A Gentle and Incomplete Introduction to Bilevel Optimization

Martin Schmidt

Belgian Mathematical Optimization Workshop
April 22, 2024

Agenda

1. What is bilevel optimization anyway?
2. How do you solve a linear bilevel problem?
3. What to do if you have an MINLP in the lower level?
4. An (in my opinion) important open problem

I hope in 2 hours, you want to read ...

A Survey on Mixed-Integer Programming Techniques in Bilevel Optimization
In: EURO Journal on Computational Optimization. 2021
Jointly with Thomas Kleinert, Martine Labbé, and Ivana Ljubic

A Gentle and Incomplete Introduction to Bilevel Optimization
Publicly available lectures notes
Jointly with Yasmine Beck

What is bilevel optimization anyway?

“Usual” single-level problems

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

- only one objective function f
- one vector of variables x
- one set of constraints g and h

“Usual” single-level problems

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

- only one objective function f
- one vector of variables x
- one set of constraints g and h

This models a situation in which a single decision maker takes all decisions, i.e., decides on the variables of the problem.

“Usual” single-level problems

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

- only one objective function f
- one vector of variables x
- one set of constraints g and h

This models a situation in which a single decision maker takes all decisions, i.e., decides on the variables of the problem.

Very often, that’s appropriate:

- a single dispatcher controls a gas transport network
- a single investment banker decides on the assets in a portfolio
- a single logistics company decides on its supply chain

Often, life's different

- Many situations in our day-to-day life are different
- Often:
  - A decision maker makes a decision ...
  - ... while anticipating the (rational, i.e., optimal) reaction of another decision maker
  - The decision of the other decision maker depends on the first decision
- Thus: the outcome (or in more mathematical terms, the objective function and/or feasible set) depends on the decision/reaction of the other decision maker

Often, life's different

- Many situations in our day-to-day life are different
- Often:
  - A decision maker makes a decision ...
  - ... while anticipating the (rational, i.e., optimal) reaction of another decision maker
  - The decision of the other decision maker depends on the first decision
- Thus: the outcome (or in more mathematical terms, the objective function and/or feasible set) depends on the decision/reaction of the other decision maker

Formalizing this situation leads to hierarchical or bilevel optimization problems

Informal example: Pricing

- A very rich class of applications of bilevel optimization
- First decision maker (leader)
    - decides on a price of a certain good
      (or maybe on different prices for multiple goods)
    - goal: maximize revenue from selling these goods

Informal example: Pricing

- A very rich class of applications of bilevel optimization
- First decision maker (leader)
  - decides on a price of a certain good
    (or maybe on different prices for multiple goods)
  - goal: maximize revenue from selling these goods
- Second decision maker (follower)
  - 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
- the decision of the follower depends on the (pricing) decisions of the leader

A bit more formal, please

Definition (Bilevel optimization problem)

A bilevel optimization problem is given by

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

A bit more formal, please

Definition (Bilevel optimization problem)

A bilevel optimization problem is given by

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

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

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

A bit more formal, please ... continued

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

... and ...

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

Wording

- First problem: so-called upper-level (or the leader's) problem
- Second Problem is the so-called lower-level (or the follower's) problem
- Both problems are parameterized by the decisions of the other player
- $x \in \mathbb{R}^{n_x}$: upper-level variables
  - decisions of the leader
- $y \in \mathbb{R}^{n_y}$: lower-level variables
  - decisions of the follower

A bit more formal, please ... continued

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

... and ...

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

Functions and dimensions

- Objective functions
  - $F, f : \mathbb{R}^{n_x} \times \mathbb{R}^{n_y} \to \mathbb{R}$
- Constraint functions
  - $G : \mathbb{R}^{n_x} \times \mathbb{R}^{n_y} \to \mathbb{R}^m$
  - $g : \mathbb{R}^{n_x} \times \mathbb{R}^{n_y} \to \mathbb{R}^\ell$
  - The sets $X \subseteq \mathbb{R}^{n_x}$ and $Y \subseteq \mathbb{R}^{n_y}$ are typically used to denote integrality constraints.
  - Example: $Y = \mathbb{Z}^{n_y}$ makes the lower-level problem an integer program

A bit more formal, please ... continued

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

... and ...

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

Definition

1. We call upper-level constraints $G_i(x, y) \geq 0$, $i \in \{1, \ldots, m\}$, coupling constraints if they explicitly depend on the lower-level variable vector y.
2. All upper-level variables that appear in the lower-level constraints are called linking variables.

Optimal value function

Instead of using the point-to-set mapping S …

Optimal value function

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

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

Optimal value function reformulation

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

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

and re-write the bilevel problem as

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

Shared constraint set, bilevel feasible set, inducible region

Definition

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.

Shared constraint set, bilevel feasible set, inducible region

Definition

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

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

Shared constraint set, bilevel feasible set, inducible region

Definition

The set

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

is called the shared constraint set.

Its projection onto the x-space is denoted by

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

Definition

The set

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

is called the bilevel feasible set or inducible region.

High-point relaxation

Definition

The problem of minimizing the upper-level objective function over the shared constraint set, i.e.,

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

is called the high-point relaxation (HPR) of the bilevel problem.

High-point relaxation

Definition

The problem of minimizing the upper-level objective function over the shared constraint set, i.e.,

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

is called the high-point relaxation (HPR) of the bilevel problem.

Remark

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

Pricing revisited

- First bilevel pricing problem with linear constraints, linear upper-level objective and bilinear lower-level objective: Bialas and Karwan (1984)

Pricing revisited

- First bilevel pricing problem with linear constraints, linear upper-level objective and bilinear lower-level objective: Bialas and Karwan (1984)
- Here: a more general version taken from Labbé et al. (1998)

$$
\begin{aligned}
\max_{x, y=(y_1, y_2)} \quad & x^\top y_1 \\
\text{s.t.} \quad & Ax \leq a \\
& y \in \arg\min_{\bar{y}} \left\{ (x + d_1)^\top \bar{y}_1 + d_2^\top \bar{y}_2 : D_1 \bar{y}_1 + D_2 \bar{y}_2 \geq b \right\}
\end{aligned}
$$

Pricing revisited

$$\begin{aligned}\max_{x, y=(y_1, y_2)} \quad & x^\top y_1 \\ \text{s.t.} \quad & Ax \le a \\ & y \in \arg\min_{\bar{y}} \left\{ (x + d_1)^\top \bar{y}_1 + d_2^\top \bar{y}_2 : D_1 \bar{y}_1 + D_2 \bar{y}_2 \ge b \right\}\end{aligned}$$

- 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

Pricing revisited

$$
\begin{aligned}
\max_{x, y=(y_1, y_2)} \quad & x^\top y_1 \\
\text{s.t.} \quad & Ax \leq a \\
& y \in \arg\min_{\bar{y}} \left\{ (x + d_1)^\top \bar{y}_1 + d_2^\top \bar{y}_2 : D_1 \bar{y}_1 + D_2 \bar{y}_2 \geq b \right\}
\end{aligned}
$$

- 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
- Upper-level player influences the activities of plan y_1 through the price vector x that is additionally imposed onto y_1

Pricing revisited

$$
\begin{aligned}
\max_{x, y=(y_1, y_2)} \quad & x^\top y_1 \\
\text{s.t.} \quad & Ax \leq a \\
& y \in \arg\min_{\bar{y}} \left\{ (x + d_1)^\top \bar{y}_1 + d_2^\top \bar{y}_2 : D_1 \bar{y}_1 + D_2 \bar{y}_2 \geq b \right\}
\end{aligned}
$$

- 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
- Upper-level player influences the activities of plan y_1 through the price vector x that is additionally imposed onto y_1
- Goal of the leader is to maximize her revenue given by $x^\top y_1$

Pricing revisited

$$
\begin{aligned}
\max_{x, y=(y_1,y_2)} \quad & x^\top y_1 \\
\text{s.t.} \quad & Ax \leq a \\
& y \in \arg\min_{\bar{y}} \left\{ (x + d_1)^\top \bar{y}_1 + d_2^\top \bar{y}_2 : D_1\bar{y}_1 + D_2\bar{y}_2 \geq b \right\}
\end{aligned}
$$

- 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
- Upper-level player influences the activities of plan y_1 through the price vector x that is additionally imposed onto y_1
- Goal of the leader is to maximize her revenue given by $x^\top y_1$
- Price vector x is subject to linear constraints that may, among others, impose lower and upper bounds on the prices

Pricing revisited

$$\begin{aligned} \max_{x, y=(y_1, y_2)} \quad & x^\top y_1 \\ \text{s.t.} \quad & Ax \leq a \\ & y \in \arg\min_{\bar{y}} \left\{ (x + d_1)^\top \bar{y}_1 + d_2^\top \bar{y}_2 \colon D_1 \bar{y}_1 + D_2 \bar{y}_2 \geq b \right\} \end{aligned}$$

- 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

Pricing revisited

$$
\begin{aligned}
\max_{x,y=(y_1,y_2)} \quad & x^\top y_1 \\
\text{s.t.} \quad & Ax \le a \\
& y \in \arg\min_{\bar{y}} \left\{ (x + d_1)^\top \bar{y}_1 + d_2^\top \bar{y}_2 : D_1 \bar{y}_1 + D_2 \bar{y}_2 \ge b \right\}
\end{aligned}
$$

- 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
- d_2 may also encompass the price for executing the activities not influenced by the leader
    - These activities may, e.g., be substitutes offered by competitors for which prices are known and fixed

Pricing revisited

$$\begin{aligned}
\max_{x,y=(y_1,y_2)} \quad & x^\top y_1 \\
\text{s.t.} \quad & Ax \le a \\
& y \in \arg\min_{\bar{y}} \left\{ (x+d_1)^\top \bar{y}_1 + d_2^\top \bar{y}_2 : D_1\bar{y}_1 + D_2\bar{y}_2 \ge b \right\}
\end{aligned}$$

- 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
- d_2 may also encompass the price for executing the activities not influenced by the leader
  - 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

Pricing revisited

$$\begin{aligned} \max_{x, y=(y_1, y_2)} \quad & x^\top y_1 \\ \text{s.t.} \quad & Ax \leq a \\ & y \in \arg\min_{\bar{y}} \left\{ (x + d_1)^\top \bar{y}_1 + d_2^\top \bar{y}_2 : D_1 \bar{y}_1 + D_2 \bar{y}_2 \geq b \right\} \end{aligned}$$

- 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
- d_2 may also encompass the price for executing the activities not influenced by the leader
  - 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
- 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 $\{y_2 : D_2 y_2 \geq b\}$ is non-empty

An Academic and Linear Example (Kleinert 2021)

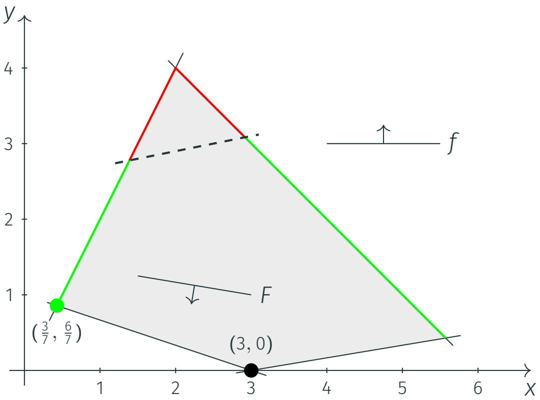
Upper-level problem

$$\begin{aligned} \min_{x,y} \quad & F(x,y) = x + 6y \\ \text{s.t.} \quad & -x + 5y \leq 12.5 \\ & x \geq 0 \\ & y \in S(x) \end{aligned}$$

Lower-level problem

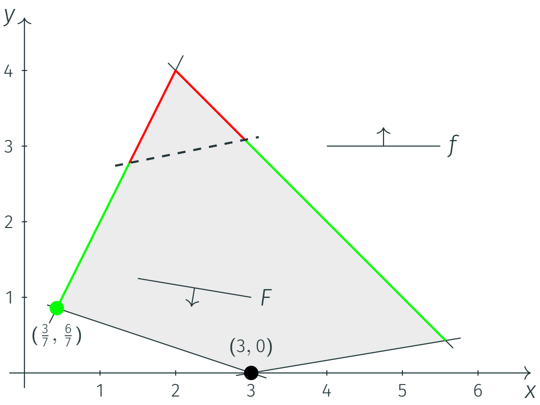
$$\begin{aligned} \min_{y} \quad & f(x,y) = -y \\ \text{s.t.} \quad & 2x - y \geq 0 \\ & -x - y \geq -6 \\ & -x + 6y \geq -3 \\ & x + 3y \geq 3 \end{aligned}$$

An Academic and Linear Example (Kleinert 2021)



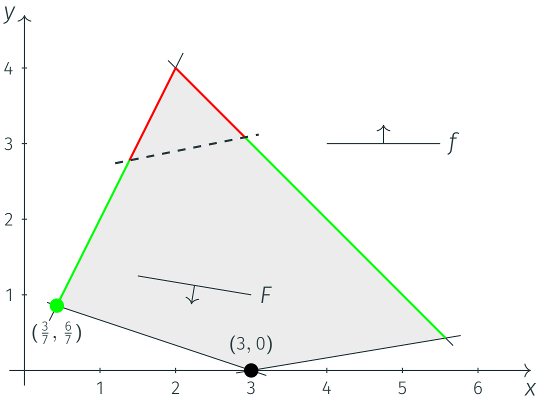
- Shared constrained set: gray area
- Green and red lines: nonconvex set of optimal follower solutions (lifted to the *x*-*y*-space)
- Green lines: Nonconvex and disconnected bilevel feasible set of the bilevel problem

An Academic and Linear Example (Kleinert 2021)



1. The feasible region of the follower problem corresponds to the gray area.
2. The follower's problem—and therefore the bilevel problem—is infeasible for certain decisions of the leader, e.g., $x = 0$.
3. 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.
4. This set is nonconvex!

An Academic and Linear Example (Kleinert 2021)

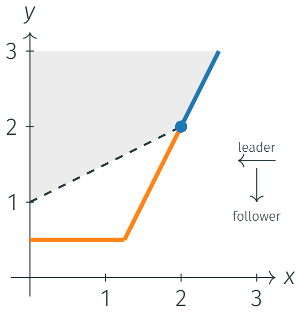


5. The single leader constraint (dashed line) renders certain optimal responses of the follower infeasible.
6. The bilevel feasible region $\mathcal{F}$ corresponds to the green facets.
7. Thus, the feasible set is not only nonconvex but also disconnected.
8. The optimal solution is $(3/7, 6/7)$ with objective function value $39/7$.
9. 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.

Independence of irrelevant constraints (Kleinert et al. 2021; Macal and Hurter 1997)

$$\min_{x,y \in \mathbb{R}} \quad x$$
$$\text{s.t.} \quad y \geq 0.5x + 1, \; x \geq 0$$
$$y \in \arg\min_{\bar{y} \in \mathbb{R}} \{\bar{y} : \bar{y} \geq 2x - 2, \; \bar{y} \geq 0.5\}$$

Optimal solution: $(2, 2)$

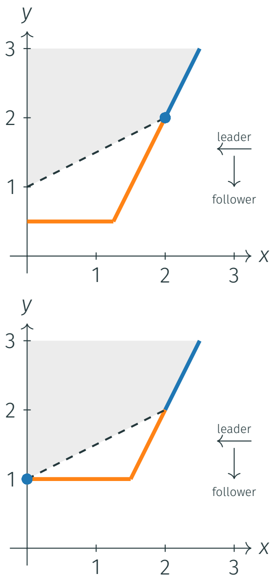


Independence of irrelevant constraints (Kleinert et al. 2021; Macal and Hurter 1997)

- Strengthening $\bar{y} \geq 0.5$ in the lower-level problem using $y \geq 0.5x + 1$ of the upper-level problem
- This yields the minimum value of $0.5x + 1$ is 1 due to $x \geq 0$
- New bound of $\bar{y}$ is $\bar{y} \geq 1$
- High-point relaxation stays the same

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

Optimal solution: $(0, 1) \neq (2, 2)$



A Brief History of Complexity Results

- Jeroslow (1985): hardness general multilevel models
- Corollary: NP-hardness of the LP-LP bilevel problem
- Hansen et al. (1992): LP-LP bilevel problems are strongly NP-hard
  - reduction from KERNEL
- Vicente et al. (1994): even checking whether a given point is a local minimum of a bilevel problem is NP-hard

How do you solve a linear bilevel problem?

Using optimality conditions

Most classic approach to obtain a single-level reformulation:

Exploit optimality conditions for the lower-level problem

Using optimality conditions

Most classic approach to obtain a single-level reformulation:

Exploit optimality conditions for the lower-level problem

- These optimality conditions need to be necessary and sufficient
- This is usually only possible for convex lower-level problems that satisfy a reasonable constraint qualification

An LP-LP Bilevel Problem

- Let's keep it simple: KKT reformulation of an LP-LP bilevel
- Consider

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

- Data: $c_x \in \mathbb{R}^{n_x}$, $c_y, d \in \mathbb{R}^{n_y}$, $A \in \mathbb{R}^{m \times n_x}$, $B \in \mathbb{R}^{m \times n_y}$, 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 $b \in \mathbb{R}^\ell$

KKT reformulation of LP-LP bilevel problems

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

KKT reformulation of LP-LP bilevel problems

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

Lower-level problem can be seen as the x-parameterized linear problem

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

KKT reformulation of LP-LP bilevel problems

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

Lower-level problem can be seen as the x-parameterized linear problem

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

Its Lagrangian function is given by

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

KKT reformulation of LP-LP bilevel problems

The KKT conditions of the lower level are given by ...

- dual feasibility

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

- primal feasibility

$$Cx + Dy \geq b$$

- and the KKT complementarity conditions

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

KKT reformulation of LP-LP bilevel problems

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

KKT reformulation of LP-LP bilevel problems

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

- We now optimize over an extended space of variables including the lower-level dual variables λ
- Since we optimize over x, y, and λ simultaneously, any global solution of the problem above corresponds to an optimistic bilevel solution
- The KKT reformulation is linear except for the KKT complementarity conditions
- Thus, the problem is a nonconvex NLP

KKT reformulation of LP-LP bilevel problems

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

- ...
- Thus, the problem is a nonconvex NLP

KKT reformulation of LP-LP bilevel problems

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

- ...
- Thus, the problem is a nonconvex NLP

It is even worse! It’s a mathematical program with complementarity constraints (an MPCC).

KKT reformulation of LP-LP bilevel problems

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

- ...
- Thus, the problem is a nonconvex NLP

It is even worse! It's a mathematical program with complementarity constraints (an MPCC).

Bad news (Ye and Zhu 1995)

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.

How to solve the KKT reformulation?

Remember

The "only" reason for the nonconvexity of the KKT reformulation are the bilinear products of the lower-level dual variables λ_i and the upper-level primal variables x in the term

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

How to solve the KKT reformulation?

Remember

The “only” reason for the nonconvexity of the KKT reformulation 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

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

How to solve the KKT reformulation?

Key idea: Linearize these terms by exploiting the combinatorial structure of the KKT complementarity conditions.

How to solve the KKT reformulation?

Key idea: Linearize these terms by exploiting the combinatorial structure of the KKT complementarity conditions.

The complementarity conditions

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

can be seen 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.

How to solve the KKT reformulation?

Key idea: Linearize these terms by exploiting the combinatorial structure of the KKT complementarity conditions.

The complementarity conditions

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

can be seen 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

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

in the following mixed-integer linear way:

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

Here, M is a sufficiently large constant.

By construction, we get the following result.

Theorem

Suppose that M is a sufficiently large constant. Then, the KKT reformulation is equivalent to the mixed-integer linear optimization problem

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

Be careful!

OPERATIONS RESEARCH

JOURNAL HOME ARTICLES IN ADVANCE CURRENT ISSUE ARCHIVES ABOUT SUBMIT SUBSCRIBE

View PDF

Tools Share

Home > Operations Research > Vol. 68, No. 6 >

Technical Note—There's No Free Lunch: On the Hardness of Choosing a Correct Big-M in Bilevel Optimization

Thomas Kleinert, Martine Labbé, Fr¨ank Plein, Martin Schmidt

Published Online: 30 Jun 2020 | https://doi.org/10.1287/opre.2019.1944

Go to Section

Abstract

Abstract

One of the most frequently used approaches to solve linear bilevel optimization problems consists in replacing the lower-level problem with its Karush–Kuhn–Tucker (KKT) conditions and by reformulating the KKT complementarity conditions using techniques from mixed-integer linear optimization. The latter step requires to determine some big-*M* constant in order to bound the lower level's dual feasible set such that no bilevel-optimal solution is cut off. In practice, heuristics are often used to find a big-*M* although it is known that these approaches may fail. In this paper, we consider the hardness of two proxies for the above mentioned concept of a bilevel-correct big-*M*. First, we prove that verifying that a given big-*M* does not cut off any feasible vertex of the lower level's dual polyhedron cannot be done in polynomial time unless P = NP. Second, we show that verifying that a given big-*M* does not cut off any optimal point of the lower level's dual problem (for any point in the projection of the high-point relaxation onto the leader's decision space) is as hard as solving the original bilevel problem.

What to do if you have an MINLP in the lower level?

The people that did the entire work ...

Convex Integer Nonlinear Bilevel Problem

$$
\begin{aligned}
\min_{x \in \mathbb{Z}^{n_1}, y \in \mathbb{Z}^{n_2}} \quad & F(x, y) \\
\text{s.t.} \quad & G(x, y) \leq 0 \\
& y \in \arg\min_{\bar{y} \in \mathbb{Z}^{n_2}} \{ f(x, \bar{y}) : g(x, \bar{y}) \leq 0 \}
\end{aligned}
$$

- All variables are integer
- All functions are continuous and jointly convex

Literature

- Mixed-integer linear bilevel problems
    - DeNegre and Ralphs (2009)
    - Wang and XU (2017)
    - Fischetti, Ljubić, Monaci, and Sinnl (2017), (2018)
- Convex mixed-integer nonlinear bilevel problems
    - Gaar, Lee, Ljubić, Sinnl, and Tanınmış (2023)
      (“but” with special structural assumptions)

Branch-and-Cut (à la Ralphs, DeNegre 2009)

- $\Omega := \{(x, y) \in \mathbb{R}^n : G(x, y) \leq 0,\ g(x, y) \leq 0\}$

Branch-and-Cut (à la Ralphs, DeNegre 2009)

- $\Omega := \{(x, y) \in \mathbb{R}^n : G(x, y) \leq 0,\ g(x, y) \leq 0\}$
- At node N_j: $\Omega_j := \Omega \cap \{(x, y) \in \mathbb{R}^n : A^j x + B^j y \leq a^j\}$

Branch-and-Cut (à la Ralphs, DeNegre 2009)

- $\Omega := \{(x, y) \in \mathbb{R}^n : G(x, y) \leq 0,\ g(x, y) \leq 0\}$
- At node N_j: $\Omega_j := \Omega \cap \{(x, y) \in \mathbb{R}^n : A^j x + B^j y \leq a^j\}$
- $\min_{(x,y)\in\Omega_j} F(x, y)$

Branch-and-Cut (à la Ralphs, DeNegre 2009)

- $\Omega := \{(x, y) \in \mathbb{R}^n : G(x, y) \leq 0,\ g(x, y) \leq 0\}$
- At node N_j: $\Omega_j := \Omega \cap \{(x, y) \in \mathbb{R}^n : A^j x + B^j y \leq a^j\}$
- $\min_{(x,y)\in\Omega_j} F(x, y)$
- (x^j, y^j) solution of node N_j

Branch-and-Cut (à la Ralphs, DeNegre 2009)

- $\Omega := \{(x, y) \in \mathbb{R}^n : G(x, y) \leq 0,\ g(x, y) \leq 0\}$
- At node N_j: $\Omega_j := \Omega \cap \{(x, y) \in \mathbb{R}^n : A^j x + B^j y \leq a^j\}$
- $\min_{(x,y)\in\Omega_j} F(x, y)$
- (x^j, y^j) solution of node N_j
- Branch on integrality constraints

Branch-and-Cut (à la Ralphs, DeNegre 2009)

- $\Omega := \{(x,y) \in \mathbb{R}^n : G(x,y) \le 0,\ g(x,y) \le 0\}$
- At node N_j: $\Omega_j := \Omega \cap \{(x,y) \in \mathbb{R}^n : A^j x + B^j y \le a^j\}$
- $\min_{(x,y)\in\Omega_j} F(x,y)$
- (x^j, y^j) solution of node N_j
- Branch on integrality constraints
- If $(x^j, y^j) \in \mathbb{Z}^n$ is not bilevel-feasible, add a local cut

Branch-and-Cut (à la Ralphs, DeNegre 2009)

- $\Omega := \{(x, y) \in \mathbb{R}^n : G(x, y) \leq 0,\ g(x, y) \leq 0\}$
- At node N_j: $\Omega_j := \Omega \cap \{(x, y) \in \mathbb{R}^n : A^j x + B^j y \leq a^j\}$
- $\min_{(x,y) \in \Omega_j} F(x, y)$
- (x^j, y^j) solution of node N_j
- Branch on integrality constraints
- If $(x^j, y^j) \in \mathbb{Z}^n$ is not bilevel-feasible, add a local cut

How to derive such a cut?

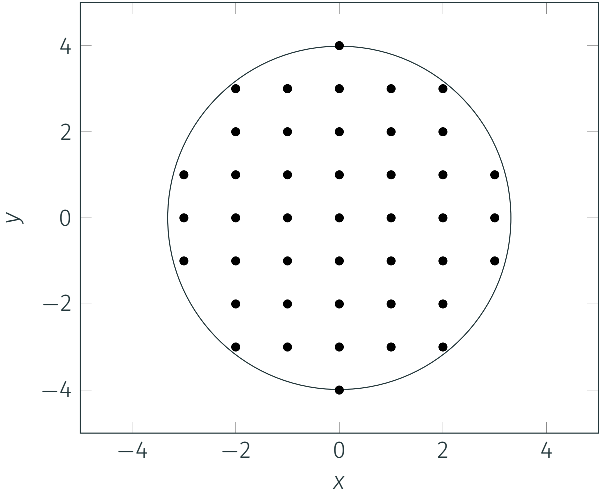
Theorem (Similar to Fischetti, Ljubić, Monaci, Sinnl 2017)

For any $\hat{y} \in \mathbb{Z}^{n_2}$, *the set*

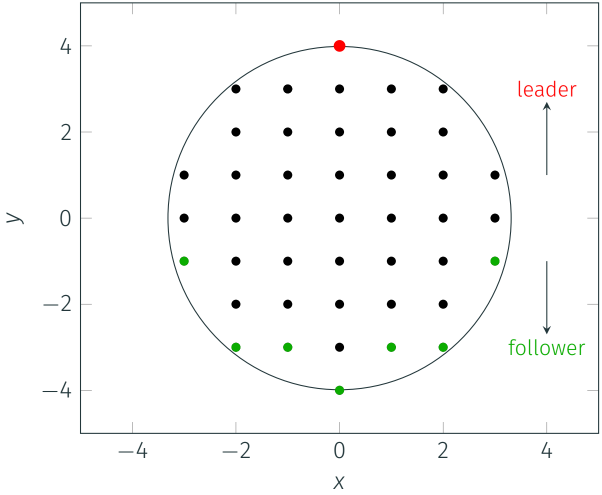
$$S(\hat{y}) := \left\{(x, y) \in \mathbb{R}^n : g(x, \hat{y}) \leq 0,\ f(x, y) > f(x, \hat{y})\right\}$$

does not contain any bilevel-feasible point.

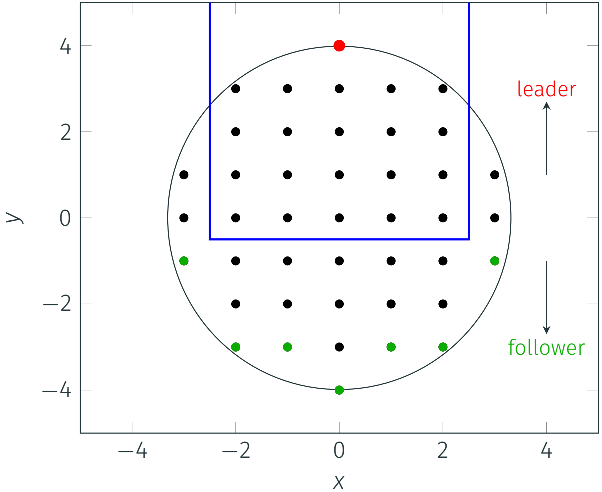
Use of Bilevel-Free Sets: Intersection Cuts



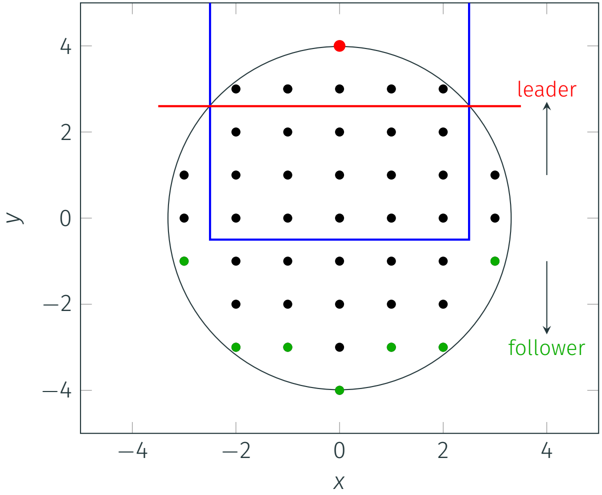
Use of Bilevel-Free Sets: Intersection Cuts



Use of Bilevel-Free Sets: Intersection Cuts



Use of Bilevel-Free Sets: Intersection Cuts



Computing Optimal Bilevel-Free Sets

Idea: Starting from y^j, go to a point $y^j + \Delta y$ such that the bilevel-free set $S(y^j + \Delta y)$ contains the point (x^j, y^j) in its interior

Idea: Starting from y^j, go to a point $y^j + \Delta y$ such that the bilevel-free set $S(y^j + \Delta y)$ contains the point (x^j, y^j) in its interior

Solve the convex mixed-integer scoop problem (Wang and Xu 2017)

$$
\begin{aligned}
\max_{\Delta y, s, t} \quad & t \\
\text{s.t.} \quad & t \leq s_i \quad \text{for all } i \in I_0 \\
& g_i(x^j, (y^j + \Delta y)) + s_i \leq 0 \quad \text{for all } i \in I := \{1, \ldots, m_2\} \\
& f(x^j, (y^j + \Delta y)) - f(x^j, y^j) + s_0 \leq 0 \\
& \Delta y \in \mathbb{Z}^{n_2} \\
& s_i \geq 0 \quad \text{for all } i \in I_0 := I \cup \{0\}
\end{aligned}
$$

General Idea: Disjunctive Cuts

- Suppose we have an Δy^j with $t^j > 0$

General Idea: Disjunctive Cuts

- Suppose we have an Δy^j with $t^j > 0$
- $S(y^j + \Delta y^j) := \left\{(x, y) \in \mathbb{R}^n : g(x, (y^j + \Delta y^j)) \leq 0, f(x, y) > f(x, (y^j + \Delta y^j))\right\}$

General Idea: Disjunctive Cuts

- Suppose we have an Δy^j with $t^j > 0$
- $S(y^j + \Delta y^j) := \left\{(x, y) \in \mathbb{R}^n : g(x, (y^j + \Delta y^j)) \leq 0, f(x, y) > f(x, (y^j + \Delta y^j))\right\}$
- $\mathcal{D}_0(y^j + \Delta y^j) := \left\{(x, y) : f(x, y) \leq f(x, (y^j + \Delta y^j))\right\}$

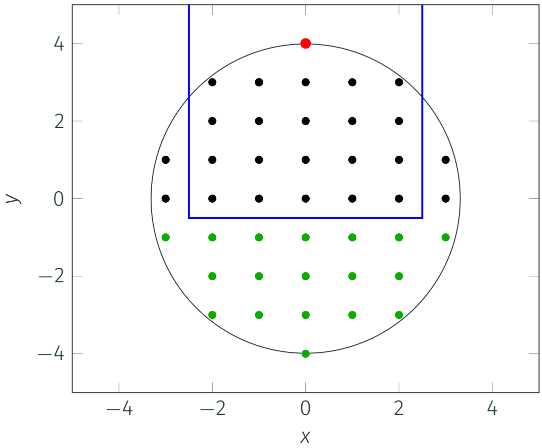
General Idea: Disjunctive Cuts

- Suppose we have an Δy^j with $t^j > 0$
- $S(y^j + \Delta y^j) := \left\{(x, y) \in \mathbb{R}^n : g(x, (y^j + \Delta y^j)) \leq 0, f(x, y) > f(x, (y^j + \Delta y^j))\right\}$
- $\mathcal{D}_0(y^j + \Delta y^j) := \left\{(x, y) : f(x, y) \leq f(x, (y^j + \Delta y^j))\right\}$
- $\mathcal{D}_i(y^j + \Delta y^j) := \left\{(x, y) : g_i(x, (y^j + \Delta y^j)) \geq 0\right\}$ for $i \in I = \{1, \ldots, m_2\}$

General Idea: Disjunctive Cuts

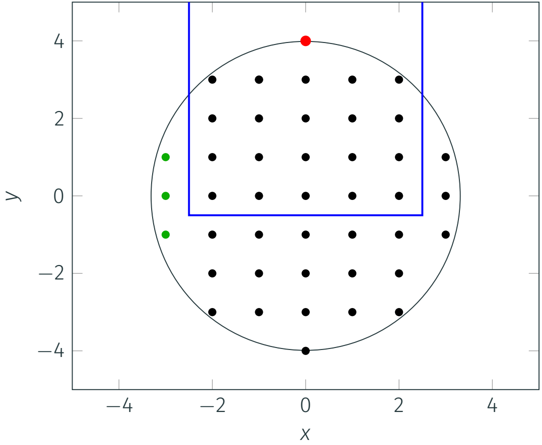
- Suppose we have an Δy^j with $t^j > 0$
- $S(y^j + \Delta y^j) := \left\{(x, y) \in \mathbb{R}^n : g(x, (y^j + \Delta y^j)) \leq 0, f(x, y) > f(x, (y^j + \Delta y^j))\right\}$
- $\mathcal{D}_0(y^j + \Delta y^j) := \left\{(x, y) : f(x, y) \leq f(x, (y^j + \Delta y^j))\right\}$
- $\mathcal{D}_i(y^j + \Delta y^j) := \left\{(x, y) : g_i(x, (y^j + \Delta y^j)) \geq 0\right\}$ for $i \in I = \{1, \dots, m_2\}$
- $\bigcup_{i \in I_0} \mathcal{D}_i(y^j + \Delta y^j) = \operatorname{int}(S(y^j + \Delta y^j))^C$, $I_0 = \{0, \dots, m_2\}$

Illustration of the Disjunction $\bigcup_{i\in I_0} \mathcal{D}_i(y^j + \Delta y^j)$



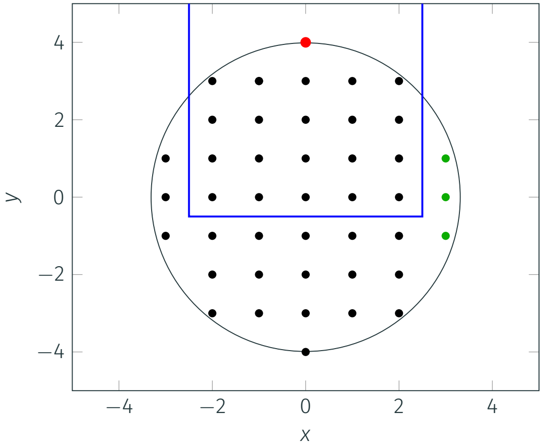
$\Omega_j \cap \mathbb{Z}^n \cap \mathcal{D}_0(y^j + \Delta y^j)$

Illustration of the Disjunction $\bigcup_{i \in I_0} \mathcal{D}_i(y^j + \Delta y^j)$



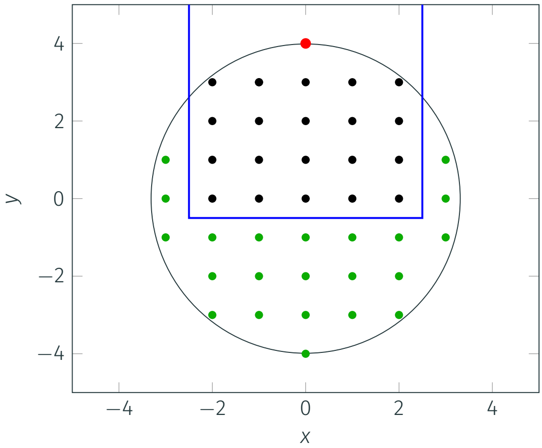
$\Omega_j \cap \mathbb{Z}^n \cap \mathcal{D}_1(y^j + \Delta y^j)$

Illustration of the Disjunction $\bigcup_{i \in I_0} \mathcal{D}_i(y^j + \Delta y^j)$



$\Omega_j \cap \mathbb{Z}^n \cap \mathcal{D}_2(y^j + \Delta y^j)$

Illustration of the Disjunction $\bigcup_{i \in I_0} \mathcal{D}_i(y^j + \Delta y^j)$



$$\Omega_j \cap \mathbb{Z}^n \cap \left(\bigcup_{i \in I_0} \mathcal{D}_i(y^j + \Delta y^j) \right)$$

Theorem (Gaar, Lee, Ljubić, Sinnl, Tanınmış 2022)

Let $(x^j, y^j) \in \mathbb{Z}^n$ be an extreme point of the convex set Ω_j which belongs to the interior of the bilevel-free set $S(y^j + \Delta y^j)$ for an appropriate Δy^j. Then, there exists a disjunctive cut that separates (x^j, y^j) from $\Omega_j \cap \mathbb{Z}^n \cap \left(\bigcup_{i \in I_0} \mathcal{D}_i\right)$ and it can be obtained by solving (CGP).

Theorem (Gaar, Lee, Ljubić, Sinnl, Tanınmış 2022)

Let $(x^j, y^j) \in \mathbb{Z}^n$ be an extreme point of the convex set Ω_j which belongs to the interior of the bilevel-free set $S(y^j + \Delta y^j)$ for an appropriate Δy^j. Then, there exists a disjunctive cut that separates (x^j, y^j) from $\Omega_j \cap \mathbb{Z}^n \cap \left(\bigcup_{i \in I_0} \mathcal{D}_i\right)$ and it can be obtained by solving (CGP).

$$
\begin{aligned}
\max_{\alpha,\beta,\tau} \quad & \alpha^\top x^j + \beta^\top y^j - \tau \\
\text{s.t.} \quad & \alpha^\top x + \beta^\top y - \tau \leq 0 \quad \text{for all } (x, y) \in \Omega_j \cap \mathbb{Z}^n \cap \left(\bigcup_{i \in I_0} \mathcal{D}_i\right) \\
& ||(\alpha, \beta, \tau)||_1 \leq 1
\end{aligned}
\tag{CGP}
$$

The Cut-Generating Problem

Theorem (Gaar, Lee, Ljubić, Sinnl, Tanınmış 2022)

Let $(x^j, y^j) \in \mathbb{Z}^n$ be an extreme point of the convex set Ω_j which belongs to the interior of the bilevel-free set $S(y^j + \Delta y^j)$ for an appropriate Δy^j. Then, there exists a disjunctive cut that separates (x^j, y^j) from $\Omega_j \cap \mathbb{Z}^n \cap \left(\bigcup_{i \in I_0} \mathcal{D}_i\right)$ and it can be obtained by solving (CGP)*.*

$$
\begin{aligned}
\max_{\alpha,\beta,\tau} \quad & \alpha^\top x^j + \beta^\top y^j - \tau \\
\text{s.t.} \quad & \alpha^\top x + \beta^\top y - \tau \leq 0 \quad \text{for all } (x, y) \in \Omega_j \cap \mathbb{Z}^n \cap \left(\bigcup_{i \in I_0} \mathcal{D}_i\right) \\
& ||(\alpha, \beta, \tau)||_1 \leq 1
\end{aligned}
\tag{CGP}
$$

Problem: integer, nonlinear, and nonconvex

The Relaxed Cut-Generating Problem

1. Solve the relaxed problem

$$\begin{aligned} \max_{\alpha,\beta,\tau} \quad & \alpha^\top x^j + \beta^\top y^j - \tau \\ \text{s.t.} \quad & \alpha^\top \tilde{x} + \beta^\top \tilde{y} - \tau \leq 0 \quad \text{for all } (\tilde{x}, \tilde{y}) \in \mathcal{Z}^k \\ & ||(\alpha, \beta, \tau)||_1 \leq 1 \end{aligned} \tag{RCGP}$$

- $k = 0, 1, 2, \ldots$
- $\mathcal{Z}^k \subseteq \Omega_j \cap \mathbb{Z}^n \cap \left(\bigcup_{i \in I_0} \mathcal{D}_i\right)$ discrete set for all k

The Relaxed Cut-Generating Problem

1. Solve the relaxed problem

$$\begin{aligned} \max_{\alpha,\beta,\tau} \quad & \alpha^\top x^j + \beta^\top y^j - \tau \\ \text{s.t.} \quad & \alpha^\top \tilde{x} + \beta^\top \tilde{y} - \tau \leq 0 \quad \text{for all } (\tilde{x}, \tilde{y}) \in \mathcal{Z}^k \\ & ||(\alpha, \beta, \tau)||_1 \leq 1 \end{aligned} \tag{RCGP}$$

   - $k = 0, 1, 2, \dots$
   - $\mathcal{Z}^k \subseteq \Omega_j \cap \mathbb{Z}^n \cap \left(\bigcup_{i \in I_0} \mathcal{D}_i\right)$ discrete set for all k

2. Check if the solution $(\alpha^k, \beta^k, \tau^k)$ separates all $(x, y) \in \Omega_j \cap \mathbb{Z}^n \cap \left(\bigcup_{i \in I_0} \mathcal{D}_i\right)$, i.e., if

$$\Psi(\alpha^k, \beta^k, \tau^k) := \max_{x,y} \left\{ (\alpha^k)^\top x + (\beta^k)^\top y - \tau^k : (x, y) \in \Omega_j \cap \mathbb{Z}^n \cap \left(\bigcup_{i \in I_0} \mathcal{D}_i \right) \right\} \leq 0$$

Cut Verification I

Solve $m_2 + 1$ subproblems

$$\begin{aligned}
\max_{x^0, y^0} \quad & (\alpha^k)^\top x^0 + (\beta^k)^\top y^0 - \tau^k \\
\text{s.t.} \quad & (x^0, y^0) \in \Omega_j \cap \mathbb{Z}^n \\
& f(x^0, (y^j + \Delta y^j)) - f(x^0, y^0) \geq 0
\end{aligned} \tag{CVP$_0$}$$

Cut Verification I

Solve $m_2 + 1$ subproblems

$$\begin{aligned} \max_{x^0, y^0} \quad & (\alpha^k)^\top x^0 + (\beta^k)^\top y^0 - \tau^k \\ \text{s.t.} \quad & (x^0, y^0) \in \Omega_j \cap \mathbb{Z}^n \\ & f(x^0, (y^j + \Delta y^j)) - f(x^0, y^0) \geq 0 \end{aligned} \tag{CVP$_0$}$$

and

$$\begin{aligned} \max_{x^i, y^i} \quad & (\alpha^k)^\top x^i + (\beta^k)^\top y^i - \tau^k \\ \text{s.t.} \quad & (x^i, y^i) \in \Omega_j \cap \mathbb{Z}^n \\ & g_i(x^i, (y^j + \Delta y^j)) \geq 0 \end{aligned} \tag{CVP$_i$}$$

$\Rightarrow$ obtain solutions $(x^i, y^i)^k$ for $i = 0, \ldots, m_2$

Cut Verification II

After obtaining solutions $(x^i, y^i)^k$ for $i = 0, \ldots, m_2$

- Check if $(\alpha^k)^\top (x^i)^k + (\beta^k)^\top (y^i)^k - \tau^k > 0$ for any $i = 0, 1, \ldots, m_2$

Cut Verification II

After obtaining solutions $(x^i, y^i)^k$ for $i = 0, \ldots, m_2$

- Check if $(\alpha^k)^\top (x^i)^k + (\beta^k)^\top (y^i)^k - \tau^k > 0$ for any $i = 0, 1, \ldots, m_2$
- true: repeat the procedure with

$$\mathcal{Z}^{k+1} \leftarrow \mathcal{Z}^k \cup \left\{ (x^i, y^i)^k : (\alpha^k)^\top (x^i)^k + (\beta^k)^\top (y^i)^k - \tau^k > 0,\ i \in I_0 \right\}$$

Cut Verification II

After obtaining solutions $(x^i, y^i)^k$ for $i = 0, \ldots, m_2$

- Check if $(\alpha^k)^\top (x^i)^k + (\beta^k)^\top (y^i)^k - \tau^k > 0$ for any $i = 0, 1, \ldots, m_2$
- true: repeat the procedure with

$$\mathcal{Z}^{k+1} \leftarrow \mathcal{Z}^k \cup \left\{ (x^i, y^i)^k : (\alpha^k)^\top (x^i)^k + (\beta^k)^\top (y^i)^k - \tau^k > 0,\ i \in I_0 \right\}$$

- false: we have a cutting plane

Cut Verification II

After obtaining solutions $(x^i, y^i)^k$ for $i = 0, \dots, m_2$

- Check if $(\alpha^k)^\top (x^i)^k + (\beta^k)^\top (y^i)^k - \tau^k > 0$ for any $i = 0, 1, \dots, m_2$
- true: repeat the procedure with

$$\mathcal{Z}^{k+1} \leftarrow \mathcal{Z}^k \cup \left\{ (x^i, y^i)^k : (\alpha^k)^\top (x^i)^k + (\beta^k)^\top (y^i)^k - \tau^k > 0,\; i \in I_0 \right\}$$

- false: we have a cutting plane

Lemma

Let f and g_i be jointly convex and linear in $\mathbf{x}$ for all $i \in \{1, \dots, m_2\}$. Then, all subproblems (CVP$_0$) *and* (CVP$_i$) *for $i \in \{1, \dots, m_2\}$ are convex.*

The Cut-Generating Method

Input: An integer-feasible node solution (x^j, y^j), which is bilevel-infeasible

1: Solve the scoop problem for given (x^j, y^j) and obtain Δy^j and t^j.
2: **if** $t^j > 0$ **then**
3: Solve the (RCGP) and obtain a valid inequality parameterized by $(\alpha^k, \beta^k, \tau^k)$.
4: Solve the subproblems (CVP$_0$) and (CVP$_i$) and obtain the solutions $(x^i, y^i)^k$.
5: **if** (CVP$_i$) are infeasible for all $i = 0, \ldots, m_2$ **then**
6: Prune the node N_j.
7: **else if** $(\alpha^k)^\top (x^i)^k + (\beta^k)^\top (y^i)^k - \tau^k \leq 0$ for all $i = 0, \ldots, m_2$ **then**
8: Add the locally valid inequality to the node problem N_j.
9: **else**
10: Update $\mathcal{Z}^k$ and set $k \leftarrow k + 1$. Go to line 3.
11: **end if**
12: **else**
13: Add an integer no-good cut to the node problem N_j.
14: **end if**

Correctness Theorem

The cut-generating procedure of the algorithm applied in a standard branch-and-cut method terminates after a finite number of steps with a globally optimal solution or with a correct indication of infeasibility.

Problem Instances

ILP-IQP bilevel problems of the form

$$
\begin{aligned}
\min_{(x,y)\in\mathbb{Z}^n} \quad & c_x^\top x + c_y^\top y \\
\text{s.t.} \quad & Ax + By \le a \\
& y \in \operatorname*{arg\,min}_{\bar{y}\in\mathbb{Z}^{n_2}} \left\{ \frac{1}{2}\bar{y}^\top Q\bar{y} + d_y^\top \bar{y} \colon (Cx + D\bar{y})_i \le b_i,\ i = 1,\dots,m_2 - 1,\ \frac{1}{2}\bar{y}^\top P\bar{y} + (Cx + D\bar{y})_{m_2} \le b_{m_2} \right\}
\end{aligned}
$$

Problem Instances

ILP-IQP bilevel problems of the form

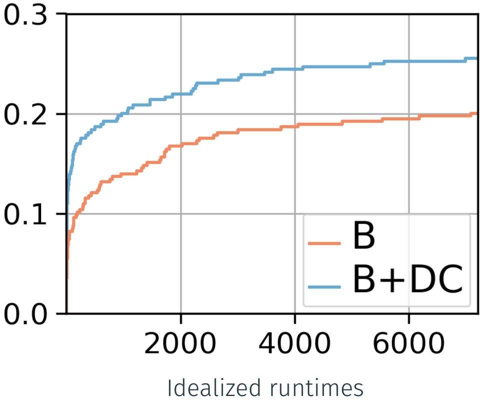
$$
\begin{aligned}
\min_{(x,y)\in\mathbb{Z}^n} \quad & c_x^\top x + c_y^\top y \\
\text{s.t.} \quad & Ax + By \leq a \\
& y \in \arg\min_{\bar{y}\in\mathbb{Z}^{n_2}} \left\{ \frac{1}{2}\bar{y}^\top Q\bar{y} + d_y^\top \bar{y} \colon (Cx + D\bar{y})_i \leq b_i,\ i = 1,\dots,m_2 - 1,\ \frac{1}{2}\bar{y}^\top P\bar{y} + (Cx + D\bar{y})_{m_2} \leq b_{m_2} \right\}
\end{aligned}
$$

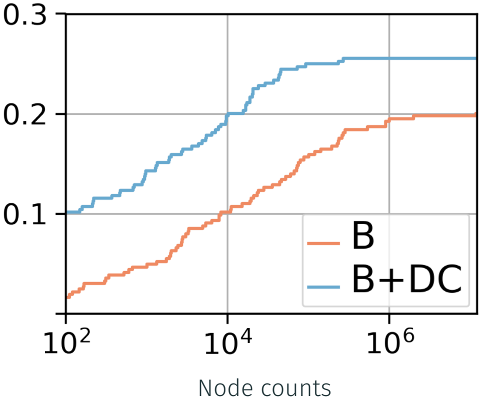
- Subset of the QBMKP instances used in Gaar et al. (2023)
  - With and without quadratic constraint in the lower level
  - Different directions for the upper-level objective function (sim/opp)
  - One lower-level constraint vs. 50% lower-level constraints
  - 1600 Instances
- Subset of the MILP-MILP instance collections by Kleinert and Schmidt (2021)
  - 180 Instances

Computational Setup

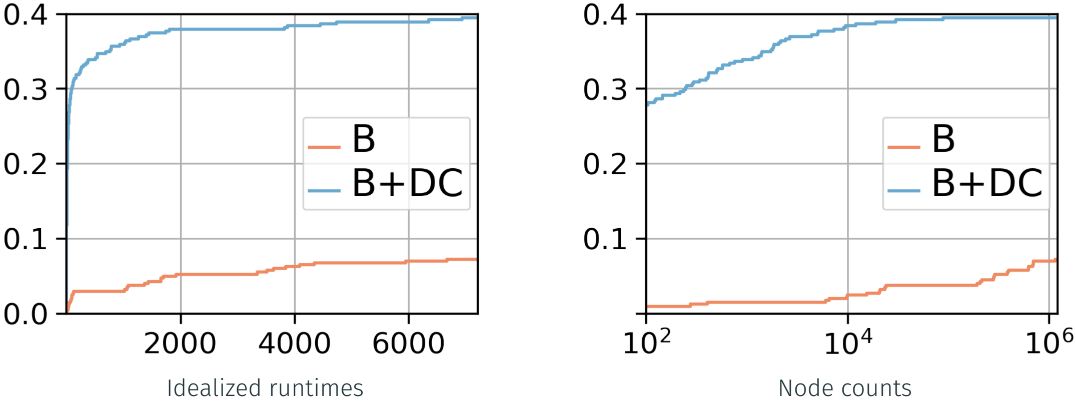
- Python 3.9.7
- CPLEX 22.1.0
  - Presolve, heuristics deactivated
  - Cuts realized with lazy constraint callbacks
- Gurobi 9.5.1
- 4 Intel XEON SP 6126 cores with 2.6 GHz and 32 GB RAM

Numerical Results QBMKP_sim

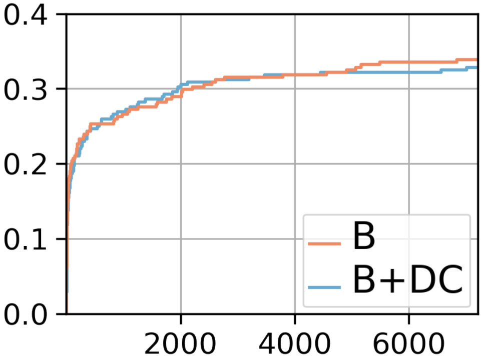




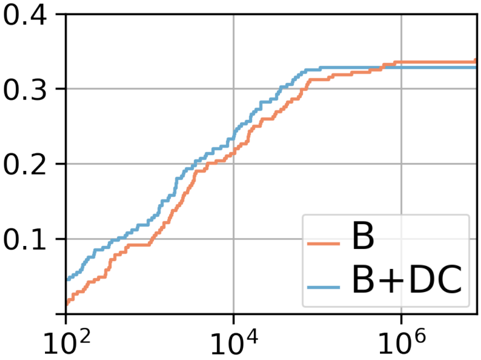
Numerical Results QBMKP_opp



Numerical Results QBMKP_50/50_sim

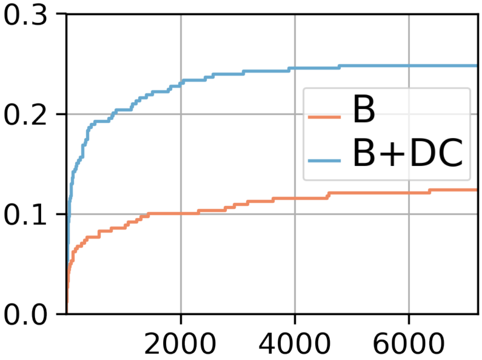


Idealized runtimes

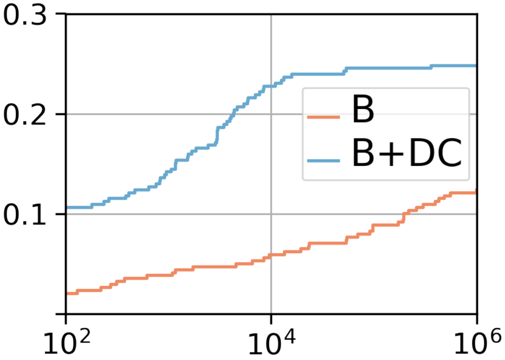


Node counts

Numerical Results QBMKP_50/50_opp



Idealized runtimes



Node counts

An (in my opinion) important open problem

Smart People

$$\begin{aligned} \min_{x,y} \quad & F(x,y) \\ \text{s.t.} \quad & G(x,y) \leq 0 \\ & x \in \mathbb{R}^{n_x}, \quad y \in \mathbb{R}^{n_y} \\ & y \in S(x) \end{aligned}$$

Bilevel Optimization

$$\begin{aligned} \min_{x,y} \quad & F(x, y) \\ \text{s.t.} \quad & G(x, y) \leq 0 \\ & x \in \mathbb{R}^{n_x}, \quad y \in \mathbb{R}^{n_y} \\ & y \in S(x) \end{aligned}$$

$S(x)$: solution set of the convex lower-level problem

$$S(x) = \arg\min_{y} \left\{ f(x, y) \colon g(x, y) \leq 0, \; y \in \mathbb{R}^{n_y} \right\}$$

Once upon a time in multilevel gas market optimization ...

A "small" extension: black-box constraint in the lower level

$$S(x) = \arg\min_{y} \left\{ f(x,y) \colon g(x,y) \le 0,\ b(y) \le 0,\ y \in \mathbb{R}^{n_y} \right\}$$

Once upon a time in multilevel gas market optimization ...

A “small” extension: black-box constraint in the lower level

$$S(x) = \arg\min_y \left\{ f(x,y) \colon g(x,y) \leq 0,\ b(y) \leq 0,\ y \in \mathbb{R}^{n_y} \right\}$$

Assumption

The black-box function b is convex and for all $(x,y) \in \{(x,y) \colon G(x,y) \leq 0,\ g(x,y) \leq 0\}$, ...

1. we can evaluate the function $b(y)$,
2. we can evaluate the gradient $\nabla b(y)$,
3. the gradient is bounded, i.e., $\|\nabla b(y)\| \leq K$ for a fixed $K \in \mathbb{R}$.

A "First-Relax-Then-Reformulate" Approach

- Block-box constraint $b(y) \leq 0$ is convex
- Construct a sequence of linear outer approximations $(E^r, e^r)_{r \in \mathbb{N}}$ of the black-box constraint $b(y) \leq 0$ with the property

$$\{y \in \mathbb{R}^{n_y} : b(y) \leq 0\} \subseteq \{y \in \mathbb{R}^{n_y} : E^{r+1} y \leq e^{r+1}\} \subseteq \{y \in \mathbb{R}^{n_y} : E^r y \leq e^r\}$$

A "First-Relax-Then-Reformulate" Approach

- Block-box constraint $b(y) \leq 0$ is convex
- Construct a sequence of linear outer approximations $(E^r, e^r)_{r \in \mathbb{N}}$ of the black-box constraint $b(y) \leq 0$ with the property

$$\{y \in \mathbb{R}^{n_y} : b(y) \leq 0\} \subseteq \{y \in \mathbb{R}^{n_y} : E^{r+1} y \leq e^{r+1}\} \subseteq \{y \in \mathbb{R}^{n_y} : E^r y \leq e^r\}$$

- For a given upper-level solution $\bar{x} \in \Omega_{\mathrm{u}}$ and $r \in \mathbb{N}$, the adapted lower-level problem reads

$$\min_{y \in \mathbb{R}^{n_y}} \quad f(\bar{x}, y) \quad \text{s.t.} \quad g(\bar{x}, y) \leq 0, \; E^r y \leq e^r$$

- This is a relaxation of the original lower-level problem

A "First-Relax-Then-Reformulate" Approach

- Block-box constraint $b(y) \leq 0$ is convex
- Construct a sequence of linear outer approximations $(E^r, e^r)_{r \in \mathbb{N}}$ of the black-box constraint $b(y) \leq 0$ with the property

$$\{y \in \mathbb{R}^{n_y} : b(y) \leq 0\} \subseteq \{y \in \mathbb{R}^{n_y} : E^{r+1} y \leq e^{r+1}\} \subseteq \{y \in \mathbb{R}^{n_y} : E^r y \leq e^r\}$$

- For a given upper-level solution $\bar{x} \in \Omega_{\mathrm{u}}$ and $r \in \mathbb{N}$, the adapted lower-level problem reads

$$\min_{y \in \mathbb{R}^{n_y}} \quad f(\bar{x}, y) \quad \text{s.t.} \quad g(\bar{x}, y) \leq 0, \; E^r y \leq e^r$$

- This is a relaxation of the original lower-level problem
- $\underline{\varphi}^r(x)$: optimal value function
- Assumption: Slater's constraint qualification holds

A "First-Relax-Then-Reformulate" Approach

- Block-box constraint $b(y) \leq 0$ is convex
- Construct a sequence of linear outer approximations $(E^r, e^r)_{r \in \mathbb{N}}$ of the black-box constraint $b(y) \leq 0$ with the property

$$\{y \in \mathbb{R}^{n_y} : b(y) \leq 0\} \subseteq \{y \in \mathbb{R}^{n_y} : E^{r+1} y \leq e^{r+1}\} \subseteq \{y \in \mathbb{R}^{n_y} : E^r y \leq e^r\}$$

- For a given upper-level solution $\bar{x} \in \Omega_{\mathrm{u}}$ and $r \in \mathbb{N}$, the adapted lower-level problem reads

$$\min_{y \in \mathbb{R}^{n_y}} \quad f(\bar{x}, y) \quad \text{s.t.} \quad g(\bar{x}, y) \leq 0, \; E^r y \leq e^r$$

- This is a relaxation of the original lower-level problem
- $\underline{\varphi}^r(x)$: optimal value function
- Assumption: Slater's constraint qualification holds

Proposition

For every $r \in \mathbb{N}$ and every upper-level decision $x \in \Omega_{\mathrm{u}}$, it holds

$$\underline{\varphi}^r(x) \leq \underline{\varphi}^{r+1}(x) \leq \varphi(x).$$

“First-Relax-Then-Reformulate”

Choose $\delta_b > 0$, set $r = 0$, $s = 0$, $\chi = \infty$, $E^0 = [0 \dots 0] \in \mathbb{R}^{1 \times n_y}$, $e^0 = 0 \in \mathbb{R}$.
while $\chi > \delta_b$ or $s > 0$ **do**
Construct E^{r+1} and e^{r+1}.
if the modified variant of the single-level reformulation is feasible **then**
Solve this problem to obtain (x^{r+1}, y^{r+1}) and set $s = 0$.
else if the feasibility problem is feasible **then**
Solve this problem to obtain (x^{r+1}, y^{r+1}, s).
else
Return “The original problem is infeasible.”
end if
Set $r \leftarrow r + 1$ and $\chi = b(y^r)$.
end while
Return $(\bar{x}, \bar{y}) = (x^r, y^r)$.

Choose $\delta_b > 0$, set $r = 0$, $s = 0$, $\chi = \infty$, $E^0 = [0 \dots 0] \in \mathbb{R}^{1 \times n_y}$, $e^0 = 0 \in \mathbb{R}$.
while $\chi > \delta_b$ or $s > 0$ **do**
  Construct E^{r+1} and e^{r+1}.
  if the modified variant of the single-level reformulation is feasible **then**
    Solve this problem to obtain (x^{r+1}, y^{r+1}) and set $s = 0$.
  else if the feasibility problem is feasible **then**
    Solve this problem to obtain (x^{r+1}, y^{r+1}, s).
  else
    Return "The original problem is infeasible."
  end if
  Set $r \leftarrow r + 1$ and $\chi = b(y^r)$.
end while
Return $(\bar{x}, \bar{y}) = (x^r, y^r)$.

Theorem: If the algorithm terminates, then $(\bar{x}, \bar{y})$ is $(0, 0, \delta_b, 0)$-feasible for original bilevel problem.

All the details ...

Open Access | Published: 13 May 2022

On convex lower-level black-box constraints in bilevel optimization with an application to gas market models with chance constraints

Holger Heitsch, René Henrion, Thomas Kleinert & Martin Schmidt

Journal of Global Optimization **84**, 651–685 (2022) | Cite this article

Abstract

Bilevel optimization is an increasingly important tool to model hierarchical decision making. However, the ability of modeling such settings makes bilevel problems hard to solve in theory and practice. In this paper, we add on the general difficulty of this class of problems by further incorporating convex black-box constraints in the lower level. For this setup, we develop a cutting-plane algorithm that computes approximate bilevel-feasible points. We apply this method to a bilevel model of the European gas market in which we use a joint chance constraint to model uncertain loads. Since the chance constraint is not available in closed form, this fits into the black-box setting studied before. For the applied model, we use further problem-specific insights to derive bounds on the objective value of the bilevel problem. By doing so, we are able to show that we solve the application problem to approximate global optimality. In our numerical case study we are thus able to evaluate the welfare sensitivity in dependence of the achieved safety level of uncertain load coverage.

Nonconvexities in the Lower Level

Upper-level problem

$$\begin{aligned} \text{“}\min_x\text{”} \quad & F(x,y) \\ \text{s.t.} \quad & G(x,y) \geq 0, \quad y \in S(x) \end{aligned}$$

Lower-level problem

$$\begin{aligned} \min_y \quad & f(x,y) \\ \text{s.t.} \quad & g(x,y) \geq 0 \end{aligned}$$

Who can solve this problem?

Upper level

$$\begin{aligned} \max_{x \in \mathbb{R}^2} \quad & F(x,y) = x_1 - 2y_{n+1} + y_{n+2} \\ \text{s.t.} \quad & (x_1, x_2) \in [\underline{x}_1, \bar{x}_1] \times [\underline{x}_2, \bar{x}_2] \\ & y \in S(x) \end{aligned}$$

- $\underline{x}$, $\bar{x} \in \mathbb{R}^2$ with $1 \leq \underline{x}_i < \bar{x}_i$, $i \in \{1, 2\}$
- Upper level is an LP with simple bound constraints
- Upper level has no coupling constraints

Who can solve this problem?

Upper level

$$\begin{aligned} \max_{x \in \mathbb{R}^2} \quad & F(x,y) = x_1 - 2y_{n+1} + y_{n+2} \\ \text{s.t.} \quad & (x_1, x_2) \in [\underline{x}_1, \bar{x}_1] \times [\underline{x}_2, \bar{x}_2] \\ & y \in S(x) \end{aligned}$$

Lower level

$$\begin{aligned} \max_{y \in \mathbb{R}^{n+2}} \quad & f(x,y) = y_1 - y_n\,(x_1 + x_2 - y_{n+1} - y_{n+2}) \\ \text{s.t.} \quad & y_1 + y_n = \frac{1}{2} \\ & y_i^2 \le y_{i+1}, \quad i \in \{1, \dots, n-1\} \\ & y_i \ge 0, \quad i \in \{1, \dots, n\} \\ & y_{n+1} \in [0, x_1] \\ & y_{n+2} \in [-x_2, x_2] \end{aligned}$$

- $\underline{x}$, $\bar{x} \in \mathbb{R}^2$ with $1 \le \underline{x}_i < \bar{x}_i$, $i \in \{1, 2\}$
- Upper level is an LP with simple bound constraints
- Upper level has no coupling constraints
- Feasible set of lower level is non-empty and compact for all feasible leader decisions
- Slater's CQ is also satisfied for all feasible leader decisions
- All constraints are linear except for some convex-quadratic inequality constraints
- The coefficients/right-hand sides are either 0, 1, or 1/2
- Bilinear objective function

Exact Feasibility

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

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

$$y_i^2 \le y_{i+1}, \quad i \in \{1, \dots, n-1\}$$

$$y_i \ge 0, \quad i \in \{1, \dots, n\}$$

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

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

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 $y_n > 0$.

Exact Feasibility

$$
\begin{aligned}
\max_{y \in \mathbb{R}^{n+2}} \quad & f(x,y) = y_1 - y_n \left(x_1 + x_2 - y_{n+1} - y_{n+2}\right) \\
\text{s.t.} \quad & y_1 + y_n = \frac{1}{2} \\
& y_i^2 \leq y_{i+1}, \quad i \in \{1, \dots, n-1\} \\
& y_i \geq 0, \quad i \in \{1, \dots, n\} \\
& y_{n+1} \in [0, x_1] \\
& y_{n+2} \in [-x_2, x_2]
\end{aligned}
$$

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 $y_n > 0$.

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 is a singleton.

Exact Feasibility

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

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

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 $y_n > 0$.

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 is a singleton.

Result #3

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

ε-feasibility

Definition

Let $0 < \varepsilon \in \mathbb{R}$, $f : \mathbb{R}^n \to \mathbb{R}$, and $g : \mathbb{R}^n \to \mathbb{R}^m$ be given. A point $x \in \mathbb{R}^n$ is called ε-feasible for the problem $\max_{x \in \mathbb{R}^n} \{f(x) \colon g(x) \leq 0\}$ if $g_i(x) \leq 0$ holds for all $i \in \{1, \ldots, m\} \setminus N$ and if $\max_{i \in N} g_i(x) \leq \varepsilon$ holds, where $N \subseteq \{1, \ldots, m\}$ denotes the index set of all nonlinear constraints.

Definition

Let $0 < \varepsilon \in \mathbb{R}$, $f : \mathbb{R}^n \to \mathbb{R}$, and $g : \mathbb{R}^n \to \mathbb{R}^m$ be given. A point $x \in \mathbb{R}^n$ is called ε-feasible for the problem $\max_{x \in \mathbb{R}^n} \{f(x) \colon g(x) \leq 0\}$ if $g_i(x) \leq 0$ holds for all $i \in \{1, \ldots, m\} \setminus N$ and if $\max_{i \in N} g_i(x) \leq \varepsilon$ holds, where $N \subseteq \{1, \ldots, m\}$ denotes the index set of all nonlinear constraints.

Result #4

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

ε-feasibility

Definition

Let $0 < \varepsilon \in \mathbb{R}$, $f : \mathbb{R}^n \to \mathbb{R}$, and $g : \mathbb{R}^n \to \mathbb{R}^m$ be given. A point $x \in \mathbb{R}^n$ is called ε-feasible for the problem $\max_{x \in \mathbb{R}^n} \{f(x) \colon g(x) \leq 0\}$ if $g_i(x) \leq 0$ holds for all $i \in \{1, \dots, m\} \setminus N$ and if $\max_{i \in N} g_i(x) \leq \varepsilon$ holds, where $N \subseteq \{1, \dots, m\}$ denotes the index set of all nonlinear constraints.

Result #4

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

Result #5

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

Result #3 (revisited)

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

ε-feasibility

Result #3 (revisited)

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

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 is given by $x_\text{o}^* = (\bar{x}_1, \bar{x}_2)$ with an optimal objective function value of $F_\text{o}^* = \bar{x}_1 + \bar{x}_2$.

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

Result #3 (revisited)

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

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 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 $F_p^* = -\underline{x}_1 - \underline{x}_2$.

- By the way: $n \geq \log_2(\log_2(1/\varepsilon^2))$
- For $\varepsilon = 10^{-8}$, the problem gets unsolvable for $n = 6$

Well ... and now?

Is this an impossibility result
for computationally solving bilevel problems
with continuous and nonconvex lower-level problems?

Well ... and now?

Is this an impossibility result
for computationally solving bilevel problems
with continuous and nonconvex lower-level problems?