, y^*, λ^*) is a stationary point of Problem* (4.16).

We close the discussion of applying an PADM to linear bilevel problems with three remarks.

Remark 6.14. *Note that Theorem 6.13 "only" makes a statement regarding stationary points of the single-level reformulation* (4.16) *and not about the original bilevel problem. In general, a stationary point of the Problem* (4.16) *does not need to be a stationary point of the bilevel problem; see Section 4.3 and Dempe and Dutta (2012) for the equivalent setting of a single-level reformulation based on KKT conditions of the lower level. Although we thus do not have any theoretical quality guarantee for the bilevel feasible points obtained by our method, it is shown in Kleinert and Schmidt (2019a) that, in practice, the quality of the obtained solutions is very good.*

Remark 6.15. *A crucial assumption of Theorem 6.12 (and that is also implicitly present in Theorem 6.13) is that one of the two PADM sub-problems always needs to have a unique solution. In the context of bilevel optimization this is strongly connected to the topic of unique lower-level solutions. It is well known that a bilevel problem can be very ill-behaved if its lower-level problem does not possess a unique solution for all possible decisions of the leader. Almost the same situation appears in the previous theorems despite the fact that dual uniqueness is required instead of primal uniqueness of the lower level if one considers the uniqueness of the second PADM sub-problem. Uniqueness of the first PADM sub-problem translates to unique solutions of the extended high-point relaxation* (6.16).

Remark 6.16. *The approach described in this section can also be applied to bilevel problems for which the upper level contains integrality constraints and a convex-quadratic objective function, i.e., problems of the form*

$$\min_{x,y} \quad \frac{1}{2} x^\top H_u x + c_x^\top x + \frac{1}{2} y^\top G_u y + c_y^\top y \tag{6.18a}$$

$$\text{s.t.} \quad Ax + By \geq a, \tag{6.18b}$$

$$x_i \in \mathcal{Z} \subset \mathbb{Z} \quad \text{for all} \quad i \in I \subseteq \{1, \dots, n\}, \tag{6.18c}$$

$$y \in \arg\min_{\bar{y}} \left\{ d^\top \bar{y} \colon Cx + D\bar{y} \geq b \right\}, \tag{6.18d}$$

with symmetric and positive semidefinite matrices H_u and G_u in appropriate dimensions. This does not affect the second PADM sub-problem at all. However, the first PADM sub-problem (6.16) *is a convex-quadratic problem (QP)* *for $I = \emptyset$ and a mixed-integer convex-quadratic problem (MIQP) for $I \neq \emptyset$. Solving (MI)QPs to global optimality in every iteration may have significant impact on the performance of the PADM. For a numerical analysis we refer to Kleinert and Schmidt (2019a).*

7

Mixed-Integer Linear Bilevel Problems

In this section, we focus on general bilevel mixed-integer linear problems (MILPs), which are defined as

$$\min_{x \in X, y} \quad c_x^\top x + c_y^\top y \tag{7.1a}$$

$$\text{s.t.} \quad Ax \geq a, \tag{7.1b}$$

$$y \in \operatorname*{arg\,min}_{\bar{y} \in Y} \left\{ d^\top \bar{y} \colon Cx + D\bar{y} \geq b \right\}, \tag{7.1c}$$

where the vectors c_x, c_y, d, a, b and matrices A, B, C, D are defined as before. The sets X and Y specify integrality constraints on a subset of x- and y-variables, respectively.

The shared constraint set of this bilevel MILP is, as usual, defined as the set of points $(x, y) \in X \times Y$ satisfying all constraints of the upper and lower level, i.e.,

$$\Omega := \{(x, y) \in X \times Y \colon Ax \geq a,\, Cx + Dy \geq b\}.$$

The bilevel-feasible set of this bilevel MILP consists of all points $(x, y) \in \Omega$ from the shared constraint set for which for a given x, the vector y is an optimal solution of the lower-level problem. This means,

$$d^\top y \leq \varphi(x)$$

holds. Here, $\varphi(x)$ again is the optimal value of the lower-level problem, which is defined as

$$\varphi(x) = \min_{y \in Y} \left\{ d^\top y \colon Dy \geq b - Cx \right\}. \tag{7.2}$$

The optimal value function $\varphi(x)$ thus corresponds to a parametric MILP in this case. Hence, it is nonconvex, not continuous, and in general very difficult to describe.

Remark 7.1 (Hardness of bilevel MILPs)**.** *In contrast to bilevel LPs, it is now NP-hard to check whether a given point* (x, y) *is a feasible solution of the bilevel MILP. Jeroslow (1985) showed that* k*-level discrete optimization problems are* Σ_k^p*-hard, even when the variables are binary and all constraints are linear. This means that, e.g., a discrete bilevel optimization problem can be solved in nondeterministic polynomial time, provided that there exists an oracle that solves problems in constant time that are in NP.*

The inducible region of the bilevel MILP is contained in the shared constraint set Ω and, therefore, minimizing the objective function of the upper level over the shared constraint set Ω (which represents another MILP) provides a valid lower bound for the bilevel MILP. Consequently, solving the LP-relaxation of the high-point relaxation provides another (and usually much weaker) lower bound of the bilevel MILP.

Moore and Bard (1990) initiated the studies of bilevel optimization problems involving discrete variables. Their illustrative example (see Figure 7.1 in Section 7.1) is frequently used in the literature to highlight the major differences and pitfalls arising in discrete bilevel optimization. Since then, studies have been carried out considering only special cases, e.g., by assuming binary variables at both levels or by considering purely linear problems at the lower level. Exact MILP-based procedures for the general case in which both the upper and the lower level are MILPs have been studied mainly in the last decade.

7.1 The Example by Moore and Bard

The following example is provided by Moore and Bard (1990). We consider the discrete bilevel problem

$$\min_{x \in \mathbb{Z}, y \in \mathbb{Z}} \left\{ -x - 10y \colon y \in \arg\min_{\bar{y} \in \mathbb{Z}} \{\bar{y} \colon (x, \bar{y}) \in \mathcal{P}\} \right\},$$

where $\mathcal{P}$ is a polytope defined by

$$-25x + 20\bar{y} \leq 30, \quad x + 2\bar{y} \leq 10, \quad 2x - \bar{y} \leq 15, \quad 2x + 10\bar{y} \geq 15.$$

The high-point relaxation of this problem is an integer linear problem, whose feasible region is depicted in Figure 7.1. The unique optimal solution for this example is the point $(x^*, y^*) = (2, 2)$ with optimal objective function value -22, which is in the interior of the convex hull of the high-point relaxation. This is in contrast to bilevel LPs, whose optimal solution is always a vertex of the shared constraint set (Theorem 5.9). The example also shows that relaxing the integrality constraints does neither provide a lower nor an upper bound for the bilevel MILP. We will discuss this in more detail later when we come back to this example to discuss how to derive a

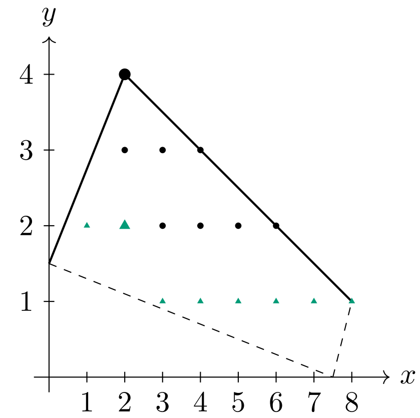

Figure 7.1: Example of a bilevel MILP (taken from Moore and Bard (1990)): Discrete points are feasible for the high-point relaxation. The point $(2, 4)$ is the optimal solution of the high-point relaxation and $(2, 2)$ is the optimal solution of the bilevel MILP. Triangles represent bilevel feasible solutions and dashed lines represent the feasible region of the bilevel LP in which the integrality constraints on the upper- and lower-level variables are relaxed.

branch-and-bound method for solving mixed-integer linear bilevel problems. In Figure 7.1, dashed lines correspond to the inducible region of the problem in which the integrality constraints for both the upper-level and the lower-level variables are relaxed. In general, this set does not even have to contain a single bilevel feasible point.

7.2 Attainability of Optimal Solutions

In Vicente, Savard, et al. (1996), the authors consider three cases of bilevel MILPs and study the following different assumptions:

(i) only upper-level variables are discrete,

(ii) all upper- and lower-level variables are discrete,

(iii) only lower-level variables can take discrete values.

Assuming that all discrete variables are bounded and that the bilevel-feasible set is non-empty, they show that for Case (i) and (ii), an optimal solution always exists and that (i) can be reduced to a linear bilevel program, whereas (ii) can be "reduced" to a linear trilevel problem. However, for Case (iii), Moore and Bard (1990) and also Vicente, Savard, et al. (1996) provide examples that demonstrate that the bilevel feasible region may not be closed and, hence, the optimal solution may not be attainable. The following simpler

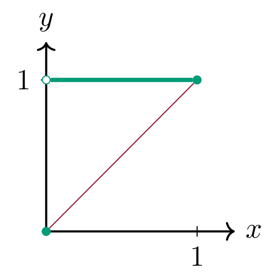

Figure 7.2: The attainability counterexample by Köppe et al. (2010)

example (see Figure 7.2) is due to Köppe et al. (2010):

$$\inf_{0\leq x\leq 1,\, y} \left\{ x - y \colon y \in \arg\min_{\bar{y}\in\mathbb{Z}} \{\bar{y} \colon \bar{y} \geq x,\ 0 \leq \bar{y} \leq 1\} \right\},$$

which is equivalent to

$$\inf_{x} \{x - \lceil x \rceil \colon 0 \leq x \leq 1\}.$$

In this problem, the infimum is -1, which is never attained. In the existing literature on bilevel MILPs, it is therefore frequently assumed that the linking variables are discrete. We recall that non-linking upper-level variables can be moved to the lower level (Bolusani and Ralphs 2020; Tahernejad et al. 2020), which effectively translates the latter assumption into "all upper-level variables are discrete".

7.3 A Branch-and-Bound Method for Mixed-Integer Bilevel Problems

We now discuss the first branch-and-bound method for solving mixed-integer linear bilevel problems. This algorithm has been published in the seminal paper by Moore and Bard (1990).

To this end, we consider the mixed-integer linear bilevel problem

$$\min_{x\in X,\, y} \quad c_x^\top x + c_y^\top y \tag{7.3a}$$

$$\text{s.t.} \quad Ax \geq a, \tag{7.3b}$$

$$y \in \arg\min_{\bar{y}\in Y} \left\{ d^\top \bar{y} \colon Cx + D\bar{y} \geq b \right\}. \tag{7.3c}$$

In what follows, the variables x and y are split in $x = (x_{C_x}, x_{I_x})$ and $y = (y_{C_y}, y_{I_y})$. Here, x_{C_x} and y_{C_y} are the subsets of upper- as well as lower-level variables, respectively, that are continuous-valued and x_{I_x} and y_{I_y} are the subsets of upper- as well as lower-level variables, respectively, that are integer-valued. This can be encoded using the sets X and Y via

$$X := \{x = (x_{C_x}, x_{I_x}) \colon x_{I_x} \in \mathbb{Z}^{n_{x_I}}\}, \quad Y := \left\{y = (y_{C_y}, y_{I_y}) \colon y_{I_y} \in \mathbb{Z}^{n_{y_I}}\right\}.$$

Thus, n_{x_I} and n_{y_I} denote the number of integer variables in the upper- as well as the lower-level problem, respectively.

For the remainder of this section, we make the following assumptions.

Assumption 7.2. *The shared constraint set Ω is non-empty and compact and its projection Ω_x onto the x-space is non-empty.*

The goal now is to design a branch-and-bound method for problems of the type given in (7.3). To this end, let us first recap the main fathoming rules that we used in the classic branch-and-bound method (Algorithm 2) for linear bilevel problems. There, we fathomed nodes according to the following three rules:

Rule 1 The problem at the current node is infeasible.

Rule 2 The problem at the current node is feasible and has a solution with an optimal objective function value that is not smaller than the current incumbent, i.e., it is not smaller than the optimal objective function value of the best solution found so far.

Rule 3 The problem at the current node is feasible w.r.t. all complementarity constraints.

As it is done in classic branch-and-bound for single-level mixed-integer linear problems, we now do not branch on complementarity constraints anymore but branch on integer variables. Thus, Rule 3 translates into ...

Rule 3 The problem at the current node is feasible w.r.t. all integrality constraints.

It turns out that only Rule 1 can be used in its original form as a fathoming rule within a branch-and-bound method for mixed-integer linear bilevel problems. Rule 2 can be adapted (which we will do later on) and Rule 3 cannot be applied at all in the context of Problem (7.3).

To see this, we revisit the example studied in Section 7.1.

Example 7.3 (The example by Moore and Bard—revisited)**.** *We have seen that the bilevel solution is given by the point $(x^*, y^*) = (2, 2)$, which leads to the optimal objective function value $F(x^*, y^*) = -22$.*[1] *The optimal solution of the problem in which we relax all integrality conditions is the point $(x, y) = (8, 1)$. Note that this point is even integer- and bilevel-feasible. The corresponding objective function value, however, is $F(x, y) = -18$, which is worse than the optimal objective function value.*

[1] We use the more general notation F for the upper-level objective function from here on again for the ease of presentation.

This leads to the following two observations.

Observation 7.4. *The solution of the continuous "relaxation" of the mixed-integer linear bilevel problem does not provide a valid lower bound on the solution of the original problem. Indeed, neglecting the integrality constraints of a mixed-integer linear bilevel problem does* not *lead to a relaxation.*

Observation 7.5. *Solutions of the continuous relaxation of the mixed-integer linear bilevel problem that are feasible for the original bilevel problem cannot, in general, be fathomed.*

These two observations already render Rule 2 and Rule 3 invalid in general. The following example, which is also taken from Moore and Bard (1990), shows what goes wrong if Rule 3 is applied although it is invalid.

Example 7.6 (See Example 2 in Moore and Bard (1990))**.** *We consider the integer linear bilevel problem*

$$\begin{aligned} \max_{x,y} \quad & F(x,y) = -x - 2y \\ s.t. \quad & y \in S(x), \end{aligned}$$

where $S(x)$ denotes the set of optimal solutions of the x-parameterized integer linear problem

$$\begin{aligned} \max_{y} \quad & f(x,y) = y \\ s.t. \quad & -x + 2.5y \leq 3.75, \\ & x + 2.5y \geq 3.75, \\ & 2.5x + y \leq 8.75, \\ & x, y \geq 0, \\ & x, y \in \mathbb{Z}. \end{aligned}$$

An illustration of the problem is given in Figure 7.3. The shared constraint set contains three integer-feasible points: $(2,1)$, $(2,2)$, and $(3,1)$. If the leader chooses $x = 2$, the follower chooses $y = 2$, leading to $F = -6$. If the leader decides for $x = 3$, the follower optimally reacts with $y = 1$, leading to an objective function value of $F = -5$. Thus, $(x^*, y^*) = (3,1)$ is the optimal solution with $F^* = -5$.

Let us now consider what a classic depth-first search branch-and-bound method would look like if we (as usual) branch on fractional integer variables and if relaxations are obtained by relaxing integrality restrictions. A possible branch-and-bound tree is given in Figure 7.4. The root node's relaxation (node 0) has the solution $(x,y) = (0,1.5)$ with $F = -3$. Adding then the constraint $y \geq 2$ leads to node 1, which has the solution $(x,y) = (1.25,2)$ with $F = -5.25$. Further branching, fathoming due to infeasibility, and

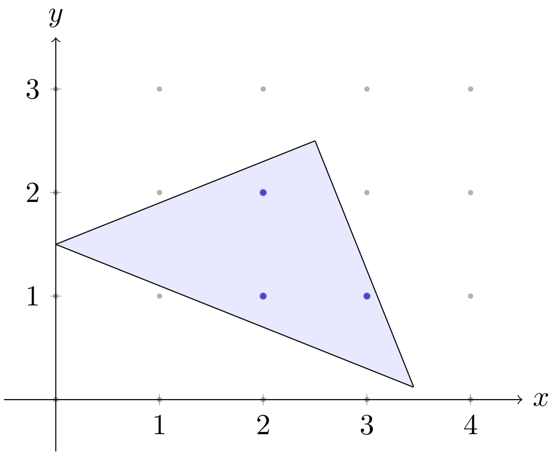

Figure 7.3: The bilevel problem of Example 7.6; taken from Moore and Bard (1990)

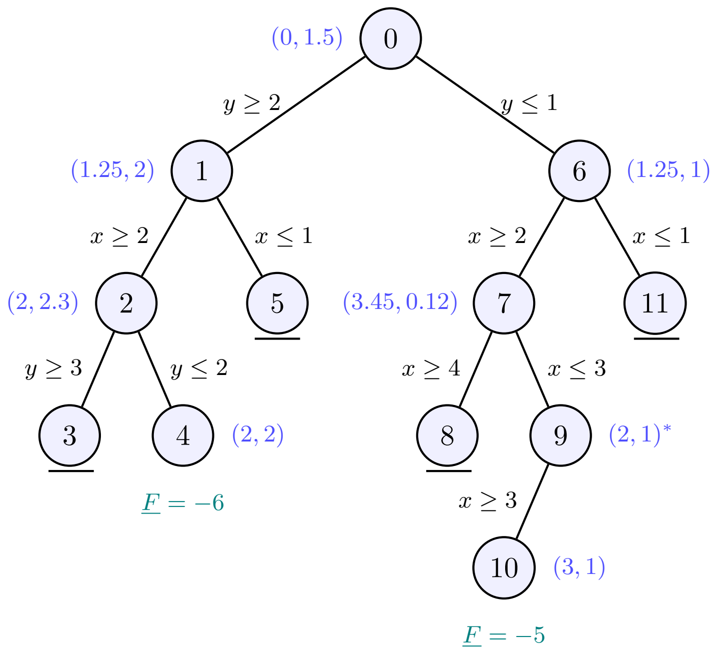

Figure 7.4: A branch-and-bound tree for Example 7.6; taken and adapted from Bard (1998). The tuples of the form (x, y) at the nodes denote the solutions of the problems at the nodes. The solution marked with a star at node 9 is integer-feasible but not bilevel-feasible.

backtracking[2] *leads us to arrive at node 9, which has the solution* $(x, y) = (2, 1)$ *with* $F = -4$. *The point lies in the shared constraint set and is integer-feasible. However, it is not bilevel feasible since* $y = 1$ *is not the optimal reaction to* $x = 2$. *Moreover, the objective function value at node 9 is* $F = -4$ *and this is not a valid lower bound. If we apply Rule 3 here, we would fathom this node and will* not *find the optimal solution* $(x, y) = (3, 1)$. *Instead, to find the optimal solution, one has to restrict the variable* x *further. One can also show that selecting* y *as the branching variable in node 7 does not help for finding the optimal solution.*

Thus, we can make the third main observation.

Observation 7.7. *An integer-feasible solution found at a node that contains branching restrictions on the follower variables cannot, in general, be used to fathom this node.*

For what follows, we need some more notation. First, we use I_x and I_y to denote the index sets of integer variables of the leader and the follower, respectively. Moreover, let U^x and U^y be the $|I_x|$ as well as $|I_y|$ dimensional vectors of upper bounds for the integer variables of the leader and of the follower, respectively. If an integer variable is not bounded from above in the original problem, the corresponding entry in U^x or U^y is set to ∞. Moreover, we assume that the initial lower bounds of all integers variables are 0, which is w.l.o.g. and which can be encoded using the sets X and Y of the original problem formulation.

The problem at node k of the branch-and-bound tree is defined by the variable bound sets

$$\begin{aligned} X_k &:= \left\{ (\underline{x}^k, \bar{x}^k) \colon 0 \le \underline{x}_j^k \le x_j \le \bar{x}_j^k \le U_j^x \text{ for } j \in I_x \right\}, \\ Y_k &:= \left\{ (\underline{y}^k, \bar{y}^k) \colon 0 \le \underline{y}_j^k \le y_j \le \bar{y}_j^k \le U_j^y \text{ for } j \in I_y \right\}. \end{aligned}$$

The notation Y_0 is used to indicate that no other bounds than the original ones are imposed on the follower's integer variables. Note further that for a node k along the path from the root to node l, the problem associated to node l is derived from the problem of the node k by additionally imposing bounds on the integer variables, i.e.,

$$X_l \subseteq X_k, \quad Y_l \subseteq Y_k$$

holds, which means that

$$\underline{x}^k \le \underline{x}^l, \quad \underline{y}^k \le \underline{y}^l$$

[2]"Backtracking" happens in a branch-and-bound method if, after solving a node's problem, no new sub-problems are generated.

as well as

$$\bar{x}^k \geq \bar{x}^l, \quad \bar{y}^k \geq \bar{y}^l$$

holds. Furthermore, the sets

$$R_k^x := \left\{ j \in I_x : \underline{x}_j^k > 0 \text{ or } \bar{x}_j^k < U_j^x \right\}$$

and

$$R_k^y := \left\{ j \in I_y : \underline{y}_j^k > 0 \text{ or } \bar{y}_j^k < U_j^y \right\}$$

denote that sets of integer variables on which additional bounds are imposed (due to branching).

Finally, for the later reference we define the problem

$$\begin{aligned} \min_{x \geq 0,\, y} \quad & c_x^\top x + c_y^\top y && (7.4a)\\ \text{s.t.} \quad & Ax \geq a, && (7.4b)\\ & \text{bounds in } X_k, \text{ i.e., } \underline{x}_j^k \leq x_j \leq \bar{x}_j^k \text{ for } j \in I_x, && (7.4c)\\ & y \in S_k(x) && (7.4d) \end{aligned}$$

with the lower-level problem

$$\begin{aligned} \min_{y \geq 0} \quad & d^\top y\\ \text{s.t.} \quad & Cx + Dy \geq b,\\ & \text{bounds in } Y_k, \text{ i.e., } \underline{y}_j^k \leq y_j \leq \bar{y}_j^k \text{ for } j \in I_y \end{aligned}$$

as the bilevel problem at node k in which the integrality constraints are relaxed. Its optimal objective function value is denoted with F_k^{cont}. The corresponding high-point relaxation is given by

$$\begin{aligned} \min_{x \geq 0,\, y \geq 0} \quad & c_x^\top x + c_y^\top y && (7.5a)\\ \text{s.t.} \quad & Ax \geq a, && (7.5b)\\ & \text{bounds in } X_k, \text{ i.e., } \underline{x}_j^k \leq x_j \leq \bar{x}_j^k \text{ for } j \in I_x, && (7.5c)\\ & Cx + Dy \geq b, && (7.5d)\\ & \text{bounds in } Y_k, \text{ i.e., } \underline{y}_j^k \leq y_j \leq \bar{y}_j^k \text{ for } j \in I_y. && (7.5e) \end{aligned}$$

Its optimal objective function value is denoted with F_k^{hpr}.

With the observations made so far and the previously collected notation, we are now able to state and prove some bounding theorems.

Theorem 7.8 (See Theorem 1 in Moore and Bard (1990)). *Consider the sub-problem at node k with the bounds given by X_k and Y_0. Let (x^k, y^k) be the global optimal solution of the continuous high-point relaxation (7.5). Then, $F_k^{hpr} = F(x^k, y^k)$ is a lower bound on the global optimal solution of the mixed-integer linear bilevel problem at node k.*

Proof. Consider any successor node l of node k in the branch-and-bound tree, i.e., $X_l \subseteq X_k$ and $Y_l \subseteq Y_0$ holds. Let (x^l, y^l) be a global optimal solution of the mixed-integer linear bilevel problem associated with node l. Assume now that $F(x^l, y^l) < F_k^{\text{hpr}}$ holds. This directly leads to a contradiction since (x^l, y^l) is also a feasible point of the continuous high-point relaxation at node k. □

The last theorem states the following. We can use the bound obtained by the continuous high-point relaxation of the mixed-integer linear bilevel problem at node k to fathom this node if no further restriction (compared to the original ones) have been imposed by branching on the integer variables of the follower, i.e., if $R_k^y = \emptyset$.

The next theorem indicates a situation in which the value F_k^{hpr} can be used as a valid lower bound if $R_k^y \neq \emptyset$.

Theorem 7.9. *Consider the sub-problem at node k with the bounds given by X_k and Y_k. Let (x^k, y^k) be the global optimal solution of the continuous high-point relaxation (7.5). Then, $F_k^{hpr} = F(x^k, y^k)$ is a lower bound on the global optimal solution of the mixed-integer linear bilevel problem at node k if $\underline{y}_j^k < y_j^k < \bar{y}_j^k$ holds for all $j \in R_k^y$.*

This means that the solution of the continuous high-point relaxation of the mixed-integer linear bilevel problem at node k can serve as a valid lower bound if the optimal integer variables of the follower at node k are not active w.r.t. their bounds imposed due to branching. Note that this is, of course, a rather strong condition, which is, for instance, violated at node 9 in Example 7.6.

Proof. Let again (x^l, y^l) be the solution of the mixed-integer linear bilevel problem associated with node l, which is a successor node of node k, i.e., $X_l \subseteq X_k$ and $Y_l \subseteq Y_k$ holds. Assume again that $F(x^l, y^l) < F_k^{\text{hpr}}$ holds. This directly implies that (x^l, y^l) cannot be feasible for the continuous high-point relaxation of the mixed-integer linear bilevel problem at node k. We consider the points (x', y') of the convex combination of (x^l, y^l) and (x^k, y^k), i.e.,

$$(x', y') = \lambda(x^k, y^k) + (1 - \lambda)(x^l, y^l)$$

holds for some $\lambda \in [0, 1]$. It holds

$$F(x', y') = \lambda F(x^k, y^k) + (1 - \lambda)F(x^l, y^l)$$

since F is linear. Using $F(x^l, y^l) < F_k^{\text{hpr}}$ we obtain

$$\begin{aligned} F(x', y') &= \lambda F(x^k, y^k) + (1 - \lambda)F(x^l, y^l) \\ &< \lambda F(x^k, y^k) + (1 - \lambda)F(x^k, y^k) \\ &= F(x^k, y^k) \end{aligned}$$

for $\lambda > 0$. This, however, contradicts the optimality of (x^k, y^k) since for sufficiently small λ, (x', y') is feasible for the continuous high-point relaxation of the mixed-integer linear bilevel problem at node k. □

As already stated, the assumptions of the last theorem are rather strict. The next corollary gives "a bit" of an improvement.

Corollary 7.10. *Consider the sub-problem at node k with the bounds given by X_k and Y_k. Let (x^k, y^k) be the global optimal solution of the continuous high-point relaxation (7.5). Then, $F_k^{hpr} = F(x^k, y^k)$ is a lower bound on the global optimal solution of the mixed-integer linear bilevel problem at node k if all restrictions in Y_k are relaxed.*

Proof. Relaxing all restrictions in Y_k is equivalent to replacing Y_k with Y_0 and, thus, Theorem 7.8 applies. □

The theoretical results so far reveal that the bounding step for mixed-integer linear bilevel problems need to be restricted significantly when compared with the bounding step for single-level mixed-integer problems or for linear bilevel problems. However, no stronger bounding schemes are known and, due to the observations and examples above, are also not to be expected.

Taking all the insights together leads to the branch-and-bound method for mixed-integer linear bilevel problems as stated in Algorithm 5.

Due to the construction of the algorithm and the insights we obtained before, we get the following results.

Proposition 7.11. *If all follower variables are integer, Algorithm 5 finds the global optimal solution of the mixed-integer linear bilevel problem (7.3).*

Proposition 7.12. *Assume that an optimum exists for the mixed-integer linear bilevel problem (7.3) and that all follower variables are continuous. If the fathoming rules 2 and 3 are used, Algorithm 5 always terminates with the global optimal solution.*

Algorithm 5 Branch-and-Bound for MILP-MILP Bilevel Problems

1: Set $k = 0$ and initialize X_k and Y_k with the bounds of the original mixed-integer linear bilevel problem. Set $R_k^x = \emptyset$, $R_k^y = \emptyset$, and $F^* = \infty$.
2: Solve the continuous high-point relaxation (7.5). If this problem is infeasible go to Step 7. Otherwise, let F_k^{hpr} be the optimal objective function value. If $F_k^{\text{hpr}} \geq F^*$ holds, go to Step 7 as well.
3: Solve the continuous relaxation (7.4). If this problem is infeasible, go to Step 7. Otherwise, denote the solution as (x^k, y^k).
4: If (x^k, y^k) is integer-feasible, go to Step 5. Otherwise, select a fractional leader variable index $j \in I_x$ or a fractional follower variable index $j \in I_y$ and place a new bound on the selected variable. Set $k \leftarrow k + 1$ and update X_k or Y_k as well as R_k^x or R_k^y accordingly. Go to Step 2.
5: Fix $x = x^k$ and solve the follower's problem to obtain the overall bilevel feasible point $(x^k, \hat{y}^k)$. Compute $F(x^k, \hat{y}^k)$ and update $F^* = \min\{F^*, F(x^k, \hat{y}^k)\}$.
6: If $\underline{x}_j^k = \bar{x}_j^k$ for all $j \in I_x$ and if $\underline{y}_j^k = \bar{y}_j^k$ for all $j \in I_y$ holds, go to Step 7. Otherwise, select an integer variable $j \in I_x$ with $\underline{x}_j^k < \bar{x}_j^k$ or a $j \in I_y$ with $\underline{y}_j^k < \bar{y}_j^k$ and place a new bound on it. Set $k \leftarrow k + 1$ and update X_k or Y_k as well as R_k^x or R_k^y accordingly. Go to Step 2.
7: If no open node exists, go to Step 8. Otherwise, branch on the lastly added open node, set $k \leftarrow k + 1$, and update X_k or Y_k as well as R_k^x or R_k^y accordingly. Go to Step 2.
8: If $F^* = \infty$, the original mixed-integer linear bilevel problem is infeasible. Otherwise, F^* is the global optimal objective function value.

8

What You Should Know Now!

1. In what situations do bilevel problems occur in general?
2. What is a pricing problem?
3. What is the toll setting problem?
4. In what situations do bilevel problems occur in energy markets?
5. What is the relation between bilevel models and critical infrastructure defense?
6. What are interdiction problems about?
7. How is a bilevel optimization problem defined formally?
8. What are coupling constraints?
9. What are linking variables?
10. What is the optimal value function?
11. How does the optimal-value-function reformulation look like?
12. Why do we have to repeat the lower-level constraints in the optimal-value-function reformulation although they are also part of the right-hand side of the optimal value function constraint?
13. What is the shared constraint set?
14. What is the bilevel feasible set/the inducible region?
15. What is the high-point relaxation? Is it really a relaxation? If yes, a relaxation of what and why?

16. How can a pricing problem by formally modeled as a bilevel optimization problem?
17. How can we formally model the knapsack interdiction problem as a bilevel problem?
18. What nasty properties do we already encounter if we consider the easiest case of bilevel problems, i.e., LP-LP bilevel problems?
19. What is an LP in standard form?
20. What is the feasible set?
21. When do we call the LP bounded?
22. What do you know about the solvability of LPs?
23. What is the dual LP?
24. What is the statement of the weak duality theorem of linear optimization?
25. What is the statement of the strong duality theorem of linear optimization?
26. What is the statement of the complementarity slackness theorem of linear optimization?
27. How can we use the strong duality theorem to re-write an LP as a system of equalities and inequalities?
28. How can we use the complementarity slackness theorem to re-write an LP as a system of equalities and inequalities?
29. What is the "standard form" of an NLP?
30. How do we define a local minimizer?
31. How do we define a strict local minimizer?
32. How do we define a global minimizer?
33. How do we define a strict global minimizer?
34. What are active inequality constraints?
35. What is the ACQ?
36. What is the Lagrangian function of an NLP?
37. How do the KKT conditions look like?

38. What is a Lagrangian multiplier?
39. What is the relation between Lagrangian multipliers and dual variables?
40. What is the statement of the KKT theorem?
41. What is the relation between the KKT theorem and the complementarity slackness theorem?
42. What is the LICQ?
43. What does the KKT theorem under LICQ say?
44. What is a convex optimization problem?
45. What is the nice thing about convex optimization problems? Why are they so special?
46. What specific properties do equality and inequality constraints need to have so that the resulting feasible set is convex?
47. How is the CQ of Slater defined?
48. What is the geometric meaning of Slater's CQ?
49. How does the KKT theorem for convex problems look like?
50. What do you know about the relationship between KKT points and global minimizers in the case of convex optimization problems?
51. What is an MPCC?
52. Why is this different from "usual" NLPs?
53. What is the relation between MPCCs and optimality conditions of optimization problems?
54. How do we define an optimistic bilevel problem?
55. How do we define a pessimistic bilevel problem without coupling constraints?
56. Why do we need to distinguish between the optimistic and the pessimistic solution at all?
57. How do we define a pessimistic bilevel problem with coupling constraints?
58. Can you illustrate the differences between these concepts using an example?

59. What is nice about uniquely defined lower-level solutions?
60. What is a local minimum of a bilevel problem?
61. What is a global minimum of a bilevel problem?
62. How many single-level reformulations do you know?
63. How does the single-level reformulation using the optimal value function look like?
64. What is the problem (in general) with the single-level reformulation using the optimal value function?
65. How does an LP-LP bilevel problem look like?
66. Why did we omit the linear term in x in the lower-level objective function of the LP-LP bilevel problem?
67. Can you derive the KKT reformulation of the LP-LP bilevel problem?
68. What makes this KKT reformulation hard to solve?
69. How do we define Slater's CQ for the lower-level problem?
70. What is the relation between the global optimal solutions of the bilevel problem with a convex follower problem and the corresponding KKT reformulations? What are the assumptions that are required?
71. What is the relation between the local optimal solutions of the bilevel problem with a convex follower problem and the corresponding KKT reformulation?
72. Is the KKT reformulation of an LP-LP bilevel problem an LP again?
73. Which ones are the nonlinear constraints of the KKT reformulation of an LP-LP bilevel problem? Can we linearize these nonlinear constraints? If yes, how? What is the price that we have to pay for it?
74. What is the problem with the big-Ms?
75. How does the strong-duality-based single-level reformulation of an LP-LP bilevel problem look like? How is it derived?
76. What makes this strong-duality-based single-level reformulation hard to solve?
77. Why can't we linearize the nonlinearities of the strong-duality-based single-level reformulation similar to the case of the KKT reformulation?

78. What is the relation between the strong-duality-based single-level reformulation and the KKT reformulation of an LP-LP bilevel problem?

79. What is the IIC property?

80. Does the IIC property hold for bilevel optimization problems?

81. What do you know about the geometrical properties of LP-LP bilevel problems?

82. What is the geometry of the bilevel-feasible set of an LP-LP bilevel problem?

83. To what points can we restrict our search for optimal solutions of an LP-LP bilevel problem? What role does the high-point relaxation play here?

84. What do you know about the relation of LP-LP bilevel problems and single-level mixed-integer linear problems? What ingredient does your LP-LP bilevel problem need to have to establish this relation?

85. What is the main idea behind the Kth-best algorithm?

86. How is the Kth-best algorithm formally defined?

87. What are the crucial parts of the Kth-best algorithm?

88. What is the main idea behind the branch-and-bound method for LP-LP bilevel problems? What do we branch on and why?

89. How is the branch-and-bound method for LP-LP bilevel problems stated formally?

90. What is a relaxation?

91. What is the claim of the bounding lemma?

92. What is the claim of the branching lemma?

93. What theoretical statement do you know about the branch-and-bound method for LP-LP bilevel problems?

94. What is an alternating direction method (ADM)?

95. What is a partial minimum?

96. What is the general convergence result for ADMs and what assumptions are required?

97. What can we gain if we impose stronger assumptions?

98. What is the idea to come from ADM to the penalty ADM?
99. What is the convergence result for the PADM?
100. How do we apply PADM to LP-LP bilevel problems and why do we do it exactly like this?
101. What is the theoretical result that we get for LP-LP bilevel problems?
102. Can you state the general form of a mixed-integer linear bilevel problem?
103. What do you know about the hardness of these problems?
104. What are the insights from the example by Moore and Bard from 1990?
105. What do you know about the attainability of solutions for mixed-integer linear bilevel problems?
106. What branch-and-bound fathoming rules can we carry over from MILP to mixed-integer linear bilevel problems?
107. What is wrong with those that we cannot carry over (directly)?
108. How does the branch-and-bound method for mixed-integer linear bilevel problems by Moore and Bard work?
109. What are the theoretical results about bounding and about the entire algorithm that you know?

Index

Bibliography

Ambrosius, M., V. Grimm, T. Kleinert, F. Liers, M. Schmidt, and G. Zöttl (2020). "Endogenous Price Zones and Investment Incentives in Electricity Markets: An Application of Multilevel Optimization with Graph Partitioning." In: *Energy Economics* 92, p. 104879. DOI: 10.1016/j.eneco.2020.104879.

Anandalingam, G. and D. White (1990). "A solution method for the linear static Stackelberg problem using penalty functions." In: *IEEE Transactions on Automatic Control* 35.10, pp. 1170–1173. DOI: 10.1109/9.58565.

Anandalingam, G. and T. Friesz (1992). "Hierarchical optimization: An introduction." In: *Annals of Operations Research* 34.1, pp. 1–11. DOI: 10.1007/BF02098169.

Audet, C., J. Haddad, and G. Savard (2006). "A note on the definition of a linear bilevel programming solution." In: *Applied Mathematics and Computation* 181.1, pp. 351–355. DOI: 10.1016/j.amc.2006.01.043.

Audet, C., J. Haddad, and G. Savard (2007). "Disjunctive Cuts for Continuous Linear Bilevel Programming." In: *Optimization Letters* 1.3, pp. 259–267. DOI: 10.1007/s11590-006-0024-3.

Audet, C., P. Hansen, B. Jaumard, and G. Savard (1997). "Links between linear bilevel and mixed 0–1 programming problems." In: *Journal of Optimization Theory and Applications* 93.2, pp. 273–300. DOI: 10.1023/A:1022645805569.

Audet, C., G. Savard, and W. Zghal (2007). "New Branch-and-Cut Algorithm for Bilevel Linear Programming." In: *Journal of Optimization Theory and Applications* 134.2, pp. 353–370. DOI: 10.1007/s10957-007-9263-4.

Bard, J. F. (1988). "Convex two-level optimization." In: *Mathematical Programming* 40.1, pp. 15–27. DOI: 10.1007/BF01580720.

Bard, J. F. (1998). *Practical bilevel optimization: algorithms and applications.* Vol. 30. Springer Science & Business Media. DOI: 10.1007/978-1-4757-2836-1.

Bard, J. F. and J. T. Moore (1990). "A branch and bound algorithm for the bilevel programming problem." In: *SIAM Journal on Scientific and Statistical Computing* 11.2, pp. 281–292. DOI: 10.1137/0911017.

Ben-Ayed, O. (1993). "Bilevel linear programming." In: *Computers & Operations Research* 20.5, pp. 485–501. DOI: 10.1016/0305-0548(93)90013-9.

Bialas, W. F. and M. H. Karwan (1978). *Multilevel linear programming.*

Bialas, W. F. and M. H. Karwan (1984). "Two-level linear programming." In: *Management Science* 30.8, pp. 1004–1020. DOI: 10.1287/mnsc.30.8.1004.

Bolusani, S. and T. K. Ralphs (2020). *A Framework for Generalized Benders' Decomposition and Its Application to Multilevel Optimization.* Tech. rep. COR@L Technical Report 20T-004. URL: http://www.optimization-online.org/DB_HTML/2020/04/7755.html.

Boyd, S., N. Parikh, E. Chu, B. Peleato, J. Eckstein, et al. (2011). "Distributed optimization and statistical learning via the alternating direction method of multipliers." In: *Foundations and Trends in Machine learning* 3.1, pp. 1–122. DOI: 10.1561/2200000016.

Bracken, J. and J. T. McGill (1973). "Mathematical programs with optimization problems in the constraints." In: *Operations Research* 21.1, pp. 37–44. DOI: 10.1287/opre.21.1.37.

Candler, W. and R. Norton (1977). *Multi-level Programming.* Discussion Papers, Development Research Center, International Bank for Reconstruction and Development. World Bank. URL: https://books.google.de/books?id=TicmAQAAMAAJ.

Caprara, A., M. Carvalho, A. Lodi, and G. J. Woeginger (2016). "Bilevel Knapsack with Interdiction Constraints." In: *INFORMS Journal on Computing* 28.2, pp. 319–333. DOI: 10.1287/ijoc.2015.0676.

Clark, F. E. (1961). "Remark on the Constraint Sets in Linear Programming." In: *The American Mathematical Monthly* 68.4, pp. 351–352. DOI: 10.2307/2311583.

Colson, B., P. Marcotte, and G. Savard (2005). "Bilevel programming: A survey." In: *4OR* 3.2, pp. 87–107. DOI: 10.1007/s10288-005-0071-0.

Colson, B., P. Marcotte, and G. Savard (2007). "An overview of bilevel optimization." In: *Annals of Operations Research* 153.1, pp. 235–256. DOI: 10.1007/s10479-007-0176-2.

Dempe, S. (2002). *Foundations of Bilevel Programming.* Springer. DOI: 10.1007/b101970.

Dempe, S. (2020). "Bilevel Optimization: Theory, Algorithms, Applications and a Bibliography." In: *Bilevel Optimization: Advances and Next Challenges.* Ed. by S. Dempe and A. Zemkoho. Springer International Publishing, pp. 581–672. DOI: 10.1007/978-3-030-52119-6_20.

Dempe, S. and J. Dutta (2012). "Is bilevel programming a special case of a mathematical program with complementarity constraints?" In: *Mathematical Programming* 131.1-2, pp. 37–48. DOI: 10.1007/s10107-010-0342-1.

Dempe, S., V. Kalashnikov, G. A. Pérez-Valdés, and N. Kalashnykova (2015). *Bilevel Programming Problems.* Springer. DOI: 10.1007/978-3-662-45827-3.

DeNegre, S. T. and T. K. Ralphs (2009). "A branch-and-cut algorithm for integer bilevel linear programs." In: *Operations research and cyber-infrastructure*. Springer, pp. 65–78. DOI: 10.1007/978-0-387-88843-9_4.

Edmunds, T. A. and J. F. Bard (1991). "Algorithms for nonlinear bilevel mathematical programs." In: *IEEE Transactions on Systems, Man, and Cybernetics* 21.1, pp. 83–89. DOI: 10.1109/21.101139.

Fortuny-Amat, J. and B. McCarl (1981). "A Representation and Economic Interpretation of a Two-Level Programming Problem." In: *The Journal of the Operational Research Society* 32.9, pp. 783–792. DOI: 10.1057/jors.1981.156.

Geißler, B., A. Morsi, L. Schewe, and M. Schmidt (2015). "Solving power-constrained gas transportation problems using an MIP-based alternating direction method." In: *Computers & Chemical Engineering* 82, pp. 303–317. ISSN: 0098-1354. DOI: 10.1016/j.compchemeng.2015.07.005.

Geißler, B., A. Morsi, L. Schewe, and M. Schmidt (2017). "Penalty Alternating Direction Methods for Mixed-Integer Optimization: A New View on Feasibility Pumps." In: *SIAM Journal on Optimization* 27.3, pp. 1611–1636. DOI: 10.1137/16M1069687.

Geißler, B., A. Morsi, L. Schewe, and M. Schmidt (2018). "Solving Highly Detailed Gas Transport MINLPs: Block Separability and Penalty Alternating Direction Methods." In: *INFORMS Journal on Computing* 30.2, pp. 309–323. ISSN: 1091-9856. DOI: 10.1287/ijoc.2017.0780.

Gorski, J., F. Pfeuffer, and K. Klamroth (2007). "Biconvex sets and optimization with biconvex functions: a survey and extensions." In: *Mathematical Methods of Operations Research* 66.3, pp. 373–407. ISSN: 1432-2994. DOI: 10.1007/s00186-007-0161-1.

Grimm, V., T. Kleinert, F. Liers, M. Schmidt, and G. Zöttl (2019). "Optimal price zones of electricity markets: a mixed-integer multilevel model and global solution approaches." In: *Optimization Methods and Software* 34.2, pp. 406–436. DOI: 10.1080/10556788.2017.1401069.

Grimm, V., A. Martin, M. Schmidt, M. Weibelzahl, and G. Zöttl (2016). "Transmission and Generation Investment in Electricity Markets: The Effects of Market Splitting and Network Fee Regimes." In: *European Journal of Operational Research* 254.2, pp. 493–509. DOI: 10.1016/j.ejor.2016.03.044.

Grimm, V., L. Schewe, M. Schmidt, and G. Zöttl (2019). "A Multilevel Model of the European Entry-Exit Gas Market." In: *Mathematical Methods of Operations Research* 89.2, pp. 223–255. ISSN: 1432-5217. DOI: 10.1007/s00186-018-0647-z.

Gurobi (2021). *The fastest solver - Gurobi*. last accessed April 29, 2021. URL: https://www.gurobi.com.

Hansen, P., B. Jaumard, and G. Savard (1992). "New branch-and-bound rules for linear bilevel programming." In: *SIAM Journal on scientific and Statistical Computing* 13.5, pp. 1194–1217. DOI: 10.1137/0913069.

IBM (2021). *CPLEX Optimizer*. last accessed April 29, 2021. URL: https://www.ibm.com/de-de/analytics/cplex-optimizer.

Jeroslow, R. G. (1985). "The polynomial hierarchy and a simple model for competitive analysis." In: *Mathematical Programming* 32.2, pp. 146–164. DOI: 10.1007/BF01586088.

Kleinert, T. (2021). "Algorithms for Mixed-Integer Bilevel Problems with Convex Followers." PhD thesis. URL: https://opus4.kobv.de/opus4-trr154/frontdoor/index/index/docId/383.

Kleinert, T., M. Labbé, I. Ljubić, and M. Schmidt (2021). *A Survey on Mixed-Integer Programming Techniques in Bilevel Optimization*. Tech. rep. URL: http://www.optimization-online.org/DB_HTML/2021/01/8187.html.

Kleinert, T., M. Labbé, F. Plein, and M. Schmidt (2020). "There's No Free Lunch: On the Hardness of Choosing a Correct Big-M in Bilevel Optimization." In: *Operations Research* 68.6, pp. 1716–1721. DOI: 10.1287/opre.2019.1944.

Kleinert, T., J. Manns, M. Schmidt, and D. Weninger (2021). *Presolving Linear Bilevel Optimization Problems*. Tech. rep. URL: http://www.optimization-online.org/DB_HTML/2021/03/8286.html.

Kleinert, T. and M. Schmidt (2019a). "Computing Feasible Points of Bilevel Problems with a Penalty Alternating Direction Method." In: *INFORMS Journal on Computing*. DOI: 10.1287/ijoc.2019.0945. Forthcoming.

Kleinert, T. and M. Schmidt (2019b). "Global Optimization of Multilevel Electricity Market Models Including Network Design and Graph Partitioning." In: *Discrete Optimization* 33, pp. 43–69. ISSN: 1572-5286. DOI: 10.1016/j.disopt.2019.02.002.

Kolstad, C. D. (1985). *A review of the literature on bi-level mathematical programming*. Tech. rep. Los Alamos National Laboratory Los Alamos, NM. URL: https://fas.org/sgp/othergov/doe/lanl/lib-www/la-pubs/00318870.pdf.

Köppe, M., M. Queyranne, and C. T. Ryan (2010). "Parametric Integer Programming Algorithm for Bilevel Mixed Integer Programs." In: *Journal of Optimization Theory and Applications* 146.1, pp. 137–150. DOI: 10.1007/s10957-010-9668-3.

Labbé, M., P. Marcotte, and G. Savard (1998). "A bilevel model of taxation and its application to optimal highway pricing." In: *Management Science* 44.12-part-1, pp. 1608–1622. DOI: 10.1287/mnsc.44.12.1608.

Land, A. H. and A. G. Doig (1960). "An Automatic Method of Solving Discrete Programming Problems." In: *Econometrica* 28.3, pp. 497–520. ISSN: 00129682. URL: http://www.jstor.org/stable/1910129.

Luo, Z.-Q., J.-S. Pang, and D. Ralph (1996). *Mathematical programs with equilibrium constraints*. Cambridge University Press. DOI: 10.1017/CBO9780511983658.

Macal, C. M. and A. P. Hurter (1997). "Dependence of bilevel mathematical programs on irrelevant constraints." In: *Computers & Operations Research* 24.12, pp. 1129–1140. DOI: 10.1016/S0305-0548(97)00025-7.

Manns, J. (2020). "Presolve of Linear Bilevel Programs." MA thesis. Friedrich-Alexander-Universität Erlangen-Nürnberg. URL: https://opus4.kobv.de/opus4-trr154/frontdoor/index/index/docId/320.

Mas-Colell, A., M. D. Whinston, J. R. Green, et al. (1995). *Microeconomic Theory*. Vol. 1. Oxford university press New York.

Mersha, A. G. and S. Dempe (2006). "Linear bilevel programming with upper level constraints depending on the lower level solution." In: *Applied Mathematics and Computation* 180.1, pp. 247–254. DOI: 10.1016/j.amc.2005.11.134.

Moore, J. T. and J. F. Bard (1990). "The mixed integer linear bilevel programming problem." In: *Operations Research* 38.5, pp. 911–921. DOI: 10.1287/opre.38.5.911.

Pineda, S., H. Bylling, and J. Morales (2018). "Efficiently solving linear bilevel programming problems using off-the-shelf optimization software." In: *Optimization and Engineering* 19.1, pp. 187–211. DOI: 10.1007/s11081-017-9369-y.

Pineda, S. and J. M. Morales (2019). "Solving Linear Bilevel Problems Using Big-Ms: Not All That Glitters Is Gold." In: *IEEE Transactions on Power Systems*. DOI: 10.1109/TPWRS.2019.2892607.

Savard, G. and J. Gauvin (1994). "The steepest descent direction for the nonlinear bilevel programming problem." In: *Operations Research Letters* 15.5, pp. 265–272. DOI: 10.1016/0167-6377(94)90086-8.

Schewe, L., M. Schmidt, and J. Thürauf (2020). *Global Optimization for the Multilevel European Gas Market System with Nonlinear Flow Models on Trees*. Tech. rep. URL: http://www.optimization-online.org/DB_HTML/2020/08/7973.html.

Tahernejad, S., T. K. Ralphs, and S. T. DeNegre (2020). "A Branch-and-Cut Algorithm for Mixed Integer Bilevel Linear Optimization Problems and Its Implementation." In: *Mathematical Programming Computation* 12, pp. 529–568. DOI: 10.1007/s12532-020-00183-6.

Vicente, L. and P. H. Calamai (1994). "Bilevel and multilevel programming: A bibliography review." In: *Journal of Global optimization* 5.3, pp. 291–306. DOI: 10.1007/BF01096458.

Vicente, L., G. Savard, and J. Júdice (1996). "Discrete Linear Bilevel Programming Problem." In: *Journal of Optimization Theory and Applications* 89.3, pp. 597–614. DOI: 10.1007/BF02275351.

Vicente, L., G. Savard, and J. Júdice (1994). "Descent approaches for quadratic bilevel programming." In: *Journal of Optimization Theory and Applications* 81.2, pp. 379–399. DOI: 10.1007/BF02191670.

von Stackelberg, H. (1934). *Marktform und Gleichgewicht*. Springer.

von Stackelberg, H. (1952). *Theory of the market economy.* Oxford University Press.

Wen, U.-P. and S.-T. Hsu (1991). "Linear Bi-Level Programming Problems – A Review." In: *The Journal of the Operational Research Society* 42.2, pp. 125–133. URL: http://www.jstor.org/stable/2583177.

Wendell, R. E. and A. P. Hurter Jr. (1976). "Minimization of a non-separable objective function subject to disjoint constraints." In: *Operations Research* 24.4, pp. 643–657. DOI: 10.1287/opre.24.4.643.

Williams, A. (1970). "Boundedness relations for linear constraint sets." In: *Linear Algebra and its Applications* 3.2, pp. 129–141. DOI: 10.1016/0024-3795(70)90009-1.

Ye, J. J. and D. L. Zhu (1995). "Optimality conditions for bilevel programming problems." In: *Optimization* 33.1, pp. 9–27. DOI: 10.1080/02331939508844060.