A Gentle and Incomplete Introduction to Bilevel Optimization
… and Some New Results

Martin Schmidt

TRR 154 Autumn School on Equilibrium Problems
Berlin — November 6, 2025

Agenda

1. What is bilevel optimization anyway?

2. Some theory of linear bilevel problems

3. How do you solve a linear bilevel problem?

4. Connections between robust and bilevel optimization

5. The burial of coupling constraints in linear bilevel optimization

Background Material

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 bank 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

Anti Drug Smuggling

- Graph models the network of drug smuggling routes
- Smugglers want to maximize the flow of drugs from an origin to a destination

Anti Drug Smuggling

- Graph models the network of drug smuggling routes
- Smugglers want to maximize the flow of drugs from an origin to a destination

- Follower: maximum flow (= amount of drugs) problem
- Leader: Interdiction of certain parts of the drug smuggling routes
- Goal of the leader: minimize the maximum flow
- Leader only has a certain budget
- … and maybe incomplete information about the follower’s problem

Anti Drug Smuggling

Canada and the Transcontinental Drug Links

Strategic Forecasting Inc
go to original

Email Page Print Page Email Us

Canadian police conducted several simultaneous raids on suspected drug traffickers in Newfoundland and Quebec provinces Oct. 11, arresting two dozen people and seizing marijuana, cocaine, weapons, cash and property. The drug-trafficking ring, which Canadian authorities believe was operated by the Quebec-based Hell's Angels motorcycle/crime gang, could have smuggled the cocaine into Canada from South America via Mexico and the United States.

The jungles of South America, where cocaine is produced, seem a long way from the St. Lawrence River. Using a sophisticated shipment and distribution network, however, criminal and militant organizations can cover the distance in a few days.

More than 70 members of the Royal Newfoundland Constabulary and Quebec's Provincial Biker Enforcement Unit carried out the raids, which represented the culmination of an 18-monthlong investigation dubbed Operation Roadrunner. The arrests were made near St. John's in Newfoundland and near the towns of Laval and La Tuque in Quebec. In Newfoundland, authorities seized $300,000 in cash, 51 pounds of marijuana and 19 pounds of cocaine, as well as vehicles, weapons and computers. In Quebec, $170,000 and four houses were seized.

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 to 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 ...

$$S(x) = \operatorname*{arg\,min}_{y \in Y} \{f(x, y) : g(x, y) \geq 0\}$$

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 ...

$$S(x) = \operatorname*{arg\,min}_{y \in Y} \{f(x, y) : g(x, y) \geq 0\}$$

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 often 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 ...

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

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

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

is called the shared constraint set.

Its projection onto the x-space is denoted by

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

Definition

The set

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

is called the bilevel feasible set or inducible region.

Single-level 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 single-level relaxation (SLR) of the bilevel problem.

Single-level 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 single-level relaxation (SLR) of the bilevel problem.

Remark

- The single-level 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)

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

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

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
- 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

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

- 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

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

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

Anti Drug Smuggling Revisited

Follower: w-parameterized maximum flow problem

$$
\begin{aligned}
\varphi(w) := \max_{f \in \mathbb{R}^{|A|}} \quad & \sum_{a \in \delta^{\text{out}}(s)} f_a - \sum_{a \in \delta^{\text{in}}(s)} f_a \\
\text{s.t.} \quad & \sum_{a \in \delta^{\text{out}}(v)} f_a - \sum_{a \in \delta^{\text{in}}(v)} f_a = 0, \quad v \in V \setminus \{s, t\} \\
& f_a \leq c_a(1 - w_a), \quad a \in A \\
& f_a \geq 0, \quad a \in A
\end{aligned}
$$

Anti Drug Smuggling Revisited

Follower: w-parameterized maximum flow problem

$$
\begin{aligned}
\varphi(w) := \max_{f \in \mathbb{R}^{|A|}} \quad & \sum_{a \in \delta^{\text{out}}(s)} f_a - \sum_{a \in \delta^{\text{in}}(s)} f_a \\
\text{s.t.} \quad & \sum_{a \in \delta^{\text{out}}(v)} f_a - \sum_{a \in \delta^{\text{in}}(v)} f_a = 0, \quad v \in V \setminus \{s, t\} \\
& f_a \leq c_a(1 - w_a), \quad a \in A \\
& f_a \geq 0, \quad a \in A
\end{aligned}
$$

Leader: Maximum flow interdiction

$$
\begin{aligned}
\min_{w \in \{0,1\}^{|A|}} \quad & \varphi(w) \\
\text{s.t.} \quad & \sum_{a \in A} w_a \leq B
\end{aligned}
$$

An Academic and Linear Example (Kleinert 2021)

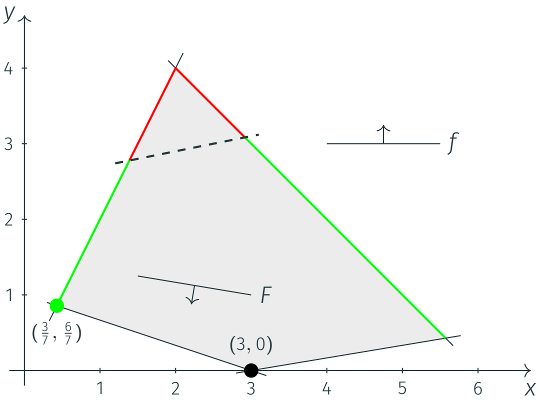
Upper-level problem

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

Lower-level problem

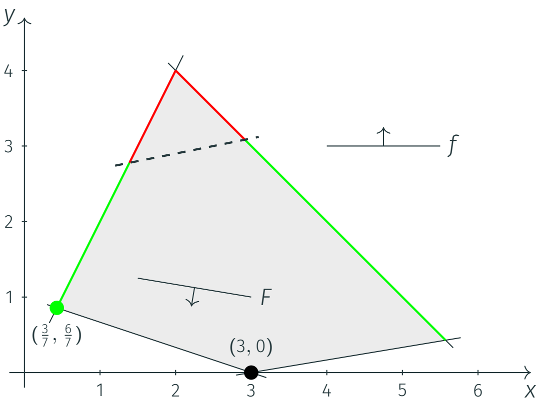
$$
\begin{aligned}
\min_{y} \quad & f(x,y) = -y \\
\text{s.t.} \quad & 2x - y \geq 0 \\
& -x - y \geq -6 \\
& -x + 6y \geq -3 \\
& x + 3y \geq 3
\end{aligned}
$$

An Academic and Linear Example (Kleinert 2021)



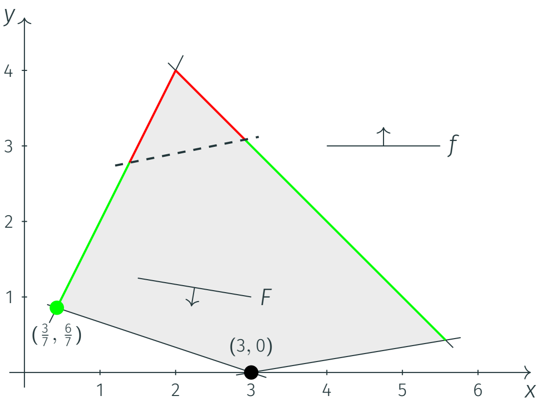
- Shared constrained set: gray area
- Green and red lines: nonconvex set of optimal follower solutions (lifted to the x-y-space)
- Green lines: Nonconvex and disconnected bilevel feasible set of the bilevel problem

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)

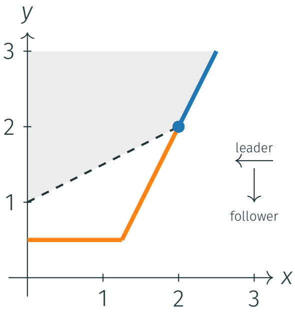


5. The single leader constraint (dashed line) renders certain optimal responses of the follower infeasible.
6. The bilevel feasible region $\mathcal{F}$ corresponds to the green facets.
7. Thus, the feasible set is not only nonconvex but also disconnected.
8. The optimal solution is $(3/7, 6/7)$ with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

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

Optimal solution: $(2, 2)$

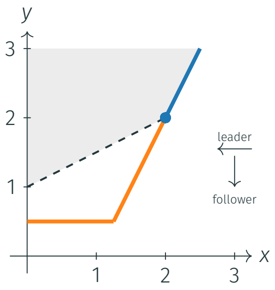


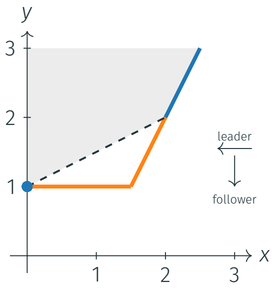
Independence of irrelevant constraints (Kleinert et al. 2021; Macal and Hurter 1997)

- 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 $\bar{y} \geq 1$
- Single-level 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 of 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

Some theory of linear bilevel problems

The linear bilevel problem

We now consider LP-LP bilevel problems of the form

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

with $c_x \in \mathbb{R}^{n_x}$, $c_y, d \in \mathbb{R}^{n_y}$, $A \in \mathbb{R}^{m \times n_x}$, and $a \in \mathbb{R}^m$ as well as $C \in \mathbb{R}^{\ell \times n_x}$, $D \in \mathbb{R}^{\ell \times n_y}$, and $b \in \mathbb{R}^\ell$.

We now consider LP-LP bilevel problems of the form

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

with $c_x \in \mathbb{R}^{n_x}$, $c_y, d \in \mathbb{R}^{n_y}$, $A \in \mathbb{R}^{m \times n_x}$, and $a \in \mathbb{R}^m$ as well as $C \in \mathbb{R}^{\ell \times n_x}$, $D \in \mathbb{R}^{\ell \times n_y}$, and $b \in \mathbb{R}^\ell$.

Remark

This problem does not contain coupling constraints to avoid the further difficulties that arise due to disconnected bilevel feasible sets.

The first structural result

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

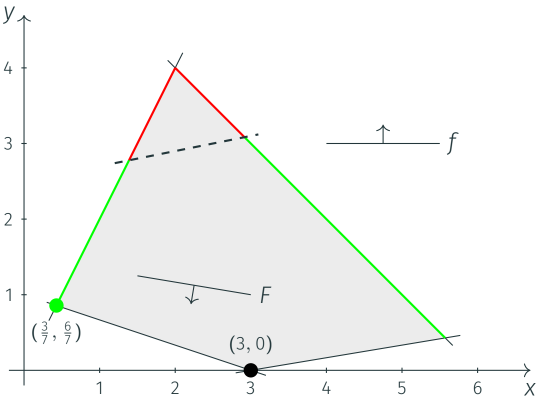
The first structural result

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

Theorem

Suppose that the shared constraint set is non-empty and bounded. The bilevel-feasible set can then be equivalently written as the intersection of the shared constraint set with the feasible points of a piecewise linear equality constraint. In particular, the bilevel-feasible set is a union of faces of the shared constraint set.

The Academic Example Revisited



The first structural result: proof

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

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

The first structural result: proof

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

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

explicitly as

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

and use the optimal-value function

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

again.

The first structural result: proof

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

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

explicitly as

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

and use the optimal-value function

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

again.

By using the strong-duality theorem, we can also express the optimal-value function by means of the dual LP as

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

The first structural result: proof

From the classic theory of linear optimization we know that the optimal solution is attained in one of the vertices of the feasible set,

The first structural result: proof

From the classic theory of linear optimization we know that the optimal solution is attained in one of the vertices of the feasible set, which, for the dual LP, does not depend on the leader's decision x anymore.

The first structural result: proof

From the classic theory of linear optimization we know that the optimal solution is attained in one of the vertices of the feasible set, which, for the dual LP, does not depend on the leader's decision x anymore.

Let $\lambda^1, \ldots, \lambda^s$ be the set of all the dual polyhedron's vertices, i.e., the set of vertices of the polyhedron defined by

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

The first structural result: proof

From the classic theory of linear optimization we know that the optimal solution is attained in one of the vertices of the feasible set, which, for the dual LP, does not depend on the leader's decision x anymore.

Let $\lambda^1, \dots, \lambda^s$ be the set of all the dual polyhedron's vertices, i.e., the set of vertices of the polyhedron defined by

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

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

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

The first structural result: proof

From the classic theory of linear optimization we know that the optimal solution is attained in one of the vertices of the feasible set, which, for the dual LP, does not depend on the leader's decision x anymore.

Let $\lambda^1, \dots, \lambda^s$ be the set of all the dual polyhedron's vertices, i.e., the set of vertices of the polyhedron defined by

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

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

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

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

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

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

The first structural result: proof

Consider now again the definition of the optimal-value function using the vertices of the dual polyhedron of the lower-level problem.

The first structural result: proof

Consider now again the definition of the optimal-value function using the vertices of the dual polyhedron of the lower-level problem.

Suppose that for a given x the corresponding solution is the vertex λ^k.

Consider now again the definition of the optimal-value function using the vertices of the dual polyhedron of the lower-level problem.

Suppose that for a given x the corresponding solution is the vertex λ^k.

By using dual feasibility, we obtain

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

The first structural result: proof

Consider now again the definition of the optimal-value function using the vertices of the dual polyhedron of the lower-level problem.

Suppose that for a given x the corresponding solution is the vertex λ^k.

By using dual feasibility, we obtain

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

Thus, for those λ_i^k, $i \in \{1, \ldots, \ell\}$, with $\lambda_i^k > 0$ we get $(Cx + Dy - b)_i = 0$.

The first structural result: proof

Consider now again the definition of the optimal-value function using the vertices of the dual polyhedron of the lower-level problem.

Suppose that for a given x the corresponding solution is the vertex λ^k.

By using dual feasibility, we obtain

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

Thus, for those λ_i^k, $i \in \{1, \dots, \ell\}$, with $\lambda_i^k > 0$ we get $(Cx + Dy - b)_i = 0$.

Hence, the bilevel-feasible set is a union of faces of the shared constraint set.

In other words ...

Corollary

Suppose that the assumptions of the last theorem hold. Then, the LP-LP bilevel problem is equivalent to minimizing the upper-level's objective function over the intersection of the shared constraint set with a piecewise linear equality constraint.

In other words ...

Corollary

Suppose that the assumptions of the last theorem hold. Then, the LP-LP bilevel problem is equivalent to minimizing the upper-level's objective function over the intersection of the shared constraint set with a piecewise linear equality constraint.

Corollary

Suppose that the assumptions of the last theorem hold. Then, a solution of the LP-LP bilevel problem is always attained at a vertex of the bilevel-feasible set.

Solutions appear at vertices of the HPR

Theorem

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

Solutions appear at vertices of the HPR: proof

Let $(x^1, y^1), \dots, (x^r, y^r)$ be the distinct vertices of the shared constraint set Ω.

Solutions appear at vertices of the HPR: proof

Let $(x^1, y^1), \dots, (x^r, y^r)$ be the distinct vertices of the shared constraint set Ω.

Since Ω is a convex polyhedron, any point in Ω can be written as a convex combination of these vertices, i.e.,

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

with

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

Solutions appear at vertices of the HPR: proof

Let $(x^1, y^1), \dots, (x^r, y^r)$ be the distinct vertices of the shared constraint set Ω.

Since Ω is a convex polyhedron, any point in Ω can be written as a convex combination of these vertices, i.e.,

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

with

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

From the proof of the last theorem it follows that the optimal-value function φ is convex and continuous.

Solutions appear at vertices of the HPR: proof

Since the bilevel solution (x^*, y^*) is, of course, bilevel feasible, we obtain

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

Solutions appear at vertices of the HPR: proof

Since the bilevel solution (x^*, y^*) is, of course, bilevel feasible, we obtain

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

By the definition of the optimal-value function we also have

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

This implies $d^\top y^i - \varphi(x^i) \geq 0$.

Solutions appear at vertices of the HPR: proof

Consequently, for all $i \in \{1, \ldots, r\}$ with $\alpha_i > 0$ it holds $d^\top y^i = \varphi(x^i)$ since we otherwise get a contradiction on the last slide.

Solutions appear at vertices of the HPR: proof

Consequently, for all $i \in \{1, \dots, r\}$ with $\alpha_i > 0$ it holds $d^\top y^i = \varphi(x^i)$ since we otherwise get a contradiction on the last slide.

Hence, for those i with $\alpha_i > 0$ we obtain $(x^i, y^i) \in \mathcal{F}$.

Solutions appear at vertices of the HPR: proof

Consequently, for all $i \in \{1, \dots, r\}$ with $\alpha_i > 0$ it holds $d^\top y^i = \varphi(x^i)$ since we otherwise get a contradiction on the last slide.

Hence, for those i with $\alpha_i > 0$ we obtain $(x^i, y^i) \in \mathcal{F}$.

From the last corollary we know that (x^*, y^*) is a vertex of the bilevel-feasible set. Suppose now that there are two indices i and j with $\alpha_i > 0$ and $\alpha_j > 0$.

Solutions appear at vertices of the HPR: proof

Consequently, for all $i \in \{1, \ldots, r\}$ with $\alpha_i > 0$ it holds $d^\top y^i = \varphi(x^i)$ since we otherwise get a contradiction on the last slide.

Hence, for those i with $\alpha_i > 0$ we obtain $(x^i, y^i) \in \mathcal{F}$.

From the last corollary we know that (x^*, y^*) is a vertex of the bilevel-feasible set. Suppose now that there are two indices i and j with $\alpha_i > 0$ and $\alpha_j > 0$.

Thus, $(x^i, y^i) \in \mathcal{F}$ and $(x^j, y^j) \in \mathcal{F}$ holds and we can write (x^*, y^*) as a proper convex combination of two bilevel feasible points, which is a contradiction to the last corollary.

Solutions appear at vertices of the HPR: proof

Consequently, for all $i \in \{1, \dots, r\}$ with $\alpha_i > 0$ it holds $d^\top y^i = \varphi(x^i)$ since we otherwise get a contradiction on the last slide.

Hence, for those i with $\alpha_i > 0$ we obtain $(x^i, y^i) \in \mathcal{F}$.

From the last corollary we know that (x^*, y^*) is a vertex of the bilevel-feasible set. Suppose now that there are two indices i and j with $\alpha_i > 0$ and $\alpha_j > 0$.

Thus, $(x^i, y^i) \in \mathcal{F}$ and $(x^j, y^j) \in \mathcal{F}$ holds and we can write (x^*, y^*) as a proper convex combination of two bilevel feasible points, which is a contradiction to the last corollary.

Thus, (x^*, y^*) is a vertex of the shared constraint set.

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

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

- 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

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

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

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

- ...
- 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

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

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.

These two cases can be modeled using binary variables

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

in the following mixed-integer linear way:

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

Here, M is a sufficiently large constant.

How to solve the KKT reformulation?

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

View PDF

Tools Share

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.

Connections between robust and bilevel optimization

Question:

“What are the best and the worst aspects of this seminar?”

Question:

“What are the best and the worst aspects of this seminar?”

Answer (by one of the participants):

“Best: Very active audience with a lot of questions.

Question:

"What are the best and the worst aspects of this seminar?"

Answer (by one of the participants):

"Best: Very active audience with a lot of questions.

Worst: Found out that robust optimization is just a special case of bilevel optimization"

Later on Twitter

Jannis Kurtz
@JannisK13

I asked #chatgpt if #robustoptimization is a special case of #bileveloptimization .

Tweet übersetzen

JA Is robust optimization a special case of bilevel optimization?

No, robust optimization is not a special case of bilevel optimization. Robust optimization and bilevel optimization are two distinct types of optimization techniques that are used to solve different types of problems.

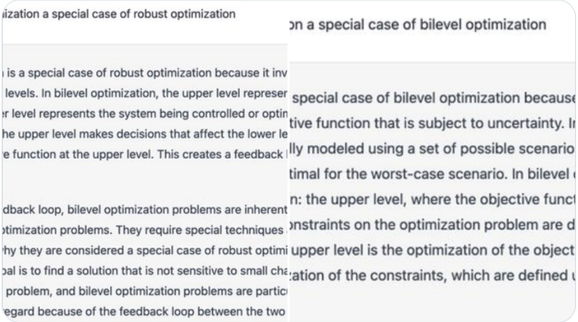
Robust optimization is a method of optimization that seeks to find solutions that are not overly sensitive to changes in the inputs or assumptions of the problem. This is achieved by introducing a certain degree of flexibility or "robustness" into the optimization process, allowing the solution to remain optimal even under a range of different conditions.

Bilevel optimization, on the other hand, is a type of optimization problem that involves two

Later on Twitter

Ryan Cory-Wright @RyanCoryWright · 9. Dez. 2022

🤔



3 5

Jannis Kurtz @JannisK13 · 9. Dez. 2022

Haha ok i think we solved P=NP now

2

Later on Twitter

R

why is bilevel optimization a special case of robust optimization

Bilevel optimization is a special case of robust optimization because it involves optimizing a problem with two levels. In bilevel optimization, the upper level represents the decision maker and the lower level represents the system being controlled or optimized. The decision maker at the upper level makes decisions that affect the lower level, which in turn affects the objective function at the upper level. This creates a feedback loop between the two levels.

Because of this feedback loop, bilevel optimization problems are inherently more complex than single-level optimization problems. They require special techniques and algorithms to solve, which is why they are considered a special case of robust optimization. In robust optimization, the goal is to find a solution that is not sensitive to small changes or uncertainties in the problem, and bilevel optimization problems are particularly challenging in this regard because of the feedback loop between the two levels.

Later on Twitter

R

why is robust optimization a special case of bilevel optimization

Robust optimization is a special case of bilevel optimization because it involves the optimization of an objective function that is subject to uncertainty. In robust optimization, the uncertainty is typically modeled using a set of possible scenarios, and the goal is to find a solution that is optimal for the worst-case scenario. In bilevel optimization, there are two levels of optimization: the upper level, where the objective function is defined, and the lower level, where the constraints on the optimization problem are defined. In the case of robust optimization, the upper level is the optimization of the objective function, and the lower level is the optimization of the constraints, which are defined using the set of possible scenarios.

We wanted to find out!

Marc Goerigk

Jannis Kurtz

Johannes Thürauf

Optimistic Bilevel Optimization (Dempe 2002)

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

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

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

Pessimistic Bilevel Optimization (Wiesemann et al. 2013)

$$\begin{aligned} \min_{x} \quad & \max_{y \in S(x)} F(x, y) \\ \text{s.t.} \quad & G(x, y') \leq 0 \quad \forall y' \in S(x) \end{aligned}$$

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

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

Robust Optimization

$$\begin{aligned} \min_x \quad & H(x) \\ \text{s.t.} \quad & h_i(x, u_i) \le 0 \quad \forall i \in I,\ \forall u_i \in U_i \\ & h_j(x) \le 0 \quad \forall j \in J \end{aligned}$$

Important Case

Decision-dependent uncertainty: $U_i = U_i(x)$

Standard Assumptions

- For every robust feasible point x and every $i \in I$, the constraint functions $h_i(x, \cdot)$ are continuous
- For $i \in I$, the uncertainty set $U_i(x)$ is non-empty and compact

Main Research Question

If P is an instance of problem class $\mathcal{P}$

and if A is an algorithm for solving instances of problem class $\mathcal{Q}$,

can then A also be used to solve P?

Main Research Question

If P is an instance of problem class $\mathcal{P}$

and if A is an algorithm for solving instances of problem class $\mathcal{Q}$,

can then A also be used to solve P?

One can use an algorithm A

for solving optimistic bilevel optimization problems $\mathcal{Q}$

for solving a strictly robust optimization problem P.

Bilevel Methods can Solve Decision-Dependent Robust Problems

Theorem

Let the standard assumptions be satisfied. Let further (x^*, u^*) be a solution to the optimistic bilevel problem

$$\begin{aligned} \min_{x,u} \quad & H(x) \\ \text{s.t.} \quad & h_i(x, u_i) \leq 0 \quad \forall i \in I, \\ & h_j(x) \leq 0 \quad \forall j \in J, \\ & u \in S(x) \end{aligned}$$

where $S(x)$ is the set of solutions to the x-parameterized lower-level problem

$$\max_{u=(u_i)_{i\in I}} \quad \sum_{i\in I} h_i(x, u_i) \quad \text{s.t.} \quad u_i \in U_i(x) \ \forall i \in I.$$

Then, x^* is a solution to the strictly robust optimization problem with decision-dependent uncertainty sets $U_i(x)$, $i \in I$.

Some Remarks

- Both “standard” as well as decision-dependent uncertainty sets can be tackled
- Uncertain constraints with concave dependence on u and non-empty interior of the uncertainty sets lead to convex lower levels satisfying Slater’s CQ

Some Remarks

- Both "standard" as well as decision-dependent uncertainty sets can be tackled
- Uncertain constraints with concave dependence on u and non-empty interior of the uncertainty sets lead to convex lower levels satisfying Slater's CQ
- The bilevel problem from the theorem is an optimistic one.
  Using a pessimistic one, the lower level can even have a constant objective function

$$
\begin{aligned}
\min_{x} \max_{u \in S(x)} \quad & H(x) \\
\text{s.t.} \quad & h_i(x, u_i) \leq 0 \quad \forall i \in I,\ \forall u = (u_i)_{i \in I} \in S(x) \\
& h_j(x) \leq 0 \quad \forall j \in J
\end{aligned}
$$

$S(x)$: set of solutions to the x-parameterized lower-level problem

$$
\min_{u=(u_i)_{i \in I}} \quad 42 \quad \text{s.t.} \quad u_i \in U_i(x) \quad \forall i \in I
$$

And the other way around?

Theorem

Let (x^*, y^*) be a solution to the strictly robust problem

$$\begin{aligned} \min_{x,y} \quad & F(x,y) \\ \text{s.t.} \quad & f(x,y) \le f(x,\tilde{y}) \quad \forall \tilde{y} \in U(x), \\ & G(x,y) \le 0, \\ & g(x,y) \le 0, \end{aligned}$$

where the decision-dependent uncertainty set is given by

$$U(x) := \left\{ \tilde{y} \in \mathbb{R}^{n_y} : g(x,\tilde{y}) \le 0 \right\}.$$

Then, (x^*, y^*) is a solution to the optimistic bilevel problem.

First Main Result

Optimistic bilevel optimization

and

strictly robust optimization with decision-dependent uncertainty sets

are equivalent!

First Main Result

Optimistic bilevel optimization

and

strictly robust optimization with decision-dependent uncertainty sets

are equivalent!

PS: Both are equivalent to generalized semi-infinite optimization.

What about pessimistic bilevel problems?

Theorem

A solution to the pessimistic bilevel problem can be computed by solving the following strictly robust problem with decision-dependent uncertainty set

$$\begin{aligned} \min_{x,y} \quad & F(x,y) \\ \text{s.t.} \quad & F(x,y) \geq F(x,y') \quad \forall y' \in U(x) \\ & G(x,y') \leq 0 \quad \forall y' \in U(x) \\ & f(x,y) \leq f(x,y') \quad \forall y' \in U(x) \\ & G(x,y) \leq 0 \\ & g(x,y) \leq 0 \end{aligned}$$

with uncertainty set

$$U(x) := \{\tilde{y} \colon f(x,\tilde{y}) \leq \chi(x),\ g(x,\tilde{y}) \leq 0\}\,.$$

Here, $\chi(x)$ is the optimal-value function of the x-parameterized lower-level problem.

- Uncertainty is still handled in a strict way
- Decisions are split
  - here-and-now
  - wait-and-see
- Ben-Tal, Goryashko, Guslitzer, Nemirovski (2004), Bertsimas, Den Hertog (2022)

$$\min_{x \in X} \max_{u \in U} \min_{y \in Y(x,u)} H(x, y)$$

with

$$Y(x, u) = \left\{ y \in \mathbb{R}^{n_y} : h(x, y, u) \leq 0 \right\}$$

Important difference:

1. Wait-and-see follower
2. Here-and-now follower

Robust bilevel problem with wait-and-see follower

$$\min_{x \in X} \max_{u \in U(x)} \min_{y} \; \{F(x, y) \colon y \in S(x, u)\}$$

$S(x, u)$: set of solutions to the (x, u)-parameterized problem

$$\min_{y} \quad f(x, y) \quad \text{s.t.} \quad g(x, y, u) \leq 0$$

Theorem

Let x^* be a solution to the optimistic robust bilevel problem with wait-and-see follower

$$\min_{x \in X} \max_{u \in U(x)} \min_{y} \{H(x, y) \colon y \in S(x, u)\}$$

where $X \subseteq \mathbb{R}^{n_x}$, $U(x) \subseteq \mathbb{R}^{n_u}$, and $S(x, u)$ is the set of solutions to the (x, u)-parameterized lower-level problem

$$\min_{y} \quad H(x, y) \quad \text{s.t.} \quad h(x, y, u) \leq 0.$$

Then, x^* is a solution to the two-stage robust problem with decision-dependent uncertainty set $U(x)$.

Theorem

Let x^* be a solution to the optimistic robust bilevel problem with wait-and-see follower

$$\min_{x \in X} \max_{u \in U(x)} \min_{y} \{H(x,y) \colon y \in S(x,u)\}$$

where $X \subseteq \mathbb{R}^{n_x}$, $U(x) \subseteq \mathbb{R}^{n_u}$, and $S(x,u)$ is the set of solutions to the (x,u)-parameterized lower-level problem

$$\min_{y} \quad H(x,y) \quad \text{s.t.} \quad h(x,y,u) \leq 0.$$

Then, x^* is a solution to the two-stage robust problem with decision-dependent uncertainty set $U(x)$.

The other direction again requires using optimal-value functions.

Min-Max-Regret Optimization

$$
\begin{aligned}
\min_{x} \quad & \max_{u \in U} \left\{ H(x,u) - \min_{\{y:\ h(y,u) \le 0\}} H(y,u) \right\} \\
\text{s.t.} \quad & h(x,u) \le 0 \quad \forall u \in U
\end{aligned}
$$

$$\begin{aligned} \min_{x} \quad & \max_{u \in U} \left\{ H(x,u) - \min_{\{y:\ h(y,u) \leq 0\}} H(y,u) \right\} \\ \text{s.t.} \quad & h(x,u) \leq 0 \quad \forall u \in U \end{aligned}$$

Theorem

Let (x^*, y^*) *be a solution to the pessimistic bilevel problem*

$$\min_{x} \left\{ \max_{(y_1,y_2) \in S(x)} H(x,y_1) - H(y_2,y_1) \colon h(x,y_1') \leq 0 \ \forall y' = (y_1', y_2') \in S(x) \right\}$$

with $y = (y_1, y_2)$ *and* $S(x) = \arg\min_y \{42 : y_1 \in U, h(y_2, y_1) \leq 0\}$.

Then, x^* *is a solution to the regret problem.*

Min-Max-Regret Optimization

Min-max regret criterion is most commonly defined with uncertainty only in the objective

- Aissi et al. (2009), Kasperski and Zieliński (2016), Kouvelis and Yu (2013)

Theorem

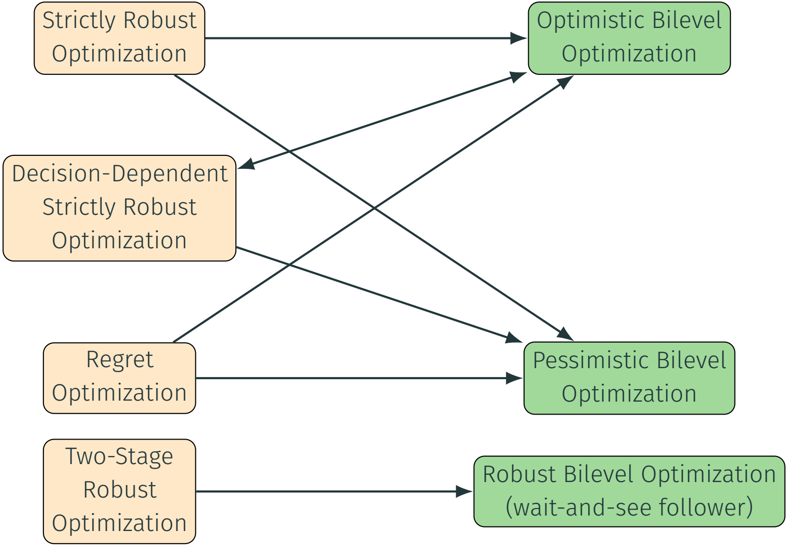
Let (x^*, y^*) *be a solution to the optimistic bilevel problem*

$$\min_{x,y \in S(x)} \{H(x, y_1) - H(y_2, y_1) \colon h(x) \leq 0\}$$

with $y = (y_1, y_2)$ *and* $S(x) = \arg\min_y \{H(y_2, y_1) - H(x, y_1) : y_1 \in U, h(y_2) \leq 0\}$.

Then, x^* *is a solution to the regret problem without uncertainty in the constraints.*

That's what we know now

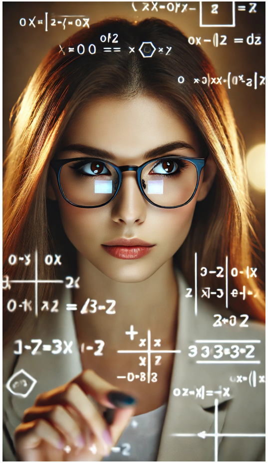


The burial of coupling constraints in linear bilevel optimization

The Team

Henri Lefebvre

Dorothee Henke



Johannes Thürauf

Do we “Really” Increase Modeling Capabilities by Using Coupling Constraints?

Spoiler: No!

Do we “Really” Increase Modeling Capabilities by Using Coupling Constraints?

Spoiler: No!

Why not?

For every given linear bilevel optimization problem with coupling constraints, we derive …

- a linear bilevel problem without coupling constraints
- that has the same set of optimal solutions

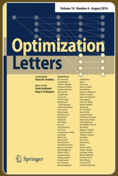
The Details

Home > Optimization Letters > Article

On coupling constraints in linear bilevel optimization

Short Communication | Open access | Published: 03 December 2024

(2024) Cite this article



Optimization Letters

Aims and scope →

Submit manuscript →

Download PDF

You have full access to this open access article

Dorothee Henke, Henri Lefebvre, Martin Schmidt & Johannes Thürauf

Sections | Figures | References

Abstract

It is well-known that coupling constraints in linear bilevel optimization can lead to disconnected feasible sets, which is not possible without coupling constraints. However, there is no difference between linear bilevel problems with and without coupling constraints w.r.t. their complexity-theoretical hardness. In this note, we prove that, although there is a clear difference between these two classes of problems in terms of their feasible sets, the classes are equivalent on the level of optimal solutions. To this end, given

Re-Writing the Problem

Upper level

$$\begin{aligned} \min_{x,y,\varepsilon} \quad & c^\top x + d^\top y \\ \text{s.t.} \quad & x \in X \\ & \varepsilon = 0 \\ & (y, \varepsilon) \in \tilde{S}(x) \end{aligned} \qquad \text{(R)}$$

Lower level

$$\begin{aligned} \min_{y,\varepsilon} \quad & f^\top y \\ \text{s.t.} \quad & Ax + By + \varepsilon e \geq a \\ & Cx + Dy \geq b \\ & \varepsilon \geq 0 \end{aligned}$$

Re-Writing the Problem

Upper level

$$\begin{aligned} \min_{x,y,\varepsilon} \quad & c^\top x + d^\top y \\ \text{s.t.} \quad & x \in X \\ & \varepsilon = 0 \\ & (y,\varepsilon) \in \tilde{S}(x) \end{aligned} \tag{R}$$

Lower level

$$\begin{aligned} \min_{y,\varepsilon} \quad & f^\top y \\ \text{s.t.} \quad & Ax + By + \varepsilon e \geq a \\ & Cx + Dy \geq b \\ & \varepsilon \geq 0 \end{aligned}$$

Lemma

For every bilevel feasible point (x, y) *of the original bilevel problem, the point* $(x, y, 0)$ *is bilevel feasible for Problem* (R) *with the same objective value. For every bilevel feasible point* (x, y, ε) *of Problem* (R)*, the point* (x, y) *is bilevel feasible for the original bilevel problem with the same objective value.*

Re-Writing the Problem & Penalize

Theorem

There is a finite and poly-sized parameter $\kappa > 0$ (in the bit-encoding length of the problem's data) so that the bilevel problem (without coupling constraints)

$$\begin{aligned} \min_{x,y,\varepsilon} \quad & c^\top x + d^\top y + \kappa\varepsilon \\ \text{s.t.} \quad & x \in X,\ (y,\varepsilon) \in \tilde{S}(x) \end{aligned} \tag{P}$$

has the same set of optimal solutions as Problem (R).

Re-Writing the Problem & Penalize

Theorem

There is a finite and poly-sized parameter $\kappa > 0$ (in the bit-encoding length of the problem's data) so that the bilevel problem (without coupling constraints)

$$\begin{aligned} \min_{x,y,\varepsilon} \quad & c^\top x + d^\top y + \kappa\varepsilon \\ \text{s.t.} \quad & x \in X, \ (y, \varepsilon) \in \tilde{S}(x) \end{aligned} \tag{P}$$

has the same set of optimal solutions as Problem (R).

That's surprising!

- Reason #1
  - Feasible region of the original problem is nonconvex and disconnected
  - Ye and Zhu (1995): no constraint qualification is satisfied
  - Exact penalization usually fails!
- Reason #2
  - Exact penalty functions are usually nonsmooth (à la ℓ_1)
  - Our penalty function is perfectly smooth (even linear)

Proof Idea

1. We derive a single-level reformulation of the bilevel problem (R), using the KKT conditions of the follower's problem.
2. We apply results from augmented Lagrangian duality theory for mixed-integer linear problems to show that a poly-sized exact penalization parameter exists.
3. We show that the resulting mixed-integer linear program is nothing but the KKT reformulation of Problem (P).

Proof

- Lower-level problem of Problem (R) is an LP
- Dempe and Dutta (2012): Replace it with its KKT conditions

$$
\begin{aligned}
\min_{x,y,\varepsilon} \quad & c^\top x + d^\top y \\
\text{s.t.} \quad & x \in X, \; \varepsilon = 0 \\
& Ax + By + \varepsilon e \geq a, \; Cx + Dy \geq b, \; \varepsilon \geq 0 \\
& B^\top \lambda + D^\top \mu = f, \; e^\top \lambda + \eta = 0 \\
& \lambda, \mu, \eta \geq 0, \\
& \lambda^\top (Ax + By + \varepsilon e - a) = 0, \; \mu^\top (Cx + Dy - b) = 0, \; \eta\varepsilon = 0
\end{aligned}
$$

- Additional binary variables z^λ, z^μ, z^η
- Sufficiently large big-M

$$
\begin{aligned}
&\lambda \leq (1 - z^\lambda)M, \quad \mu \leq (1 - z^\mu)M, \quad \eta \leq (1 - z^\eta)M \\
&Ax + By + \varepsilon e - a \leq z^\lambda M, \quad Cx + Dy - b \leq z^\mu M, \quad \varepsilon \leq z^\eta M
\end{aligned}
$$

Wait! Are we Cheating?

- Pineda and Morales (2019): Heuristics for computing big-M values usually fail
- Kleinert et al. (2020): Validating the correctness of a given big-M is as hard as the original bilevel problem

Wait! Are we Cheating?

- Pineda and Morales (2019): Heuristics for computing big-M values usually fail
- Kleinert et al. (2020): Validating the correctness of a given big-M is as hard as the original bilevel problem

But ...

- Buchheim (2023): valid and poly-sized M can be computed in polynomial time

Proof ... Continued

We have the MILP

$$
\begin{aligned}
\min_{x,y,\varepsilon,z^\lambda,z^\mu,z^\eta} \quad & c^\top x + d^\top y \\
\text{s.t.} \quad & x \in X,\ \varepsilon = 0 \\
& Ax + By + \varepsilon e \geq a,\ Cx + Dy \geq b,\ \varepsilon \geq 0 \\
& B^\top \lambda + D^\top \mu = f,\ e^\top \lambda + \eta = 0,\ \lambda, \mu, \eta \geq 0 \\
& \lambda \leq (1 - z^\lambda)M,\ \mu \leq (1 - z^\mu)M,\ \eta \leq (1 - z^\eta)M \\
& Ax + By + \varepsilon e - a \leq z^\lambda M,\ Cx + Dy - b \leq z^\mu M,\ \varepsilon \leq z^\eta M
\end{aligned}
$$

Proof ... Continued

We have the MILP

$$
\begin{aligned}
\min_{x,y,\varepsilon,z^\lambda,z^\mu,z^\eta} \quad & c^\top x + d^\top y \\
\text{s.t.} \quad & x \in X, \ \varepsilon = 0 \\
& Ax + By + \varepsilon e \geq a, \ Cx + Dy \geq b, \ \varepsilon \geq 0 \\
& B^\top \lambda + D^\top \mu = f, \ e^\top \lambda + \eta = 0, \ \lambda, \mu, \eta \geq 0 \\
& \lambda \leq (1 - z^\lambda)M, \ \mu \leq (1 - z^\mu)M, \ \eta \leq (1 - z^\eta)M \\
& Ax + By + \varepsilon e - a \leq z^\lambda M, \ Cx + Dy - b \leq z^\mu M, \ \varepsilon \leq z^\eta M
\end{aligned}
$$

ℓ_∞ penalization of the coupling constraint $\varepsilon = 0$

$$
\begin{aligned}
\min_{x,y,\varepsilon} \quad & c^\top x + d^\top y + \kappa\varepsilon \\
\text{s.t.} \quad & \text{all constraints except from } \varepsilon = 0
\end{aligned}
$$

Proof … Continued: What About κ?

Home > Mathematical Programming > Article

Exact augmented Lagrangian duality for mixed integer linear programming

Full Length Paper | Series A | Published: 21 April 2016

Volume 161, pages 365–387, (2017) Cite this article

Mohammad Javad Feizollahi, Shabbir Ahmed & Andy Sun

Abstract

We investigate the augmented Lagrangian dual (ALD) for mixed integer linear programming (MIP) problems. ALD modifies the classical Lagrangian dual by appending a nonlinear penalty function on the violation of the dualized constraints in order to reduce the duality gap. We first provide a primal characterization for ALD for MIPs and prove that ALD is able to asymptotically achieve zero duality gap when the weight on the penalty function is allowed to go to infinity. This provides an alternative characterization and proof of a recent result in Boland and Eberhard (Math Program 150(2):491–509, 2015, Proposition 3). We further show that, under some mild conditions, ALD using any norm as

Feizollahi et al. (2016)

- Theorem 4: duality gap for the augmented Lagrangian dual of a solvable (mixed-integer) linear optimization problem can be closed by using a norm as the augmenting function and a sufficiently large but finite penalty parameter.
- Proposition 1: Optimal solutions of MILP reformulation and the ℓ_∞ penalty problem are the same

Gu et al. (2020)

- Theorem 22: Penalty parameter can be chosen to be of polynomial size in case of the ℓ_∞-norm

Proof … Continued: Back to the Slide Before

We have the MILP

$$\begin{aligned}
\min_{x,y,\varepsilon,z^\lambda,z^\mu,z^\eta} \quad & c^\top x + d^\top y \\
\text{s.t.} \quad & x \in X, \ \varepsilon = 0, \\
& Ax + By + \varepsilon e \geq a, \ Cx + Dy \geq b, \ \varepsilon \geq 0 \\
& B^\top \lambda + D^\top \mu = f, \ e^\top \lambda + \eta = 0, \ \lambda, \mu, \eta \geq 0 \\
& \lambda \leq (1 - z^\lambda)M, \ \mu \leq (1 - z^\mu)M, \ \eta \leq (1 - z^\eta)M \\
& Ax + By + \varepsilon e - a \leq z^\lambda M, \ Cx + Dy - b \leq z^\mu M, \ \varepsilon \leq z^\eta M
\end{aligned}$$

ℓ_∞ penalization of the coupling constraint $\varepsilon = 0$

$$\begin{aligned}
\min_{x,y,\varepsilon} \quad & c^\top x + d^\top y + \kappa\varepsilon \\
\text{s.t.} \quad & \text{all constraints except from } \varepsilon = 0
\end{aligned}$$

This is the KKT reformulation of the bilevel problem from the theorem!

Feizollahi et al. (2016) & Gu et al. (2020)

Existence of finite and poly-sized
exact penalty parameter.

Open (until last year)

Can it be computed in polynomial time?

κ ... One More Time!

Feizollahi et al. (2016) & Gu et al. (2020)

Existence of finite and poly-sized exact penalty parameter.

Open (until last year)

Can it be computed in polynomial time?

Yes! Lemma 4 of Lefebvre and Schmidt (2024)

Exact Augmented Lagrangian Duality for Nonconvex Mixed-Integer Nonlinear Optimization

Henri Lefebvre, Martin Schmidt

Abstract. In the context of mixed-integer nonlinear problems (MINLPs), it is well-known that strong duality does not hold in general if the standard Lagrangian dual is used. Hence, we consider the augmented Lagrangian dual (ALD), which adds a nonlinear penalty function to the classic Lagrangian function. For this setup, we study conditions under which the ALD leads to a zero duality gap for general MINLPs. In particular, under mild assumptions and for a large class of penalty functions, we show that the ALD yields zero duality gaps if the penalty parameter goes to infinity. If the penalty function is a norm, we also show that the ALD leads to zero duality gaps for a finite penalty parameter. Moreover, we show that such a finite penalty parameter can be computed in polynomial time in the mixed-integer linear case. This generalizes the recent results on linearly constrained mixed-integer problems by Bhardwaj et al. (2024), Boland and Eberhard (2014), Feizollahi et al. (2016), and Gu et al. (2020).

And what about the pessimistic case?

Optimistic bilevel problem without coupling constraints

$$\min_{x \in X} \quad F_o(x) := c^\top x + \min_y \left\{ d^\top y : y \in S(x) \right\}$$

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

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

Optimistic Bilevel Problems with and without Coupling Constraints

Optimistic bilevel problem without coupling constraints

$$\min_{x \in X} \quad F_{\text{o}}(x) := c^\top x + \min_y \left\{ d^\top y \colon y \in S(x) \right\}$$

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

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

Optimistic bilevel problem with coupling constraints

$$\min_{x \in X} \quad F_{\text{oc}}(x) := c^\top x + \min_y \left\{ d^\top y \colon y \in S(x),\ Ax + By \geq a \right\}$$

Pessimistic Bilevel Problems with and without Coupling Constraints

Pessimistic bilevel problem without coupling constraints

$$\min_{x \in \bar{X}} \quad F_{\mathrm{p}}(x) := c^\top x + \max_y \left\{ d^\top y : y \in S(x) \right\}$$

with

$$\bar{X} := X \cap \{x : S(x) \neq \emptyset\}$$

Pessimistic Bilevel Problems with and without Coupling Constraints

Pessimistic bilevel problem without coupling constraints

$$\min_{x \in \bar{X}} \quad F_{\mathrm{p}}(x) := c^\top x + \max_y \left\{ d^\top y : y \in S(x) \right\}$$

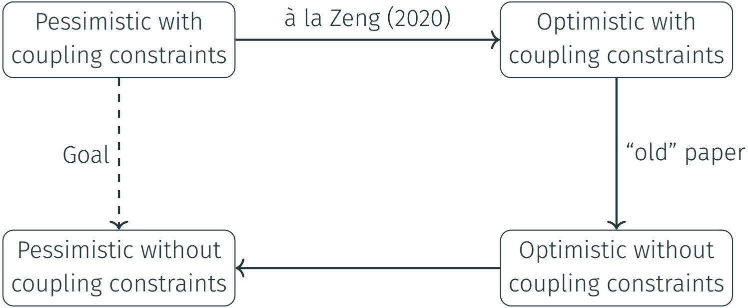
with

$$\bar{X} := X \cap \{x : S(x) \neq \emptyset\}$$

Pessimistic bilevel problem with coupling constraints

$$\begin{aligned} \min_{x \in \bar{X}} \quad & F_{\mathrm{pc}}(x) := c^\top x \\ \text{s.t.} \quad & Ax + By \geq a \quad \text{for all } y \in S(x) \end{aligned}$$

Bye bye, coupling constraints ...



The pessimistic coupling constraint

$$Ax + By \geq a \quad \text{for all } y \in S(x)$$

is equivalent to

$$A_{i\cdot}x + B_{i\cdot}y \geq a_i \quad \text{for all } y \in S(x) \text{ and all } i \in [m] := \{1, \ldots, m\}\,.$$

The pessimistic coupling constraint

$$Ax + By \geq a \quad \text{for all } y \in S(x)$$

is equivalent to

$$A_{i\cdot}x + B_{i\cdot}y \geq a_i \quad \text{for all } y \in S(x) \text{ and all } i \in [m] := \{1, \ldots, m\}.$$

Lemma (à la Zeng (2020))

Let $x \in X$ be given and consider a fixed $i \in [m]$. Then, x satisfies the i-th coupling constraint if and only if there exist $\bar{y}$ and

$$y^i \in \arg\min \left\{ B_{i\cdot}y \colon Dy \geq b - Cx,\ f^\top y \leq f^\top \bar{y} \right\}$$

that satisfy

$$D\bar{y} \geq b - Cx, \quad B_{i\cdot}y^i \geq a_i - A_{i\cdot}x.$$

Theorem

Let $\mathcal{S}$ be the set of globally optimal solutions to the pessimistic bilevel problem with coupling constraints. Moreover, let $\tilde{\mathcal{S}}$ be the set of globally optimal solutions to the optimistic single-leader multi-follower problem

$$\min_{(x,\bar{y})\in\tilde{X}} c^\top x + \min_y \left\{0 \colon y^i \in \tilde{S}^i(x,\bar{y}),\ B_{i\cdot}y^i \geq a_i - A_{i\cdot}x \text{ for all } i \in [m]\right\}$$

with $\tilde{X} = \{(x,\bar{y}) \colon x \in X, D\bar{y} \geq b - Cx\}$, $\tilde{S}^i(x,\bar{y}) = \arg\min_{y'}\{B_{i\cdot}y' \colon Dy' \geq b - Cx, f^\top y' \leq f^\top \bar{y}\}$, and $y = (y^i)_{i=1}^m$. Let $\hat{\mathcal{S}}$ be the set of globally optimal solutions to the optimistic bilevel problem

$$\min_{(x,\bar{y})\in\tilde{X}} c^\top x + \min_y \left\{0 \colon y \in \hat{S}(x,\bar{y}),\ B_{i\cdot}y^i \geq a_i - A_{i\cdot}x \text{ for all } i \in [m]\right\},$$

where $\hat{S}(x,\bar{y})$ denotes the set of optimal solutions to the aggregated lower-level problem

$$\min_y \quad \sum_{i=1}^m B_{i\cdot}y^i \quad \text{s.t.} \quad Dy^i \geq b - Cx,\ f^\top y^i \leq f^\top \bar{y} \quad \text{for all } i \in [m].$$

Then, $\mathcal{S} = \operatorname{proj}_x(\tilde{\mathcal{S}}) = \operatorname{proj}_x(\hat{\mathcal{S}})$ holds and all optimal objective function values coincide.

Theorem (Simply apply the "old" optimistic result ...)

There is a poly-sized penalty parameter $\kappa > 0$ so that the optimistic bilevel problem with coupling constraints has the same set of globally optimal solutions as the optimistic bilevel problem

$$\min_{x \in X} \quad c^\top x + \min_{y,\varepsilon} \left\{ d^\top y + \kappa\varepsilon \colon (y, \varepsilon) \in S'(x) \right\}$$

without coupling constraints, where $S'(x)$ is the set of optimal solutions to the x-parameterized lower-level problem

$$\begin{aligned} \min_{y,\varepsilon} \quad & f^\top y \\ \text{s.t.} \quad & Ax + By + \varepsilon e \geq a, \\ & Cx + Dy \geq b, \\ & \varepsilon \geq 0, \end{aligned}$$

where e is the vector of all ones in appropriate dimension. Moreover, both bilevel problems have the same optimal objective function value.

Let's consider

$$\min_{(x,\bar{y})\in\tilde{X}} \quad F_{\text{oa}}(x,\bar{y}) := c^\top x + d^\top \bar{y} + \min_{y,\varepsilon} \left\{0 \colon \varepsilon = 0,\ (y,\varepsilon) \in \tilde{S}(x,\bar{y})\right\} \tag{AUX-UL}$$

with a single coupling constraint. Again, we use $\tilde{X} = \{(x,\bar{y}) \colon x \in X, D\bar{y} \geq b - Cx\}$ and $\tilde{S}(x,\bar{y})$ denotes the set of optimal points to

$$\begin{aligned} \min_{y,\varepsilon} \quad & f^\top y \\ \text{s.t.} \quad & Cx + Dy \geq b, \quad f^\top \bar{y} - f^\top y = \varepsilon, \quad \varepsilon \geq 0. \end{aligned} \tag{AUX-LL}$$

Let's consider

$$\min_{(x,\bar{y})\in\tilde{X}} \quad F_{\text{oa}}(x,\bar{y}) := c^\top x + d^\top \bar{y} + \min_{y,\varepsilon} \left\{ 0 \colon \varepsilon = 0,\ (y,\varepsilon) \in \tilde{S}(x,\bar{y}) \right\} \tag{AUX-UL}$$

with a single coupling constraint. Again, we use $\tilde{X} = \{(x,\bar{y}) \colon x \in X, D\bar{y} \geq b - Cx\}$ and $\tilde{S}(x,\bar{y})$ denotes the set of optimal points to

$$\begin{aligned} \min_{y,\varepsilon} \quad & f^\top y \\ \text{s.t.} \quad & Cx + Dy \geq b, \quad f^\top \bar{y} - f^\top y = \varepsilon, \quad \varepsilon \geq 0. \end{aligned} \tag{AUX-LL}$$

Lemma

For every bilevel feasible point x of the optimistic bilevel problem without coupling constraints, the point $(x,\bar{y})$ with $\bar{y} \in \arg\min_y \{d^\top y \colon y \in S(x)\}$ is also bilevel feasible for the optimistic bilevel problem (AUX-UL) *with the same objective value. Moreover, for every globally optimal point $(x,\bar{y})$ to Problem* (AUX-UL)*, x is bilevel feasible for the optimistic bilevel problem without coupling constraints with the same objective value.*

Lemma

There is a poly-sized parameter $\kappa > 0$ so that Problem (AUX-UL) *has the same set of globally optimal solutions as the optimistic bilevel problem*

$$\min_{(x,\bar{y}) \in \tilde{X}} \quad F_{\mathrm{o}\kappa}(x,\bar{y}) := c^\top x + d^\top \bar{y} + \min_{y,\varepsilon} \left\{ \kappa\varepsilon \colon (y,\varepsilon) \in \tilde{S}(x,\bar{y}) \right\}$$

without coupling constraints. Here, we again use $\tilde{X} = \{(x,\bar{y}) \colon x \in X, D\bar{y} \geq b - Cx\}$ and $\tilde{S}(x,\bar{y})$ is the set of optimal solutions of (AUX-LL).

From Optimistic to Pessimistic Bilevel Optimization without Coupling Constraints

Lemma

There is a poly-sized parameter $\kappa > 0$ *so that Problem* (AUX-UL) *has the same set of globally optimal solutions as the optimistic bilevel problem*

$$\min_{(x,\bar{y}) \in \tilde{X}} \quad F_{\mathrm{o}\kappa}(x, \bar{y}) := c^\top x + d^\top \bar{y} + \min_{y,\varepsilon} \left\{ \kappa\varepsilon \colon (y, \varepsilon) \in \tilde{S}(x, \bar{y}) \right\}$$

without coupling constraints. Here, we again use $\tilde{X} = \{(x, \bar{y}) \colon x \in X, D\bar{y} \geq b - Cx\}$ *and* $\tilde{S}(x, \bar{y})$ *is the set of optimal solutions of* (AUX-LL).

Theorem

For any κ*, the optimistic bilevel problem without coupling constraints from the last lemma and its pessimistic version*

$$\min_{(x,\bar{y}) \in \tilde{X}} \quad F_{\mathrm{p}\kappa}(x, \bar{y}) := c^\top x + d^\top \bar{y} + \max_{y,\varepsilon} \left\{ \kappa\varepsilon \colon (y, \varepsilon) \in \tilde{S}(x, \bar{y}) \right\}$$

have the same set of feasible and globally optimal solutions.

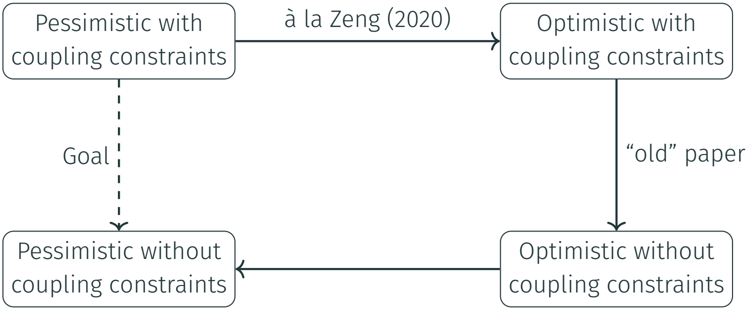
Corollary

There is a poly-sized parameter $\kappa > 0$ *so that the optimistic bilevel problem without coupling constraints has the same set of globally optimal solutions as the pessimistic bilevel problem*

$$\min_{(x,\bar{y})\in\tilde{X}} \quad F_{\mathrm{p}\kappa}(x,\bar{y}) := c^\top x + d^\top \bar{y} + \max_{y,\varepsilon}\left\{\kappa\varepsilon \colon (y,\varepsilon) \in \tilde{S}(x,\bar{y})\right\}$$

without coupling constraints.

Bye bye, coupling constraints ...



Proof of the “à la Zeng (2020)” lemma

Let $x \in X$ be given and consider a fixed $i \in [m]$. Then, the i-th coupling constraint is equivalent to $\min_y\{B_{i\cdot}y : y \in S(x)\} \geq a_i - A_{i\cdot}x$, which can be reformulated as $B_{i\cdot}y^i \geq a_i - A_{i\cdot}x$ with $y^i \in \arg\min_y \{B_{i\cdot}y : y \in S(x)\}$. Now, let φ denote the optimal-value function of the lower-level problem. It follows that x satisfies the i-th coupling constraint if and only if

$$B_{i\cdot}y^i \geq a_i - A_{i\cdot}x \quad \text{with} \quad y^i \in \arg\min_y \left\{B_{i\cdot}y : Dy \geq b - Cx,\ f^\top y \leq \varphi(x)\right\}. \tag{1}$$

We now show that the latter is equivalent to the stated conditions in the lemma. First, let us assume that (1) holds. Then, there exists $\bar{y}$ such that $\varphi(x) = f^\top \bar{y}$ and $D\bar{y} \geq b - Cx$ is satisfied. Hence, the conditions of the lemma hold.

Conversely, assume that the conditions of the lemma are satisfied. The feasibility of $\bar{y}$ implies

$$\min_y \left\{B_{i\cdot}y : Dy \geq b - Cx,\ f^\top y \leq f^\top \bar{y}\right\} \leq \min_y \left\{B_{i\cdot}y : Dy \geq b - Cx,\ f^\top y \leq \varphi(x)\right\}.$$

Hence, (1) is satisfied, which concludes the proof.

The End

There is a lot more to discover and to study!

- Bilevel optimization with discrete variables
- Bilevel optimization with nonlinear lower-level problems
- Stochastic bilevel optimization
- Robust bilevel optimization
- Bounded rationality
- etc. etc. etc.

The End

There is a lot more to discover and to study!

- Bilevel optimization with discrete variables
- Bilevel optimization with nonlinear lower-level problems
- Stochastic bilevel optimization
- Robust bilevel optimization
- Bounded rationality
- etc. etc. etc.

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

BOBILib: Bilevel Optimization (Benchmark) Instance Library

- More than 2600 instances of mixed-integer linear bilevel optimization problems
- Well-curated set of test instances
- Freely available for the research community
- Testing of new methods + comparison with other ones
- Different types of instances
  - Interdiction
  - Mixed-integer
  - Pure integer
- Benchmark sets for all of them
- Extensive numerical results
- New data + solution format
- All best known solutions available

`https://bobilib.org`