A Gentle and Incomplete Introduction to Linear Bilevel Optimization
... and a Tiny New Result

Martin Schmidt

The 5th EUROYoung Workshop
Naples — October 15, 2025

Agenda

1. What is Bilevel Optimization Anyway?
2. Some Theory on Linear Bilevel Problems
3. How to Solve a Linear Bilevel Problem?
4. Coupling Constraints … or not

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 is 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 is 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

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: A Real-World Application

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 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 ...

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

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.

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.

Anti Drug Smuggling Revisited

Follower: w-parameterized maximum flow problem

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

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

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

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

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

Special properties: min-max problem & no coupling constraints

An Academic and Linear Example (Kleinert 2021)

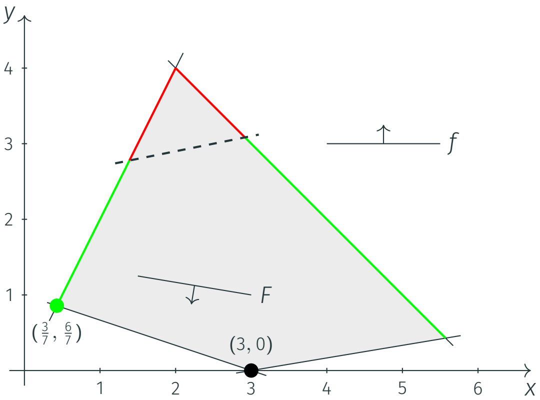
Upper-level problem

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

Lower-level problem

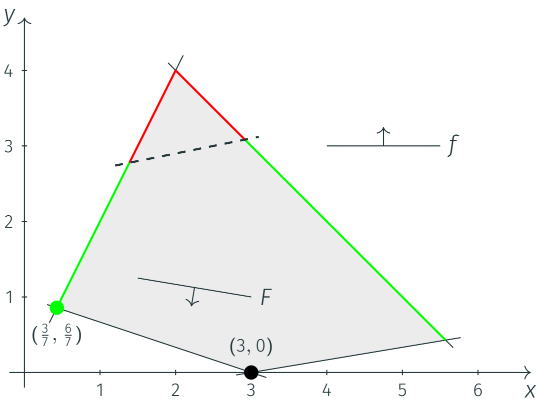
$$
\begin{aligned}
\min_{y} \quad & f(x,y) = -y \\
\text{s.t.} \quad & 2x - y \geq 0 \\
& -x - y \geq -6 \\
& -x + 6y \geq -3 \\
& x + 3y \geq 3
\end{aligned}
$$

An Academic and Linear Example (Kleinert 2021)



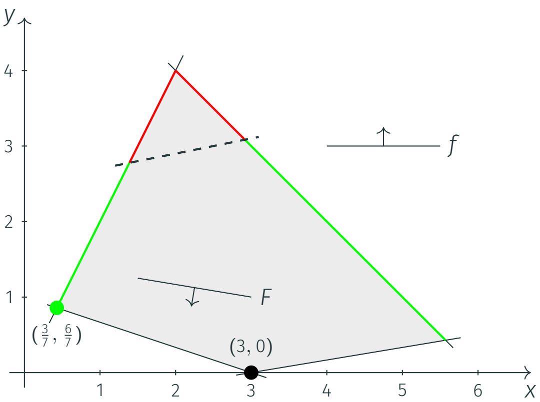
- Shared constrained set: gray area
- Green and red lines: nonconvex set of optimal follower solutions (lifted to the x-y-space)
- Green lines: Nonconvex and disconnected bilevel feasible set of the bilevel problem

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)

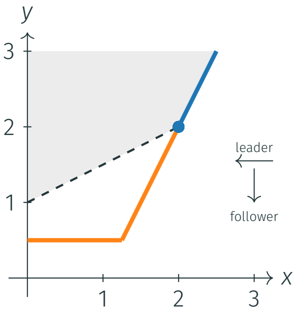


5. The single leader constraint (dashed line) renders certain optimal responses of the follower infeasible.
6. The bilevel feasible region $\mathcal{F}$ corresponds to the green facets.
7. Thus, the feasible set is not only nonconvex but also disconnected.
8. The optimal solution is $(3/7, 6/7)$ with 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: $(2, 2)$

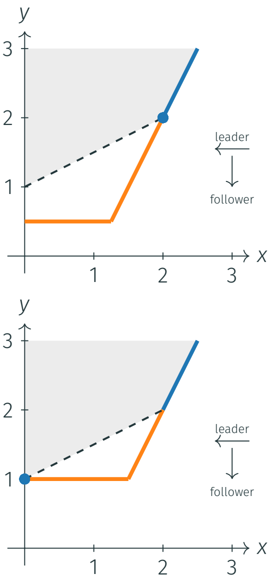


Independence of Irrelevant Constraints (Kleinert et al. 2021; Macal and Hurter 1997)

- 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 on 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 $b \in \mathbb{R}^\ell$.

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 $b \in \mathbb{R}^\ell$.

Remark

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

A First Structural Result

- Our goal now is to understand the geometric properties of LP-LP bilevel problems.

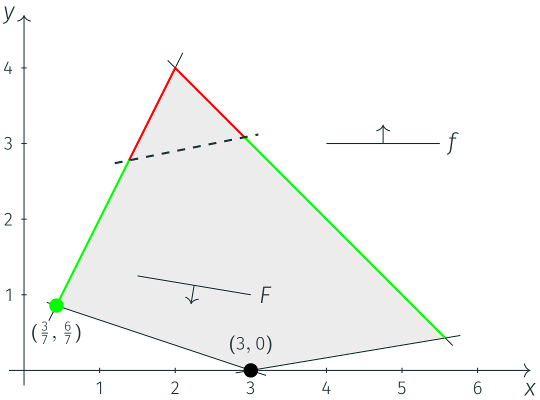
A First Structural Result

- Our goal now is to understand the geometric properties of LP-LP bilevel problems.

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



A First Structural Result: Proof

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

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

A 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.

A 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

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

A 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,

A 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.

A 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

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

A 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

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

A 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.

A 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.

A 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.

A 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

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

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 $(Cx + Dy - b)_i = 0$.

A 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.

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.

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.

How to 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

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

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

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

- 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

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

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.

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

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

Abstract

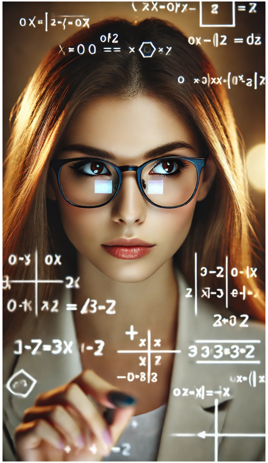
One of the most frequently used approaches to solve linear bilevel optimization problems consists in replacing the lower-level problem with its Karush–Kuhn–Tucker (KKT) conditions and by reformulating the KKT complementarity conditions using techniques from mixed-integer linear optimization. The latter 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.

Coupling Constraints ... or not

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 …

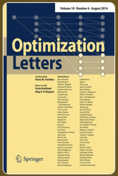
- a linear bilevel problem without coupling constraints
- that has the same set of optimal solutions

The Details

On coupling constraints in linear bilevel optimization

Short Communication | Open access | Published: 03 December 2024

(2024) Cite this article



Dorothee Henke, Henri Lefebvre, Martin Schmidt & Johannes Thürauf

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

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

Lower level

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

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.

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).

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`