On a Computationally Ill-Behaved Bilevel Problem with a Continuous and Nonconvex Lower Level

Yasmine Beck, Daniel Bienstock, Martin Schmidt, Johannes Thürauf

ABSTRACT. It is well known that bilevel optimization problems are hard to solve both in theory and practice. In this paper, we highlight a further computational difficulty when it comes to solving bilevel problems with continuous but nonconvex lower levels. Even if the lower-level problem is solved to ε-feasibility regarding its nonlinear constraints for an arbitrarily small but positive ε, the obtained bilevel solution as well as its objective value may be arbitrarily far away from the actual bilevel solution and its actual objective value. This result even holds for bilevel problems for which the nonconvex lower level is uniquely solvable, for which the strict complementarity condition holds, for which the feasible set is convex, and for which Slater's constraint qualification is satisfied for all feasible upper-level decisions. Since the consideration of ε-feasibility cannot be avoided when solving nonconvex problems to global optimality, our result shows that computational bilevel optimization with continuous and nonconvex lower levels needs to be done with great care. Finally, we illustrate that the nonlinearities in the lower level are the key reason for the observed bad behavior by showing that linear bilevel problems behave much better at least on the level of feasible solutions.

1. INTRODUCTION

Bilevel optimization problems are known to be notoriously hard to solve and this holds true both in theory and in practice. In theory, bilevel problems are strongly NP-hard even if all objective functions and constraints are linear and all variables are continuous; see Hansen et al. (1992). This, of course, is also reflected in computational practice since linear bilevel problems are inherently nonsmooth and nonconvex problems. Moreover, the single-level reformulations used to solve linear bilevel problems in practice are nonconvex, complementarity-constrained problems. Their linearization requires big-M parameters that are hard to obtain in general (Kleinert et al. 2020) and that often lead to numerically badly posed problems, which are hard to tackle even for state-of-the-art commercial solvers.

In the last years, algorithmic research on bilevel optimization focused on more and more complicated lower-level problems such as mixed-integer linear models (Fischetti et al. 2017, 2018), nonlinear but still convex models (Kleinert et al. 2021a), or problems in the lower level with uncertain data (Beck et al. 2023; Buchheim and Henke 2022; Burtscheidt and Claus 2020). When it comes to the situation of a lower-level problem with continuous nonlinearities there is not too much literature—in particular in comparison to the case in which the lower-level problem is convex; see, e.g., Kleniati and Adjiman (2011, 2014a,b, 2015), Mitsos (2010), Mitsos et al. (2008), Paulavičius and Adjiman (2020), Paulavičius et al. (2020), and Paulavičius et al. (2016). Due to the brevity of this article, we do not go into the details of

the literature but refer to the seminal book by Dempe (2002) as well as the recent survey by Kleinert et al. (2021b) for further discussions of the relevant literature.

There is one important difference when crossing the border from (mixed-integer) convex to (mixed-integer) nonconvex lower-level problems. The lower-level problem can, in general, not be solved to global optimality anymore in an exact sense in finite time since we need to exploit techniques such as spatial branching to tackle nonconvexities. These techniques only lead to finite algorithms for prescribed and strictly positive feasibility tolerances; see, e.g., Locatelli and Schoen (2013) for more detailed discussions. Note that this is in clear contrast to, e.g., linear optimization. Here, the simplex method is an exact method in the sense that, if applied using exact arithmetic, the method computes a global optimal solution without any errors; see, e.g., Applegate et al. (2007). The same applies to simplex-based branch-and-bound methods for solving mixed-integer linear optimization problems. Algorithms as such are not available for continuous but nonconvex problems. This means that, just due to algorithmic necessities, we cannot expect to get exact feasible solutions of the lower-level problem anymore when doing computations for continuous but nonconvex lower-level problems. Instead, we have to deal with ε-feasible solutions—at least for the nonlinear constraints of the lower-level problem.

The aim of this paper is to present an exemplary bilevel optimization problem with continuous variables and a nonconvex lower-level problem, where the latter algorithmic aspect leads to the following severe issue:

> *Even if the feasibility tolerance for the lower level can be made extremely small, the exact bilevel solution can be arbitrarily far away from the bilevel solution that one obtains for ε-feasibility in the lower level, which in particular can be superoptimal out of proportion to ε. The same is true for the optimal objective function values.*

The main idea for the construction of this exemplary bilevel problem is based on a constraint set presented first by Bienstock et al. (2021). We explicitly note here that this construction does not make use of large constraint coefficient ranges (all coefficients are 1) or arbitrarily large degrees of polynomials (we only use quadratic or linear terms). Moreover, when considered in an exact sense, (i) the example's lower-level problem is uniquely solvable, (ii) strict complementarity holds, (iii) its convex constraint set satisfies Slater's constraint qualification for all feasible upper-level decisions, (iv) the upper level does not contain coupling constraints, and (v) the overall problem has a unique global solution as well. Thus, the bilevel program does not look like a badly-modeled problem but is shown to behave very badly in a computational sense, i.e., if only ε-feasible points for the nonlinear constraints of the lower-level problem can be considered. We also show that the observed pathological behavior arises due to the nonlinearities as we show that linear bilevel problems behave much better at least on the level of feasible points.

The example is presented in Section 2 and discussed in an exact sense in Section 3. Afterward, the example is analyzed in Section 4 for the case of ε-feasibility of nonlinear constraints. Section 5 presents an analysis of the linear bilevel case. Our final conclusions are drawn in Section 6.

2. Problem Statement

Let us consider the bilevel problem

$$\max_{x \in \mathbb{R}^2} \quad F(x, y) = x_1 - 2y_{n+1} + y_{n+2} \tag{1a}$$

$$\text{s.t.} \quad (x_1, x_2) \in [\underline{x}_1, \bar{x}_1] \times [\underline{x}_2, \bar{x}_2], \tag{1b}$$

$$y \in S(x), \tag{1c}$$

where $\underline{x}, \bar{x} \in \mathbb{R}^2$ with $1 \leq \underline{x}_i < \bar{x}_i$, $i \in \{1, 2\}$, denote lower and upper bounds on the variables x. Here, $S(x)$ is the set of optimal solutions of the x-parameterized problem

$$\max_{y \in \mathbb{R}^{n+2}} \quad f(x, y) = y_1 - y_n \left(x_1 + x_2 - y_{n+1} - y_{n+2}\right) \tag{2a}$$

$$\text{s.t.} \quad y_1 + y_n = \frac{1}{2}, \tag{2b}$$

$$y_i^2 \leq y_{i+1}, \quad i \in \{1, \ldots, n-1\}, \tag{2c}$$

$$y_i \geq 0, \quad i \in \{1, \ldots, n\}, \tag{2d}$$

$$y_{n+1} \in [0, x_1], \tag{2e}$$

$$y_{n+2} \in [-x_2, x_2]. \tag{2f}$$

We refer to Problem (1) as the upper-level (or the leader's) problem and to Problem (2) as the lower-level (or the follower's) problem. Let us point out that the lower-level constraints (2b) and (2c) together with $y_1 \geq 0$ have already been considered in Bienstock et al. (2021) in the context of approximately feasible solutions for single-level optimization problems. Let us further emphasize that the number of variables and constraints of the lower-level problem is linear in n.

Problem (1) is a linear problem in both the leader's and the follower's variables. The only constraints that occur in this problem are variable bounds for the leader's variables x. In particular, there are no upper-level constraints that explicitly depend on the follower's variables y, i.e., there are no coupling constraints.

Moreover, the feasible set of the lower-level problem is bounded due to the following. From (2d) and (2b), we obtain $0 \leq y_1 \leq 1/2$ as well as $0 \leq y_n \leq 1/2$ for any feasible follower's decision y. Using Constraints (2c), we further obtain $0 \leq y_i \leq 1$ for all $i \in \{1, \ldots, n\}$. Finally, we have $0 \leq y_{n+1} \leq \bar{x}_1$ as well as $-\bar{x}_2 \leq y_{n+2} \leq \bar{x}_2$ by (2e) and (2f) because the leader's variables x are bounded. Since all finitely many lower-level constraints are continuous, the feasible set of the follower's problem is compact. In addition to the compactness, the feasible set of the lower-level problem (2) is non-empty for every feasible leader's decision $(x_1, x_2) \in [\underline{x}_1, \bar{x}_1] \times [\underline{x}_2, \bar{x}_2]$. For instance, the point

$$y_i = \frac{i}{2^{2^n}}, \ i \in \{1, \ldots, n-1\}, \quad y_n = \frac{1}{2} - \frac{1}{2^{2^n}}, \quad y_{n+1} = \frac{1}{2}, \quad \text{and} \quad y_{n+2} = 0$$

is strictly feasible w.r.t. the inequality constraints (2d), (2e) as well as (2f) and feasible w.r.t. the equality constraint (2b). Here, we exploit the assumption that $1 \leq \underline{x}_1, \underline{x}_2$ holds to obtain strict feasibility w.r.t. the variable bounds in (2e) and (2f). Moreover, y is also strictly feasible w.r.t. the inequality constraints (2c) due to the following. For all $i \in \{1, \ldots, n-2\}$, we have

$$y_{i+1} - y_i^2 = \frac{i+1}{2^{2^n}} - \left(\frac{i}{2^{2^n}}\right)^2 = \frac{2^{2^n}(i+1) - i^2}{\left(2^{2^n}\right)^2} = \frac{2^{2^n} + 2^{2^n} i \left(1 - \frac{i}{2^{2^n}}\right)}{\left(2^{2^n}\right)^2} > 0.$$

Furthermore, we have

$$y_n - y_{n-1}^2 = \frac{1}{2} - \frac{1}{2^{2^n}} - \left(\frac{n-1}{2^{2^n}}\right)^2 = \frac{\left(2^{2^n}\right)^2 - 2 \cdot 2^{2^n} - 2(n-1)^2}{2 \cdot \left(2^{2^n}\right)^2}$$

$$= \frac{2^{2^n} \left(2^{2^n} - 2 - \frac{2(n-1)^2}{2^{2^n}}\right)}{2 \cdot \left(2^{2^n}\right)^2} > 0.$$

In particular, this means that the problem satisfies Slater's constraint qualification. Moreover, the gradient of the single equality constraint (2b) is not the null vector. Hence, the Mangasarian–Fromovitz constraint qualification (MFCQ) is also satisfied

at every feasible decision of the follower. Let us further point out that all lower-level constraints are linear except for the quadratic but convex inequality constraints in (2c). Therefore, the feasible set of the lower-level problem (2) is convex. Nevertheless, the overall lower-level problem is nonconvex since the follower's objective function contains bilinear terms.

Before we solve the bilevel problem (1) and (2) in the following sections, let us briefly summarize the nice properties of the problem. The upper-level problem is linear and does not contain coupling constraints. The feasible set of the lower-level problem is convex and compact. For every feasible leader's decision, the lower-level problem further satisfies Slater's constraint qualification and the MFCQ is satisfied for every feasible follower's decision.

3. Exact Feasibility

In this section, we determine the unique exact solution of the bilevel problem (1) and (2). To this end, we start by solving the lower-level problem (2) analytically for an arbitrary but fixed feasible leader's decision $(x_1, x_2) \in [\underline{x}_1, \bar{x}_1] \times [\underline{x}_2, \bar{x}_2]$.

First, we note that any feasible follower's decision y satisfies $y_n > 0$. The reasons are as follows. Let us contrarily assume that $y_n = 0$ holds. Then, Constraint (2b) yields $y_1 = 1/2$. From $y_n = 0$ and (2c), it follows that $y_i = 0$ holds for all $i \in \{1, \ldots, n\}$, which contradicts $y_1 = 1/2$. Consequently, $y_n > 0$ holds. For later reference, let us briefly summarize the previous observation.

Result 1. *For every feasible leader's decision $(x_1, x_2) \in [\underline{x}_1, \bar{x}_1] \times [\underline{x}_2, \bar{x}_2]$, a feasible follower's decision y satisfies $y_n > 0$.*

The equality constraint (2b) thus yields $y_1 < 1/2$. From (2e) and (2f), we additionally obtain
$$y_n (x_1 + x_2 - y_{n+1} - y_{n+2}) \geq 0.$$
In particular, the latter term is minimized for $(y_{n+1}, y_{n+2}) = (x_1, x_2)$. Therefore, the lower-level objective function value can be bounded from above by
$$f(x, y) = y_1 - y_n (x_1 + x_2 - y_{n+1} - y_{n+2}) \leq y_1 < \frac{1}{2}.$$
It is thus evident that an optimal follower's decision y^* satisfies $(y_{n+1}^*, y_{n+2}^*) = (x_1, x_2)$. Here, we can fix (y_{n+1}^*, y_{n+2}^*) since these variables are subject to simple bound constraints and, in particular, they are not coupled to the other variables of the follower. Hence, the follower's problem can be reduced to the convex problem

$$\max_{y} \quad y_1 \tag{3a}$$

$$\text{s.t.} \quad y_1 + y_n = \frac{1}{2}, \tag{3b}$$

$$y_i^2 \leq y_{i+1}, \quad i \in \{1, \ldots, n-1\}, \tag{3c}$$

$$y_i \geq 0, \quad i \in \{1, \ldots, n\}. \tag{3d}$$

As shown above, Problem (3) satisfies Slater's constraint qualification.[1] Again as shown above, the feasible set is compact. Therefore, Problem (3) has an optimal solution y^*. Because of the equality constraint (3b), the lower-level objective function value y_1^* is maximized by minimizing y_n^*. From Constraints (3c) and the optimality of y^*, we obtain
$$y_i^* = (y_1^*)^{2^{i-1}} \qquad \text{for all } i \in \{2, \ldots, n\},$$

---

[1]It can also be shown that the linear independence constraint qualification is valid at all solutions of the follower's problem (for any given leader's decision x). For the latter, see Appendix A.

where y_1^* denotes the root of the function

$$h : \left[0, \frac{1}{2}\right] \to \mathbb{R}, \quad z \mapsto z + z^{2^{n-1}} - \frac{1}{2}. \qquad (4)$$

In particular, one can show that y_1^* is the unique root of (4). The function h is continuous and strictly increasing on $[0, 1/2]$. Moreover, we have $h(0) < 0$ and $h(1/2) > 0$. Consequently, there is a unique point $y_1^* \in (0, 1/2)$ such that $h(y_1^*) = 0$ holds. Furthermore, the follower's decision y^* is the unique solution of Problem (3). To see this, let us assume that there is another feasible follower's decision $\hat{y} \neq y^*$ for which the optimal objective function value y_1^* is obtained, i.e., $\hat{y}_1 = y_1^*$. Then, there must be at least one quadratic inequality constraint in (3c) that is not satisfied with equality for $\hat{y}$. Otherwise, we have $y^* = \hat{y}$. However, if there is slack in Constraints (3c), we obtain $\hat{y}_n > y_n^*$. Then, (3b) yields

$$y_1^* = \frac{1}{2} - y_n^* > \frac{1}{2} - \hat{y}_n = \hat{y}_1,$$

which is a contradiction to the optimality of $\hat{y}$.

Result 2. *For every feasible leader's decision $(x_1, x_2) \in [\underline{x}_1, \bar{x}_1] \times [\underline{x}_2, \bar{x}_2]$, the set of optimal solutions of the lower-level problem (2) is a singleton.*

In particular, Result 2 means that there is no need to distinguish between the optimistic and the pessimistic approach to bilevel optimization; see, e.g., Dempe (2002). Thus, we can finally determine an optimal leader's decision for the overall bilevel problem (1) and (2). As $(y_{n+1}^*, y_{n+2}^*) = (x_1, x_2)$ holds in the optimal follower's decision y^*, the leader actually solves the linear problem

$$\max_x \quad -x_1 + x_2 \quad \text{s.t.} \quad (x_1, x_2) \in [\underline{x}_1, \bar{x}_1] \times [\underline{x}_2, \bar{x}_2].$$

The unique optimal solution is given by $x^* = (\underline{x}_1, \bar{x}_2)$.

Result 3. *The bilevel problem (1) and (2) has a unique solution given by $x^* = (\underline{x}_1, \bar{x}_2)$ with an optimal objective function value of $F^* = -\underline{x}_1 + \bar{x}_2$.*

To sum up, the bilevel problem (1) and (2) not only has nice properties such as a convex and bounded lower-level feasible set as well as a lower-level problem that satisfies Slater's constraint qualification, but also has a unique optimal solution. Moreover, the strict complementarity condition holds for which we give a proof in Appendix B. Overall, the bilevel problem (1) and (2) is thus well-behaved.

4. ε-FEASIBILITY

In what follows, we determine an optimal solution of the bilevel problem (1) and (2) under the assumption that we allow for small violations of the nonlinear lower-level constraints according to the following notion, which is motivated by the necessary special treatment of nonlinear (and, in particular, nonconvex) constraints in global optimization as we discussed it in the introduction.

Definition 1. *Let $0 < \varepsilon \in \mathbb{R}$, $f : \mathbb{R}^n \to \mathbb{R}$, $g : \mathbb{R}^n \to \mathbb{R}^m$, and $h : \mathbb{R}^n \to \mathbb{R}^p$ be given. A point $x \in \mathbb{R}^n$ is called ε-feasible for the problem $\max_{x \in \mathbb{R}^n}\{f(x) : g(x) \leq 0, h(x) = 0\}$ if $g_i(x) \leq 0$ and $h_j(x) = 0$ holds for all $i \in \{1, \ldots, m\} \setminus N$ as well as for all $j \in \{1, \ldots, p\} \setminus M$ and if $\max\{\max\{g_i(x) : i \in N\}, \max\{|h_j(x)| : j \in M\}\} \leq \varepsilon$ holds, where $N \subseteq \{1, \ldots, m\}$ and $M \subseteq \{1, \ldots, p\}$ denote the index sets of all nonlinear inequality and equality constraints.*

A follower's decision of the form

$$y_i = 2^{-2^{i-1}}, \; i \in \{1, \ldots, n-1\}, \; y_n = 0, \; y_{n+1} \in [0, x_1], \text{ and } y_{n+2} \in [-x_2, x_2] \quad (5)$$

is ε-feasible with $\varepsilon = 2^{-2^{n-1}}$ for every feasible leader's decision $(x_1, x_2) \in [\underline{x}_1, \bar{x}_1] \times [\underline{x}_2, \bar{x}_2]$ due to the following. The constraints

$$
\begin{aligned}
y_1 + y_n &= \frac{1}{2}, \\
y_i^2 &\leq y_{i+1}, \quad i \in \{1, \ldots, n-2\}, \\
y_i &\geq 0, \quad i \in \{1, \ldots, n\}, \\
y_{n+1} &\in [0, x_1], \\
y_{n+2} &\in [-x_2, x_2]
\end{aligned}
$$

are (exactly) satisfied, whereas only the constraint $y_{n-1}^2 \leq y_n$ is violated by $\varepsilon = 2^{-2^{n-1}}$. Moreover, the lower-level objective function value is $1/2$.

Result 4. *If $\varepsilon \geq 2^{-2^{n-1}}$, there is an ε-feasible follower's decision y with $y_n = 0$ for every feasible leader's decision $(x_1, x_2) \in [\underline{x}_1, \bar{x}_1] \times [\underline{x}_2, \bar{x}_2]$.*

It can easily be seen that by increasing n, we can obtain arbitrarily small values for ε. In particular, there is no ε-feasible follower's decision that yields a better objective function value than $1/2$. The reasons are as follows. Using the equality constraint (2b), the lower-level objective function can be re-written as

$$
f(x, y) = \frac{1}{2} - y_n - y_n \left(x_1 + x_2 - y_{n+1} - y_{n+2} \right).
$$

For all ε-feasible follower's decisions, we have

$$
x_1 + x_2 - y_{n+1} - y_{n+2} \geq 0
$$

because of the linear constraints (2e) and (2f). Consequently, a lower-level objective function value larger than $1/2$ could only be obtained if $y_n < 0$. However, this is not ε-feasible w.r.t. the variable bounds (2d). In addition, a follower's decision of the form stated in (5) is thus an ε-feasible solution of the lower-level problem (2). Let us point out that, in contrast to the exact case, the follower's variables y_{n+1} and y_{n+2} do not affect the lower-level objective function value in this setting and can thus be chosen arbitrarily. Therefore, the set of ε-feasible follower's solutions is not a singleton anymore.

Result 5. *If $\varepsilon \geq 2^{-2^{n-1}}$, the set of ε-feasible follower's solutions is not a singleton for every feasible leader's decision $(x_1, x_2) \in [\underline{x}_1, \bar{x}_1] \times [\underline{x}_2, \bar{x}_2]$.*

Due to Result 5, we need to distinguish between optimistic and pessimistic solutions. Following the optimistic approach, the follower chooses $y_{n+1} = 0$ as well as $y_{n+2} = x_2$ such as to favor the leader w.r.t. the leader's objective function value. Therefore, the leader actually solves the linear problem

$$
\max_x \quad x_1 + x_2 \quad \text{s.t.} \quad (x_1, x_2) \in [\underline{x}_1, \bar{x}_1] \times [\underline{x}_2, \bar{x}_2].
$$

The optimistic optimal leader's decision is thus given by $x^* = (\bar{x}_1, \bar{x}_2)$. In the pessimistic case, the follower chooses $y_{n+1} = x_1$ as well as $y_{n+2} = -x_2$ such as to adversely affect the leader's decision. In this setting, the leader solves the linear problem

$$
\max_x \quad -x_1 - x_2 \quad \text{s.t.} \quad (x_1, x_2) \in [\underline{x}_1, \bar{x}_1] \times [\underline{x}_2, \bar{x}_2].
$$

Hence, the pessimistic optimal leader's decision is given by $x^* = (\underline{x}_1, \underline{x}_2)$. To sum up, let us state the main observations of this section.

Result 6. *Let $\varepsilon \geq 2^{-2^{n-1}}$ and suppose that we allow for ε-feasible follower's solutions. Then, the optimistic optimal solution of the bilevel problem (1) and (2) is given by $x_o^* = (\bar{x}_1, \bar{x}_2)$ with an optimal objective function value of $F_o^* = \bar{x}_1 + \bar{x}_2$.*

The pessimistic optimal solution is given by $x_p^* = (\underline{x}_1, \underline{x}_2)$ with an optimal objective function value of $F_p^* = -\underline{x}_1 - \underline{x}_2$.

We now finally compare the results of the exact bilevel solution with the results for the optimistic and pessimistic setting for the case of only ε-feasibility of the lower level. In the optimistic setting, the distance between the solutions is $\bar{x}_1 - \underline{x}_1$ and the difference between the corresponding objective function values is $\bar{x}_1 + \underline{x}_1$. Two aspects are remarkable. First, by enlarging the feasible interval for the variable x_1, we get an arbitrarily large error and, second, this error is independent of ε, i.e., this arbitrarily large error occurs independent of how accurate one solves the lower-level problem.

For the pessimistic setting, the distance between the solution is $\bar{x}_2 - \underline{x}_2$ and the difference between the objective function values is $\bar{x}_2 + \underline{x}_2$. Hence, we obtain the same qualitative behavior but now in dependence of the variable x_2 instead of x_1.

In summary, we obtain the following two main observations. First, we can be arbitrarily far away from the overall exact bilevel solution. Second, we also obtain arbitrarily large errors regarding the optimal objective function value of the leader. The latter is very much in contrast to the situation in single-level optimization for which sensitivity results are available; see, e.g., Proposition 4.2.2 in Bertsekas (2016). This is particularly the case for linear optimization problems, where standard sensitivity analysis results (see, e.g., Theorem 5.5 in Chvátal (1983)) apply as well and state that a small change in the right-hand side of the problem's constraints can only lead to a small change in the optimal objective function value.

Lastly, let us comment on that only very moderate values of n are required to get the wrong solution. Taking the inequality for ε from Result 6, it is easy to see that for a given tolerance ε, the parameter n needs to satisfy $n \geq \log_2(\log_2(1/\varepsilon^2))$ so that numerically computed solutions do not coincide with the exact solution for the given ε. For instance, a tolerance of $\varepsilon = 10^{-8}$ already leads to a wrong result for $n = 6$. This particularly means that the considered bilevel problem is moderate in size w.r.t. the number of constraints and variables. For $n = 6$, we only have 16 constraints and 8 variables on the lower level. We further note that the used constraint coefficients are all 1 and that the coefficients are independent from n and the given tolerance ε.

A Python code for the example considered in this paper is publicly available at https://github.com/m-schmidt-math-opt/ill-behaved-bilevel-example and can be used to verify the discussed results.

5. Analysis of the ε-Feasible Linear Case

In this section, we analyze the linear bilevel case, i.e., we study the problem

$$\min_{x,y} \quad c_x^\top x + c_y^\top y \tag{6a}$$

$$\text{s.t.} \quad Ax \geq a, \tag{6b}$$

$$y \in \arg\min_{\bar{y}} \left\{ d^\top \bar{y} \colon Cx + D\bar{y} \geq b \right\} \tag{6c}$$

with $c_x \in \mathbb{R}^{n_x}$, $c_y, d \in \mathbb{R}^{n_y}$, $A \in \mathbb{R}^{m \times n_x}$, $a \in \mathbb{R}^m$, $C \in \mathbb{R}^{\ell \times n_x}$, $D \in \mathbb{R}^{\ell \times n_y}$, and $b \in \mathbb{R}^\ell$. We assume that the set $\{(x,y) \in \mathbb{R}^{n_x} \times \mathbb{R}^{n_y} \colon Ax \geq a, Cx + Dy \geq b\}$ is non-empty and compact and that for every feasible upper-level decision x, there exists a feasible lower-level decision y. This implies that the lower-level problem is bounded for every feasible upper-level decision and that the dual problem of the lower level is feasible. We also assume that the set $\{x \in \mathbb{R}^{n_x} \colon Ax \geq a\}$ is bounded. Moreover, we consider the setting in which the underlying linear algebra and linear optimization routines are of finite precision only.

When finite-precision procedures are used, an algorithm that solves Problem (6) will output a pair $(\hat{x}, \hat{y})$ that may be slightly infeasible. The concern, should that happen, is that the solution being output can be *superoptimal* to a degree that is not proportional to its infeasibility. As discussed in the previous section, such an outcome can be observed for general, i.e., nonlinear, bilevel problems. In this section, however, we show that linear bilevel problems behave better in some sense. To this end, we assume that our underlying solver can ensure the following properties:

- $A\hat{x} \geq a - \varepsilon e_m$ and $C\hat{x} + D\hat{y} \geq b - \varepsilon e_\ell$,
- $d^\top \hat{y} \geq \min\{d^\top y \colon C\hat{x} + Dy \geq b\} - \varepsilon$.

Here and in what follows, $0 < \varepsilon < 1$ is a given tolerance, $e_k \in \mathbb{R}^k$ is the vector of all ones, and $(\hat{x}, \hat{y})$ is used to denote a nearly feasible solution of the bilevel problem (6).

Prior to our analysis, we present a general result that will be used below. This result can be read from Theorem 3.38 (Page 112) of Conforti et al. (2014). It can also be obtained from Corollary 3.2b (Page 20) of Schrijver (1986) or from Theorem 10.2 (Page 121) of Schrijver (1986). We will use the term *size* to refer to the (bit) encoding length of a matrix, vector, or formulation, as appropriate.

Definition 2. *Let $P = \{x \in \mathbb{R}^n \colon Hx = h, \, x \geq 0\}$ with $H \in \mathbb{R}^{m \times n}$ and $h \in \mathbb{R}^m$. Given $z \in \mathbb{R}^n$, we say that z is* basic *if $Hz = h$ and, defining $B = B(z) = \{j \colon z_j \neq 0\}$, the submatrix H_B of H corresponding to the columns in B has rank $|B|$. Furthermore, if in addition $z \geq 0$, we say that z is* basic feasible.

Remark 1. *Let P be as in Definition 2. The extreme points of P are precisely the vectors z that are basic feasible.*

Theorem 1. *Let $P = \{x \in \mathbb{R}^n \colon Hx = h, \, x \geq 0\}$ with $H \in \mathbb{R}^{m \times n}$ and $h \in \mathbb{R}^m$. There is a constant $\kappa(H) > 0$ of size polynomial in the size (of the bit-encoding) of H such that, for any basic vector v, we have*

$$\|v\|_\infty \leq \kappa(H)\|h\|_\infty.$$

Proof. Let v be basic and set $J = \{j \colon v_j \neq 0\}$. Since v is basic, there is a subset of rows I of H with $|I| = |J|$ such that the following holds:

(i) The submatrix $H_{I,J}$ of H indexed by rows I and columns J is invertible.
(ii) As a consequence, it holds $v_J = H_{I,J}^{-1} h_I$, where v_J is the subvector of v indexed by J and h_I is the subvector of h indexed by I.

Using submultiplicativity of the norm, we get

$$\|v\|_\infty \leq \|H_{I,J}^{-1}\|_\infty \|h_I\|_\infty.$$

The result now follows by defining $\kappa(H)$ to be the maximum over all $\|B^{-1}\|_\infty$ for B being an invertible submatrix of H. $\qquad\square$

In Theorem 1, we use what is usually termed the *standard* representation of a polyhedron. Similar statements can be derived using other representations of polyhedra, e.g., $\{x \in \mathbb{R}^n \colon Hx \leq h\}$, via well-known reformulations.

5.1. **Linear Optimization with Errors.** We start with some simple observations for classic, i.e., single-level, linear problems of the form

$$v^* := \min_{x \in \mathbb{R}^{n_x}} \left\{ v^\top x \colon Mx \geq f \right\} \tag{7}$$

with $v \in \mathbb{R}^{n_x}$, $0 \neq M \in \mathbb{R}^{m \times n_x}$, and $f \in \mathbb{R}^m$. Throughout this section, we assume that the feasible region for problem (7) is non-empty and bounded. Moreover, we denote the corresponding dual problem by

$$\max_{z \in \mathbb{R}^m} \left\{ f^\top z \colon M^\top z = v, \, z \geq 0 \right\}.$$

Next, we will derive estimates involving near-feasible and near-optimal points for Problem (7).

Lemma 1. *Suppose that there is a point $\hat{x} \in \mathbb{R}^{n_x}$ that is nearly feasible for Problem (7), i.e., $M\hat{x} \geq f - \varepsilon e_m$. Then, the following holds.*

  (a) *It holds*
$$v^\top \hat{x} \geq v^* - \varepsilon \kappa(M) \|v\|_\infty,$$
   where $\kappa(M) > 0$ is a constant of polynomial size.
  (b) *There exists x^* feasible for Problem (7) such that*
$$\|x^* - \hat{x}\|_\infty \leq \varepsilon \kappa_1(M)$$
   holds for a certain constant $\kappa_1(M) > 0$ of polynomial size.

Proof.   (a) Let z^* be an optimal solution of the dual problem of (7). Then,
$$v^\top \hat{x} = (z^*)^\top M\hat{x} \geq (z^*)^\top (f - \varepsilon e_m) = v^* - \varepsilon \|z^*\|_1$$

holds. In particular, this equation applies to any dual optimal z^*. Since $z \geq 0$ is a constraint of the dual problem, the dual feasible region is a pointed polyhedron, and, w.l.o.g., z^* is an extreme point. The result now follows from Theorem 1.

  (b) Consider the linear optimization problem

$$\min_{x,\delta} \quad \delta \tag{8a}$$

$$\text{s.t.} \quad \|x - \hat{x}\|_\infty \leq \delta, \tag{8b}$$

$$Mx \geq f. \tag{8c}$$

That this is indeed a linear program follows by reformulating (8b) as

$$x_j - \delta \leq \hat{x}_j, \quad -x_j - \delta \leq -\hat{x}_j, \quad 1 \leq j \leq n_x. \tag{9}$$

Clearly, the resulting problem is both feasible and bounded since (7) is. Moreover, $(\hat{x}, 0)$ satisfies the constraints of this problem with additive error of at most ε. We can therefore apply (a) to this problem to obtain x^* being feasible for Problem (7) and such that

$$\|x^* - \hat{x}\|_\infty \leq \varepsilon \kappa_1(M)$$

holds, where $\kappa_1(M)$ is the κ-constant (of polynomial size in M) that applies to the matrix for Constraints (9) and (8c). □

Let us emphasize that the result in Lemma 1 applies for any $\varepsilon > 0$, no matter how large. In particular, it is not required that ε is "sufficiently small".

Lemma 2. *Suppose that there is a nearly primal-dual feasible and nearly primal-dual optimal pair $(\hat{x}, \hat{z}) \in \mathbb{R}^{n_x} \times \mathbb{R}^m$ for Problem (7), i.e., $(\hat{x}, \hat{z})$ satisfies*

  (i) *$M\hat{x} \geq f - \varepsilon e_m$,*
  (ii) *$\|M^\top \hat{z} - v\|_\infty \leq \varepsilon$, $\hat{z} \geq -\varepsilon e_m$, and*
  (iii) *$v^\top \hat{x} - f^\top \hat{z} \leq \varepsilon$.*

Then, there exists an optimal solution x^* for Problem (7) such that

$$\|x^* - \hat{x}\|_\infty \leq \varepsilon \kappa_3(M, v) \max\{1, \|f\|_\infty\}$$

holds for a certain constant $\kappa_3(M, v) > 0$, whose size is polynomial in the size of the input data M and v.

Proof. First, we note that Condition (ii) simply states that $\hat{z}$ is feasible for the dual of (7) up to an error of ε. We can thus apply Part (a) of Lemma 1 to obtain

$$f^\top \hat{z} \leq v^* + \varepsilon \kappa_2(M)\|f\|_\infty, \tag{10}$$

where $\kappa_2(M)$ is the κ-constant for the dual of (7), which is of polynomial size in M. Together with (iii), this implies

$$v^\top \hat{x} \leq v^* + \varepsilon(1 + \kappa_2(M)\|f\|_\infty). \tag{11}$$

Next, we consider the polyhedron given by

$$Mx \geq f, \tag{12a}$$

$$v^\top x \leq v^*, \tag{12b}$$

which is feasible and bounded. By (i) and (11), $\hat{x}$ satisfies these inequalities with feasibility error of at most $\varepsilon(1+\kappa_2(M)\|f\|_\infty)$. By applying Part (b) of Lemma 1, we obtain that there is a feasible point x^* for (12), i.e., an optimal solution x^* for (7), such that

$$\|x^* - \hat{x}\|_\infty \leq \varepsilon \kappa_1(M,v)(1 + \kappa_2(M)\|f\|_\infty)$$

holds. $\square$

5.2. **Application to Linear Bilevel Problems.** We now return to the bilevel setup as stated in (6). To this end, note that for a given upper-level decision $x \in \mathbb{R}^{n_x}$, the dual of the lower-level problem (6c) reads

$$\max_z \quad (b - Cx)^\top z \tag{13a}$$

$$\text{s.t.} \quad D^\top z = d, \tag{13b}$$

$$z \geq 0. \tag{13c}$$

Lemma 3. *Let $x \in \mathbb{R}^{n_x}$, $\hat{y} \in \mathbb{R}^{n_y}$, and $\hat{z} \in \mathbb{R}^\ell$ be such that*

(i) $Ax \geq a$,
(ii) $D\hat{y} \geq b - Cx - \varepsilon e_\ell$,
(iii) $\|D^\top \hat{z} - d\|_\infty \leq \varepsilon$, $\hat{z} \geq -\varepsilon e_\ell$, and
(iv) $d^\top \hat{y} - (b - Cx)^\top \hat{z} \leq \varepsilon$.

Then, there exists an optimal solution y^* for the x-parameterized lower-level problem (6c) such that

$$\|y^* - \hat{y}\|_\infty \leq \varepsilon \kappa_4(A, C, D, a, b, d) \tag{14}$$

holds for a constant $\kappa_4(A, C, D, a, b, d) > 0$, whose size is polynomial in the size of the input data A, C, D, a, b, and d.

Proof. By assumption, for a given x, the lower-level problem is feasible and bounded. We can thus apply Lemma 2 since Conditions (ii)–(iv) correspond to Conditions (i)–(iii) of Lemma 2. Thus, there exists an optimal point y^* for the lower-level problem such that

$$\|y^* - \hat{y}\|_\infty \leq \varepsilon \kappa_3(D, d) \max\{1, \|b - Cx\|_\infty\}$$

holds. Using the triangle inequality and the submultiplicativity of the norm, we obtain

$$\|b - Cx\|_\infty \leq \|b\|_\infty + \|C\|_\infty \|x\|_\infty.$$

Since the feasible region for the upper-level problem is bounded, $\|x\|_\infty$ is upper bounded by the ∞-norm of some extreme point. We can apply Theorem 1 to obtain

$$\|x\|_\infty \leq \kappa'(A)\|a\|_\infty,$$

where $\kappa'(A)$ is the κ-constant (of polynomial size in A) for the system $Ax \geq a$. The proof is now concluded by appropriately defining $\kappa_4(A, C, D, a, b, d)$. $\square$

Now, we consider the entire bilevel problem (6) and recall a basic definition from linear optimization.

Definition 3. *Let $z \in \mathbb{R}^{\ell}$ satisfy $D^{\top}z = d$ and define $B = B(z) = \{j \colon z_j \neq 0\}$. We say that z is* dual basic *if the submatrix $D_B^{\top}$ of $D^{\top}$ corresponding to the columns in B has rank $|B|$.*

Note that z is dual basic and feasible (i.e., $z \geq 0$) if and only if z is an extreme point of the dual polyhedron to any lower-level problem.

Theorem 2. *Let $\hat{x} \in \mathbb{R}^{n_x}$, $\hat{y} \in \mathbb{R}^{n_y}$, and $\hat{z} \in \mathbb{R}^{\ell}$ be such that*

    (i) $A\hat{x} \geq a - \varepsilon e_m$,
    (ii) $D\hat{y} \geq b - C\hat{x} - \varepsilon e_{\ell}$,
    (iii) $\|D^{\top}\hat{z} - d\|_{\infty} \leq \varepsilon$, $\hat{z} \geq -\varepsilon e_{\ell}$,
    (iv) $d^{\top}\hat{y} - (b - C\hat{x})^{\top}\hat{z} \leq \varepsilon$,
    (v) $\|\hat{z} - \tilde{z}\|_{\infty} \leq \varepsilon$ *for some dual basic $\tilde{z}$.*

Then, there exists a pair (x^*, y^*) that is feasible for the bilevel problem (6) such that

$$\|(x^*, y^*)^{\top} - (\hat{x}, \hat{y})^{\top}\|_{\infty} \leq \varepsilon \kappa_5(A, C, D, a, b, d),$$

$$|c_x^{\top}x^* + c_y^{\top}y^* - (c_x^{\top}\hat{x} + c_y^{\top}\hat{y})| \leq \varepsilon \kappa_6(A, C, D, a, b, c, d)$$

hold for certain constants $\kappa_5(A, C, D, a, b, d)$ and $\kappa_6(A, C, D, a, b, c, d) > 0$, whose sizes are polynomial in the size of the input data.

Proof. By Assumptions (i) and (ii), the pair $(\hat{x}, \hat{y}) \in \mathbb{R}^{n_x} \times \mathbb{R}^{n_y}$ is nearly feasible for the upper- and the lower-level problem of (6). Applying Part (b) of Lemma 1 to $(\hat{x}, \hat{y})$ and the system

$$\begin{bmatrix} A & 0 \\ C & D \end{bmatrix} \begin{pmatrix} x \\ y \end{pmatrix} \geq \begin{pmatrix} a \\ b \end{pmatrix} \tag{15}$$

yields (x^*, y') with $Ax^* \geq a$, $Dy' \geq b - Cx^*$, and

$$\|(x^*, y')^{\top} - (\hat{x}, \hat{y})^{\top}\|_{\infty} \leq \varepsilon \kappa_7(A, C, D), \tag{16}$$

where $\kappa_7(A, C, D)$ is the κ-constant for System (15). Next, we use (iv) and obtain

$$\begin{aligned}
& d^{\top}y' - (b - Cx^*)^{\top}\hat{z} \\
& \leq \varepsilon + d^{\top}(y' - \hat{y}) - (C(\hat{x} - x^*))^{\top}\hat{z} \\
& \leq \varepsilon + \left|d^{\top}(\hat{y} - y')\right| + \left|(C(\hat{x} - x^*))^{\top}\hat{z}\right|.
\end{aligned} \tag{17}$$

Note that, if $u, v \in \mathbb{R}^n$, then $|u^{\top}v| \leq \sum_{j=1}^{n}|u_j v_j| \leq n\|u\|_{\infty}\|v\|_{\infty}$ holds. Moreover, if $Q \in \mathbb{R}^{m \times n}$ and $u \in \mathbb{R}^n$, then, for $1 \leq i \leq m$, $|(Qu)_i| \leq (\sum_{j=1}^{n}|q_{ij}|)\|u\|_{\infty} \leq \|Q\|_{\infty}\|u\|_{\infty}$ by definition of the infinity-norm of a matrix. Hence, using (17) we obtain

$$d^{\top}y' - (b - Cx^*)^{\top}\hat{z} \leq \varepsilon + n_y\|d\|_{\infty}\|\hat{y} - y'\|_{\infty} + \ell\|C\|_{\infty}\|\hat{x} - x^*\|_{\infty}\|\hat{z}\|_{\infty}. \tag{18}$$

Further, since $\tilde{z}$ is dual basic, $\|\tilde{z}\|_{\infty}$ is upper bounded by $\kappa(D)\|d\|_{\infty}$ due to Theorem 1. Thus, using (v) yields $\|\hat{z}\|_{\infty} \leq \varepsilon + \kappa(D)\|d\|_{\infty}$. These facts, together with (16) and (18), yield

$$d^{\top}y' - (b - Cx^*)^{\top}\hat{z} \leq \varepsilon \kappa_8(A, C, D, d),$$

with $\kappa_8(A, C, D, d) \geq 1$ being appropriately defined and of polynomial size. To sum up, x^*, y', and $\hat{z}$ satisfy

    (a) $Ax^* \geq a$,
    (b) $Dy' \geq b - Cx^*$,
    (c) $\|D^{\top}\hat{z} - d\|_{\infty} \leq \varepsilon$, $\hat{z} \geq -\varepsilon e_{\ell}$,
    (d) $d^{\top}y' - (b - Cx^*)^{\top}\hat{z} \leq \varepsilon \kappa_8(A, C, D, d)$.

Thus, by Lemma 3 applied to the error $\varepsilon\kappa_8(A, C, D, d) \geq \varepsilon$, there exists an optimal solution y^* for the x^*-parameterized lower-level problem (6c) such that

$$\|y^* - y'\|_\infty \leq \varepsilon\kappa_8(A, C, D, d)\kappa_4(A, C, D, a, b, d) \tag{19}$$

holds. Using this inequality and (16) concludes the proof. □

Remark 2. *Assumption (v) in Theorem 2 states that the distance between a nearly feasible and nearly optimal solution for the dual of the lower-level problem and a basic solution for the dual is small, which is a reasonable assumption in our setting.*

To summarize the statement of the theorem, the distance to feasibility and the superoptimality of a nearly feasible pair $(\hat{x}, \hat{y})$ for the bilevel problem (6) is linear in ε with coefficients κ that have polynomial size in the input data. This type of guarantee with polynomial sized coefficients is simply unavailable in the nonlinear case as we have seen in the previous sections.

6. Conclusion

In this paper, we consider an exemplary bilevel problem with continuous variables and a nonconvex lower-level problem and illustrate that numerically obtained solutions can be arbitrarily far away from an exact solution. The discrepancy between exact and numerically computed solutions is based on the fact that we cannot exactly satisfy all constraints of the nonconvex lower level when using global optimization techniques such as spatial branching. The considered problem itself is well-posed in the sense that we do not use large constraint coefficient ranges or high-degree polynomials. Moreover, we show that the constraint set of the lower-level problem is convex, compact, and that it satisfies Slater's constraint qualification. In an exact sense, we prove that the lower-level problem as well as the overall bilevel problem possess unique solutions. It is further established that LICQ holds in every follower's solution for every feasible leader's decision. While working computationally, however, we can only expect to obtain ε-feasible solutions of the nonconvex lower-level problem. Furthermore, the set of ε-feasible follower solutions is not a singleton anymore. Thus, we determine both an optimal solution for the optimistic and the pessimistic variant of the bilevel problem. By doing so, we establish that not only the obtained ε-feasible bilevel solutions can be arbitrarily far away from the overall exact bilevel solution but that there can also be an arbitrarily large error in the objective function value of the leader.

We also show that the pathological behavior observed for nonlinear lower-level problems seems to be due to the nonlinearities by showing that linear bilevel problems behave better at least on the level of feasible points. As an important question for future research, it is still open if one can prove that the bad behavior can also not appear for more general problems than linear ones, such as convex problems, in the lower level.

Finally, our results show that computational bilevel optimization with continuous but nonconvex lower levels needs to be done with great care and that ex-post checks may be needed to avoid considering arbitrarily bad points as "solutions" of the given bilevel problem.

Acknowledgements

References

Applegate, D. L., W. Cook, S. Dash, and D. G. Espinoza (2007). "Exact solutions to linear programming problems." In: *Operations Research Letters* 35.6, pp. 693–699. DOI: 10.1016/j.orl.2006.12.010.

Beck, Y., I. Ljubić, and M. Schmidt (2023). "A Survey on Bilevel Optimization Under Uncertainty." In: *European Journal on Operational Research*. DOI: 10.1016/j.ejor.2023.01.008.

Bertsekas, D. P. (2016). *Nonlinear Programming*. Athena scientific Belmont.

Bienstock, D., A. D. Pia, and R. Hildebrand (2021). "Complexity, Exactness, and Rationality in Polynomial Optimization." In: *Integer Programming and Combinatorial Optimization*. Ed. by M. Singh and D. P. Williamson. Vol. 12707. IPCO 2021. Springer, Cham., pp. 58–72. DOI: 10.1007/978-3-030-73879-2_5.

Buchheim, C. and D. Henke (2022). "The robust bilevel continuous knapsack problem with uncertain coefficients in the follower's objective." In: *Journal of Global Optimization*. DOI: 10.1007/s10898-021-01117-9.

Burtscheidt, J. and M. Claus (2020). "Bilevel Linear Optimization Under Uncertainty." In: *Bilevel Optimization: Advances and Next Challenges*. Ed. by S. Dempe and A. Zemkoho. Springer International Publishing, pp. 485–511. DOI: 10.1007/978-3-030-52119-6_17.

Chvátal, V. (1983). *Linear Programming*. A Series of books in the mathematical sciences. New York (N. Y.): Freeman.

Conforti, M., G. Cornuéjols, and G. Zambelli (2014). *Integer Programming*. Springer International Publishing. DOI: 10.1007/978-3-319-11008-0.

Dempe, S. (2002). *Foundations of Bilevel Programming*. Springer. DOI: 10.1007/b101970.

Fischetti, M., I. Ljubić, M. Monaci, and M. Sinnl (2017). "A New General-Purpose Algorithm for Mixed-Integer Bilevel Linear Programs." In: *Operations Research* 65.6, pp. 1615–1637. DOI: 10.1287/opre.2017.1650.

– (2018). "On the use of intersection cuts for bilevel optimization." In: *Mathematical Programming* 172.1-2, pp. 77–103. DOI: 10.1007/s10107-017-1189-5.

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.

Kleinert, T., V. Grimm, and M. Schmidt (2021a). "Outer Approximation for Global Optimization of Mixed-Integer Quadratic Bilevel Problems." In: *Mathematical Programming (Series B)*. DOI: 10.1007/s10107-020-01601-2.

Kleinert, T., M. Labbé, I. Ljubić, and M. Schmidt (2021b). "A Survey on Mixed-Integer Programming Techniques in Bilevel Optimization." In: *EURO Journal on Computational Optimization*. DOI: 10.1016/j.ejco.2021.100007.

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.

Kleniati, P.-M. and C. S. Adjiman (2011). "Branch-and-Sandwich: An Algorithm for Optimistic Bi-Level Programming Problems." In: *21st European Symposium on Computer Aided Process Engineering*. Ed. by E. Pistikopoulos, M. Georgiadis, and A. Kokossis. Vol. 29. Computer Aided Chemical Engineering. Elsevier, pp. 602–606. DOI: 10.1016/B978-0-444-53711-9.50121-8.

– (2014a). "Branch-and-Sandwich: a deterministic global optimization algorithm for optimistic bilevel programming problems. Part I: Theoretical development." In: *Journal of Global Optimization* 60.3, pp. 425–458. DOI: 10.1007/s10898-013-0121-7.

Kleniati, P.-M. and C. S. Adjiman (2014b). "Branch-and-Sandwich: a deterministic global optimization algorithm for optimistic bilevel programming problems. Part II: Convergence analysis and numerical results." In: *Journal of Global Optimization* 60.3, pp. 459–481. DOI: 10.1007/s10898-013-0120-8.

– (2015). "A generalization of the Branch-and-Sandwich algorithm: From continuous to mixed-integer nonlinear bilevel problems." In: *Computers & Chemical Engineering* 72, pp. 373–386. DOI: 10.1016/j.compchemeng.2014.06.004.

Locatelli, M. and F. Schoen (2013). *Global Optimization*. Philadelphia, PA: Society for Industrial and Applied Mathematics. DOI: 10.1137/1.9781611972672.

Mitsos, A. (2010). "Global solution of nonlinear mixed-integer bilevel programs." In: *Journal of Global Optimization* 47.4, pp. 557–582. DOI: 10.1007/s10898-009-9479-y.

Mitsos, A., P. Lemonidis, and P. I. Barton (2008). "Global solution of bilevel programs with a nonconvex inner program." In: *Journal of Global Optimization* 42.4, pp. 475–513. DOI: 10.1007/s10898-007-9260-z.

Paulavičius, R. and C. S. Adjiman (2020). "New bounding schemes and algorithmic options for the Branch-and-Sandwich algorithm." In: *Journal of Global Optimization*, pp. 1–29. DOI: 10.1007/s10898-020-00874-3.

Paulavičius, R., J Gao, P.-M. Kleniati, and C. Adjiman (2020). "BASBL: Branch-And-Sandwich BiLevel solver. Implementation and computational study with the BASBLib test set." In: *Computers & Chemical Engineering*, p. 106609. DOI: 10.1016/j.compchemeng.2019.106609.

Paulavičius, R., P.-M. Kleniati, and C. S. Adjiman (2016). "Global optimization of nonconvex bilevel problems: implementation and computational study of the Branch-and-Sandwich algorithm." In: *26th European Symposium on Computer Aided Process Engineering*. Ed. by Z. Kravanja and M. Bogataj. Vol. 38. Computer Aided Chemical Engineering. Elsevier, pp. 1977–1982. DOI: 10.1016/B978-0-444-63428-3.50334-9.

Schrijver, A. (1986). *Theory of Linear and Integer Programming*. John Wiley & Sons, Chichester.

APPENDIX A. PROOF OF THE LINEAR INDEPENDENCE CONSTRAINT QUALIFICATION

Let $(x_1, x_2) \in [\underline{x}_1, \bar{x}_1] \times [\underline{x}_2, \bar{x}_2]$ with $1 \leq \underline{x}_i < \bar{x}_i$, $i \in \{1, 2\}$, be arbitrary but fixed. Further, let y^* be the exact optimal solution of the follower for the given leader's decision x. As shown in Section 3, a follower's solution y^* satisfies $y_i^* > 0$ for all $i \in \{1, \ldots, n + 2\}$. This means that the non-negativity constraints (2d) as well as the lower bound constraints in (2e) and (2f) are inactive in an optimal follower's decision. Conversely, all quadratic constraints (2c) as well as the upper bound constraints in (2e) and (2f) are active. Hence, the Jacobian matrix of the single equality constraint and the active inequality constraints in an optimal decision of the follower is given by

$$
\begin{bmatrix}
1 & & & & & & & 1 \\
2y_1^* & -1 & & & & & & \\
& 2y_2^* & -1 & & & & & \\
& & \ddots & \ddots & & & & \\
& & & 2y_{n-2}^* & -1 & & & \\
& & & & 2y_{n-1}^* & -1 & & \\
& & & & & & 1 & \\
& & & & & & & 1
\end{bmatrix}.
$$

All matrix entries that are left blank here correspond to zeros. It is easy to verify that the Jacobian matrix has full rank, i.e., the linear independence constraint qualification holds.

APPENDIX B. PROOF OF THE STRICT COMPLEMENTARITY CONDITION

Let $(x_1, x_2) \in [\underline{x}_1, \bar{x}_1] \times [\underline{x}_2, \bar{x}_2]$ with $1 \leq \underline{x}_i < \bar{x}_i$, $i \in \{1, 2\}$, be arbitrary but fixed. For the x-parameterized lower-level problem (2), the Lagrangian function reads

$$\mathcal{L}(y, \alpha, \beta, \gamma, \delta^{\pm}, \pi) = - y_1 + y_n (x_1 + x_2 - y_{n+1} - y_{n+2})$$
$$- \sum_{i=1}^{n-1} \alpha_i \left(y_{i+1} - y_i^2\right) - \sum_{i=1}^{n+1} \beta_i y_i$$
$$- \gamma (x_1 - y_{n+1}) - \delta^- (y_{n+2} + x_2)$$
$$- \delta^+ (x_2 - y_{n+2}) - \pi \left(y_1 + y_n - \frac{1}{2}\right)$$

with the Lagrange multipliers $\alpha \in \mathbb{R}_{\geq 0}^{n-1}$, $\beta \in \mathbb{R}_{\geq 0}^{n+1}$, $\gamma, \delta^{\pm} \in \mathbb{R}_{\geq 0}$, and $\pi \in \mathbb{R}$. The KKT complementarity conditions of Problem (2) are given by

$$\alpha_i \left(y_{i+1} - y_i^2\right) = 0, \quad i \in \{1, \ldots, n-1\}, \tag{20a}$$
$$\beta_i y_i = 0, \quad i \in \{1, \ldots, n+1\}, \tag{20b}$$
$$\gamma (x_1 - y_{n+1}) = 0, \tag{20c}$$
$$\delta^- (y_{n+2} + x_2) = 0, \tag{20d}$$
$$\delta^+ (x_2 - y_{n+2}) = 0. \tag{20e}$$

Let y^* be the exact optimal solution of the follower for the given leader's decision x. Further, let α^*, β^*, γ^*, $(\delta^{\pm})^*$, and π^* be the corresponding Lagrange multipliers so that $(y^*, \alpha^*, \beta^*, \gamma^*, (\delta^{\pm})^*, \pi^*)$ is a KKT point of the lower-level problem. As shown in Appendix A, the linear independence constraint qualification is valid at all solutions of the follower's problem (for any given leader's decision x). Hence, the Lagrange multipliers α^*, β^*, γ^*, $(\delta^{\pm})^*$, and π^* are uniquely determined. Also, as observed in Section 3, the follower's optimal decision y^* satisfies $(y_{n+1}^*, y_{n+2}^*) = (x_1, x_2)$ and

$$y_i^* = (y_1^*)^{2^{i-1}} \quad \text{for all } i \in \{2, \ldots, n\},$$

where y_1^* is the unique root of the function h as given in (4). We now show that the strict complementarity condition is satisfied.

Observation 1. *The point* $(y^*, \alpha^*, \beta^*, \gamma^*, (\delta^{\pm})^*, \pi^*)$ *satisfies* $\beta_i^* = 0$ *for all* $i \in \{1, \ldots, n+1\}$ *as well as* $(\delta^-)^* = 0$.

Proof. For all $i \in \{1, \ldots, n\}$, we have $y_i^* > 0$. By (20b), we thus obtain $\beta_i^* = 0$ for all $i \in \{1, \ldots, n\}$. From (20b), we further obtain $\beta_{n+1}^* = 0$ since $y_{n+1}^* = x_1 \geq \underline{x}_1 \geq 1 > 0$ holds. Finally, due to $y_{n+2}^* + x_2 = 2x_2 \geq 2\underline{x}_2 \geq 2 > 0$, (20d) yields $(\delta^-)^* = 0$. □

Observation 2. *The point* $(y^*, \alpha^*, \beta^*, \gamma^*, (\delta^{\pm})^*, \pi^*)$ *satisfies* $\gamma^*, (\delta^+)^* > 0$.

Proof. Since $(y^*, \alpha^*, \beta^*, \gamma^*, (\delta^{\pm})^*, \pi^*)$ is a KKT point,

$$\nabla_{y_{n+1}} \mathcal{L}(y^*, \alpha^*, \beta^*, \gamma^*, (\delta^{\pm})^*, \pi^*) = -y_n^* - \beta_{n+1}^* + \gamma^* = 0,$$
$$\nabla_{y_{n+2}} \mathcal{L}(y^*, \alpha^*, \beta^*, \gamma^*, (\delta^{\pm})^*, \pi^*) = -y_n^* - (\delta^-)^* + (\delta^+)^* = 0$$

are satisfied. By Observation 1 and $y_n^* > 0$, we obtain $\gamma^*, (\delta^+)^* > 0$. □

Observation 3. *The point* $(y^*, \alpha^*, \beta^*, \gamma^*, (\delta^{\pm})^*, \pi^*)$ *satisfies* $2\alpha_1^* y_1^* = 1 + \pi^*$.

Proof. Since $(y^*, \alpha^*, \beta^*, \gamma^*, (\delta^{\pm})^*, \pi^*)$ is a KKT point,

$$\nabla_{y_1}\mathcal{L}(y^*, \alpha^*, \beta^*, \gamma^*, (\delta^{\pm})^*, \pi^*) = -1 + 2\alpha_1^* y_1^* - \beta_1^* - \pi^* = 0$$

is satisfied. Since, by Observation 1, we have $\beta_1^* = 0$, we obtain $2\alpha_1^* y_1^* = 1 + \pi^*$. $\quad\square$

Observation 4. *The point* $(y^*, \alpha^*, \beta^*, \gamma^*, (\delta^{\pm})^*, \pi^*)$ *satisfies* $2\alpha_i^* y_i^* = \alpha_{i-1}^*$ *for all* $i \in \{2, \ldots, n-1\}$.

Proof. Since $(y^*, \alpha^*, \beta^*, \gamma^*, (\delta^{\pm})^*, \pi^*)$ is a KKT point,

$$\nabla_{y_i}\mathcal{L}(y^*, \alpha^*, \beta^*, \gamma^*, (\delta^{\pm})^*, \pi^*) = -\alpha_{i-1}^* + 2\alpha_i^* y_i^* - \beta_i^* = 0$$

is satisfied for all $i \in \{2, \ldots, n-1\}$. By Observation 1, we obtain $2\alpha_i^* y_i^* = \alpha_{i-1}^*$ for all $i \in \{2, \ldots, n-1\}$. $\quad\square$

Observation 5. *The point* $(y^*, \alpha^*, \beta^*, \gamma^*, (\delta^{\pm})^*, \pi^*)$ *satisfies* $\pi^* = -\alpha_{n-1}^*$.

Proof. Since $(y^*, \alpha^*, \beta^*, \gamma^*, (\delta^{\pm})^*, \pi^*)$ is a KKT point,

$$\nabla_{y_n}\mathcal{L}(y^*, \alpha^*, \beta^*, \gamma^*, (\delta^{\pm})^*, \pi^*) = x_1 + x_2 - y_{n+1}^* - y_{n+2}^* - \alpha_{n-1}^* - \beta_n^* - \pi^* = 0$$

is satisfied. By Observation 1, we have $\beta_n^* = 0$. Moreover, $(y_{n+1}^*, y_{n+2}^*) = (x_1, x_2)$ holds. Hence, we obtain $\pi^* = -\alpha_{n-1}^*$. $\quad\square$

Observation 6. *The point* $(y^*, \alpha^*, \beta^*, \gamma^*, (\delta^{\pm})^*, \pi^*)$ *satisfies* $\alpha_i^* > 0$ *for all* $i \in \{1, \ldots, n-1\}$.

Proof. We show this by contradiction. Suppose that $\alpha_1^* = 0$ holds. By Observation 4 and $y_i^* > 0$ for all $i \in \{1, \ldots, n\}$, we obtain $\alpha_i^* = 0$ for all $i \in \{1, \ldots, n-1\}$. Then, Observation 5 yields $\pi^* = -\alpha_{n-1}^* = 0$. By Observation 3, however, we obtain $0 = 2\alpha_1^* y_1^* = 1 + \pi^*$, i.e., $\pi^* = -1$ which is a contradiction to $\pi^* = 0$. Consequently, $\alpha_1^* > 0$ needs to holds and, thus, we have $\alpha_i^* > 0$ for all $i \in \{1, \ldots, n-1\}$ by Observation 4. $\quad\square$

To sum up, we have

$$\begin{aligned}
y_{i+1}^* &= (y_i^*)^2 \quad \text{and} \quad \alpha_i^* > 0 \quad \text{for all } i \in \{1, \ldots, n-1\}, \\
y_i^* &> 0 \quad \text{and} \quad \beta_i^* = 0 \quad \text{for all } i \in \{1, \ldots, n+1\}, \\
y_{n+1}^* &= x_1 \quad \text{and} \quad \gamma^* > 0, \\
y_{n+2}^* &> -x_2 \quad \text{and} \quad (\delta^-)^* = 0, \\
y_{n+2}^* &= x_2 \quad \text{and} \quad (\delta^+)^* > 0,
\end{aligned}$$

i.e., strict complementarity is satisfied.

(Y. Beck, M. Schmidt, J. Thürauf) Trier University, Department of Mathematics, Universitätsring 15, 54296 Trier, Germany

Email address: yasmine.beck@uni-trier.de
Email address: martin.schmidt@uni-trier.de
Email address: johannes.thuerauf@uni-trier.de

(D. Bienstock) Columbia University, Department of Industrial Engineering and Operations Research, 500 W. 120th Street #315, New York, NY 10027, United States

Email address: dano@columbia.edu