A Gentle and Incomplete Introduction to Bilevel Optimization

Martin Schmidt

@schmaidt

June 2021

JPOC Spring School 2021 on MINLPs and Bilevel Problems “in Paris”

What you will learn today

You will learn . . .

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

Have fun!

There should be no crying in this compact course!

I will teach principles, not formulas!

I will teach principles, not formulas!

You will not remember the last ε,
but I hope you remember the core ideas!

A little advertising

Martin Schmidt, Yasmine Beck:

A Gentle and Incomplete Introduction to Bilevel Optimization

http://www.optimization-online.org/DB_FILE/2021/06/8450.pdf

SCAN ME

Overview

1. Introduction: Overview

“Usual” single-level problems

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

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

"Usual" single-level problems

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

- 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

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

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

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

- Very often, that's appropriate:
  - a single dispatcher controls a gas transport network
  - a single investment banker decides on the assets in a portfolio
  - a single logistics company decides on its supply chain

Often, life's different

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

Formalizing this situation leads to hierarchical or bilevel optimization problems.

Informal example #1: pricing

- One of the richest 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 #1: pricing

- One of the richest 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

Informal example #2: toll setting

- Imagine a transportation network
  - Example: the German highway network
- Some drivers want to reach their destination, starting from their origin
- The objective of these drivers is to travel from their origin to their destination at minimum costs
- Costs can be travel time, toll costs, or a combination of both

Informal example #2: toll setting

- Imagine a transportation network
  - Example: the German highway network
- Some drivers want to reach their destination, starting from their origin
- The objective of these drivers is to travel from their origin to their destination at minimum costs
- Costs can be travel time, toll costs, or a combination of both

- Toll setting agency decides on the tolls imposed on certain parts of the highway system
- Goal of the leader: maximize the revenues based on the tolls
- Goal of the followers: minimize their traveling costs

As before . . .

- the leader anticipates the optimal reaction of the followers
- the followers' decisions obviously depend on the decision of the leader

Informal example #2: toll setting

As before . . .

- the leader anticipates the optimal reaction of the followers
- the followers' decisions obviously depend on the decision of the leader

. . . but: one vs. multiple followers

- Pricing example: one follower
  - single-leader single-follower game
- Toll setting example: multiple followers
  - single-leader multi-follower game

Informal example #3: energy markets

- Energy sector: another very rich class of applications for bilevel optimization
- Especially the sub-field of energy market modeling
- In many countries of the world, electricity is traded via auctions at an energy exchange
- Auction rules determine the way of trading
- Usually decided on by the state government or some regulatory authority (leader)
- Aim: obtain market outcomes that are optimal in terms of social welfare
- Depending on these rules, producers, and consumers (follower) trade electricity at the exchange

Informal example #3: energy markets

- Energy sector: another very rich class of applications for bilevel optimization
- Especially the sub-field of energy market modeling
- In many countries of the world, electricity is traded via auctions at an energy exchange
- Auction rules determine the way of trading
- Usually decided on by the state government or some regulatory authority (leader)
- Aim: obtain market outcomes that are optimal in terms of social welfare
- Depending on these rules, producers, and consumers (follower) trade electricity at the exchange

As before . . .

- decision of the leader depends on the anticipation of the followers' decisions
- the firms' decisions depend on the market regime, i.e., on the decision of the leader.

The European Energy Exchange (EEX)

https://www.eex.com/en

Leipzig, Germany

Informal example #4: interdiction problems

- Important example in discrete bilevel optimization
- Leader is an interdictor that interdicts certain resources of the follower so that they cannot be used anymore by the follower

Informal example #4: interdiction problems

- Important example in discrete bilevel optimization
- Leader is an interdictor that interdicts certain resources of the follower so that they cannot be used anymore by the follower
- Often defined on graphs
- Example: shortest path
  - Follower wants to find a shortest path in a graph from an origin to a destination
  - Leader (interdictor) can interdict some of the arcs in the graph so that they cannot be part of a feasible path of the follower
  - Number of interdicted arcs is constrained by an interdiction budget of the leader

Informal example #4: interdiction problems

- Important example in discrete bilevel optimization
- Leader is an interdictor that interdicts certain resources of the follower so that they cannot be used anymore by the follower
- Often defined on graphs
- Example: shortest path
  - Follower wants to find a shortest path in a graph from an origin to a destination
  - Leader (interdictor) can interdict some of the arcs in the graph so that they cannot be part of a feasible path of the follower
  - Number of interdicted arcs is constrained by an interdiction budget of the leader
- Applications: vulnerability analysis of networks
  - Supply chains, transportation, ...

A bit more formal, please

Definition (Bilevel optimization problem)

A bilevel optimization problem is given by

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

Definition (Bilevel optimization problem)

A bilevel optimization problem is given by

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

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

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

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 . . .

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

Wording

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

A bit more formal, please . . . continued

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

. . . and . . .

$$S(x) = \arg\min_{y \in Y} \quad f(x, y)$$
$$\text{s.t.} \quad g(x, y) \geq 0$$

Functions and dimensions

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

A bit more formal, please . . . continued

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

. . . and . . .

$$S(x) = \arg\min_{y \in Y} \quad f(x, y)$$
$$\text{s.t.} \quad g(x, y) \geq 0$$

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

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.

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

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

Bilevel feasible set, inducible region

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.

Definition

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

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

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

High-point relaxation

Definition

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

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

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

Remark

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

Pricing revisited

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

Pricing revisited

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

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

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 from plan y_1 through the price vector x that is additionally imposed onto y_1

Pricing revisited

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

- 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 from 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

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

- 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 from 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

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

- 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

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

- The vectors d_1 and d_2 represent linear disutilities faced by the lower-level player when executing the activity plans y_1 as well as y_2
- d_2 may also encompass the price for executing the activities not influenced by the leader
  - These activities may, e.g., be substitutes offered by competitors for which prices are known and fixed
- The lower-level player determines his activity plans y_1 and y_2 to minimize the sum of total disutility and the price paid for plan y_1 subject to linear constraints

Pricing revisited

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

- The vectors d_1 and d_2 represent linear disutilities faced by the lower-level player when executing the activity plans y_1 as well as y_2
- d_2 may also encompass the price for executing the activities not influenced by the leader
  - These activities may, e.g., be substitutes offered by competitors for which prices are known and fixed
- The lower-level player determines his activity plans y_1 and y_2 to minimize the sum of total disutility and the price paid for plan y_1 subject to linear constraints
- To avoid the situation in which the leader would maximize her profit by setting prices to infinity for these activities y_1 that are essential, one may assume that the set $\{y_2 \colon D_2 y_2 \ge b\}$ is non-empty

Interdiction problems revisited: knapsack interdiction

- Problem formulation taken from Caprara et al. (2016)
- Follower owns a knapsack
- She fills the knapsack with items from a set of items $[n] := \{1, \ldots, n\}$.
- p_i: corresponding profit
- w_i: item's weights for the follower
- Leader's aim: minimize the follower's maximum profit by prohibiting the usage of certain items by the follower (at costs v_i)
- To this end, the leader first selects a subset of items respecting her so-called interdiction budget B
- Then, the follower can choose from the remaining items maximizing her profit considering the knapsack capacity C

Interdiction problems revisited: knapsack interdiction

$$\begin{aligned} \min_x \quad & p^\top y \\ \text{s.t.} \quad & v^\top x \leq B \\ & x \in \{0,1\}^n \\ & y \in \arg\max_{y'} \left\{ p^\top y' : y' \in Y(x) \right\} \end{aligned}$$

with

- $B, C \in \mathbb{R}$
- $p, v, w \in \mathbb{R}^n$
- feasible decisions of the follower (parameterized by the leader's decision x)

$$Y(x) = \{y \in \{0,1\}^n : w^\top y \leq C,\ y_i \leq 1 - x_i,\ i \in [n]\}$$

Interdiction problems revisited: knapsack interdiction

$$\begin{aligned}
\min_{x} \quad & p^\top y \\
\text{s.t.} \quad & v^\top x \leq B \\
& x \in \{0,1\}^n \\
& y \in \arg\max_{y'} \left\{ p^\top y' : y' \in Y(x) \right\}
\end{aligned}$$

with

- $B, C \in \mathbb{R}$
- $p, v, w \in \mathbb{R}^n$
- feasible decisions of the follower (parameterized by the leader's decision x)

$$Y(x) = \{ y \in \{0,1\}^n : w^\top y \leq C,\ y_i \leq 1 - x_i,\ i \in [n] \}$$

Note: same objective functions but different directions

An Academic and Linear Example (Kleinert 2021)

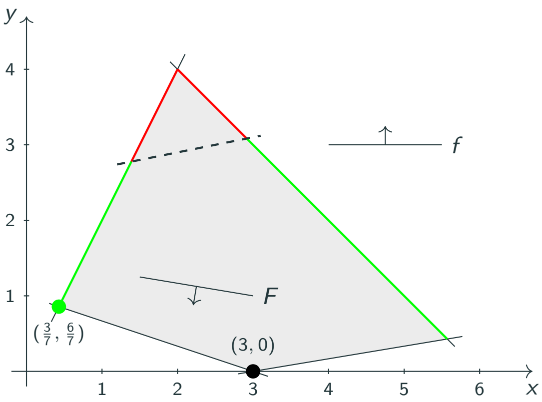
Upper-level problem

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

Lower-level problem

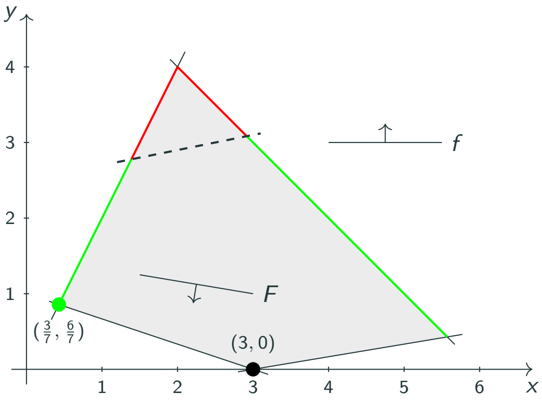
$$
\begin{aligned}
\min_{y} \quad & f(x, y) = -y \\
\text{s.t.} \quad & 2x - y \geq 0 \\
& -x - y \geq -6 \\
& -x + 6y \geq -3 \\
& x + 3y \geq 3
\end{aligned}
$$

An Academic and Linear Example (Kleinert 2021)



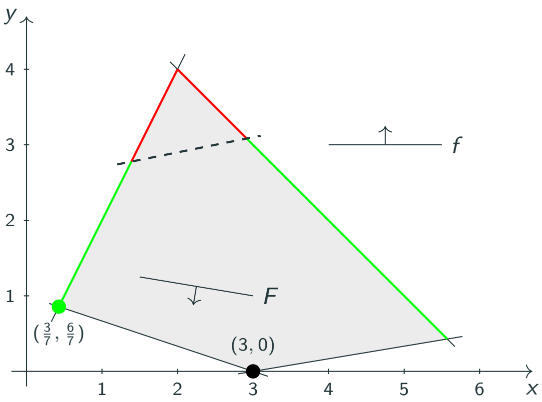
- Shared constrained set: gray area
- Green and red lines: nonconvex set of optimal follower solutions (lifted to the x-y-space)
- Green lines: Nonconvex and disconnected bilevel feasible set of the bilevel problem

An Academic Example; see Kleinert (2021)



1. The feasible region of the follower problem corresponds to the gray area.
2. The follower 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 Example; see Kleinert (2021)



5. The single leader constraint (dashed line) renders certain optimal responses of the follower infeasible.
6. The bilevel feasible region $\mathcal{F}$ corresponds to the green facets.
7. Thus, the feasible set is not only nonconvex but also disconnected.
8. The optimal solution is $(3/7, 6/7)$ with objective function value $39/7$.
9. In contrast, ignoring the follower's objective, i.e., solving the high-point relaxation, yields the optimal solution $(3, 0)$ with objective function value 3. Note that the latter point is not bilevel feasible.

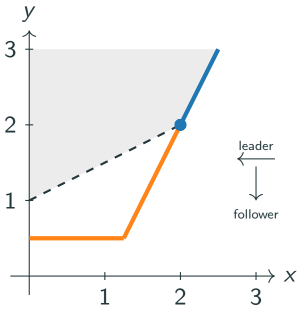
Independence of irrelevant constraints (Kleinert et al. 2021; Macal and Hurter 1997)

$$\min_{x,y\in\mathbb{R}} \quad x$$

$$\text{s.t.} \quad y \geq 0.5x + 1, \; x \geq 0$$

$$y \in \arg\min_{\bar{y}\in\mathbb{R}} \{\bar{y} : \bar{y} \geq 2x - 2, \; \bar{y} \geq 0.5\}$$

Optimal solution: $(2, 2)$

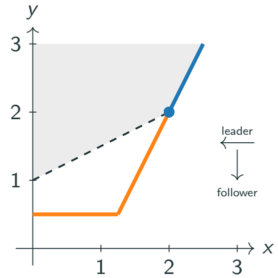


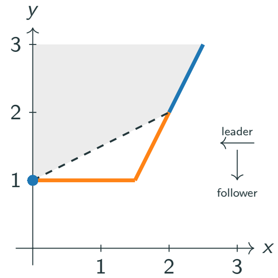
Independence of irrelevant constraints (Kleinert et al. 2021; Macal and Hurter 1997)

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

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

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





History

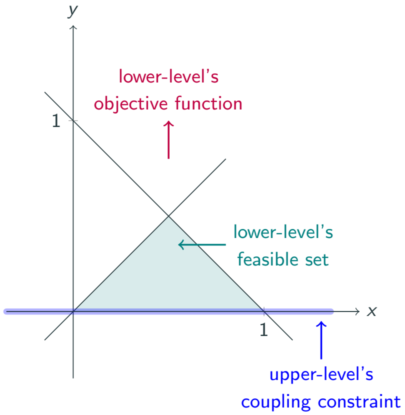
- Jeroslow (1985): general multilevel models
- Corollary: NP-hardness of the LP-LP bilevel problem
- Hansen, Jaumard, and Gilles Savard (1992): LP-LP bilevel problems are strongly NP-hard
  - reduction from KERNEL
- Vicente, Gilles Savard, and Joaquim Júdice (1994): even checking whether a given point is a local minimum of a bilevel problem is NP-hard

Bilevel vs. mixed-integer optimization

Audet et al. (1997)

- Binary constraint $x \in \{0, 1\}$
- Can be modeled by an additional variable y,
- the upper-level constraint $y = 0$,
- and the lower-level problem

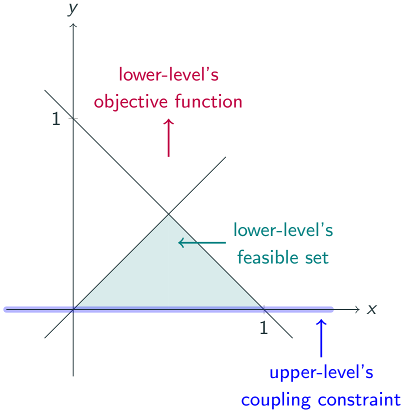
$$y = \arg\max_{\bar{y}} \{\bar{y} \colon \bar{y} \leq x, \ \bar{y} \leq 1 - x\}$$



Audet et al. (1997)

- Binary constraint $x \in \{0, 1\}$
- Can be modeled by an additional variable y,
- the upper-level constraint $y = 0$,
- and the lower-level problem

$$y = \arg\max_{\bar{y}} \{\bar{y} : \bar{y} \leq x, \ \bar{y} \leq 1 - x\}$$



Consequence: linear optimization problems with binary variables are a special case of bilevel LPs.

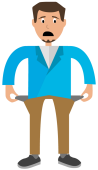
2. Solution Concepts: Overview

2. Solution Concepts

Leader: Alice x

decides first

anticipates follower (Bob)



Follower: Bob y

decides second (of course)

Bilevel optimization

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

Bilevel optimization

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

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

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

A different problem?

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

with lower level

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

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

with lower level

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

Example: Dempe, Kalashnikov, et al. (2015)

Consider the bilevel problem

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

with

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

Example: Dempe, Kalashnikov, et al. (2015)

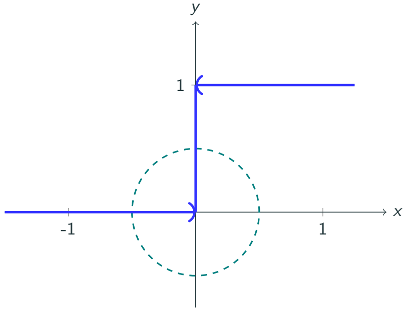
Consider the bilevel problem

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

with

$$S(x) = \arg\min_y \{-xy : 0 \leq y \leq 1\}$$

Best response of the follower

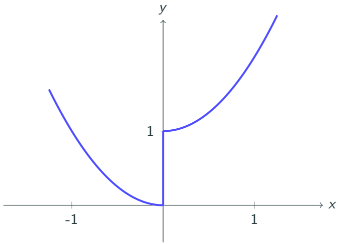


Example: Dempe, Kalashnikov, et al. (2015)

Best response of the follower

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

Mapping $x \mapsto F(x, S(x))$



- This is not a function and its minimum is unclear since it depends on the response $y \in S(x)$ of the follower if the leader chooses $x = 0$.
- For the follower, all responses $y \in S(0) = [0,1]$ are optimal
- The optimal lower-level solution is not unique.
- If the follower chooses $y = 0$, the optimal leader's decision is $x = 0$, leading to an objective function value of the leader of 0.
- However, if the follower chooses $y = 1$, the objective function value of the leader is 1, which is worse than 0 from the point of view of the leader.

Definition (Optimistic bilevel problem)

The problem

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

with the lower-level problem given by

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

is called the optimistic bilevel problem.

The pessimistic world: without coupling constraints

Definition (Pessimistic bilevel problem without coupling constraints)

The Problem

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

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

The pessimistic world: with coupling constraints

Definition (Pessimistic bilevel problem with coupling constraints)

The Problem

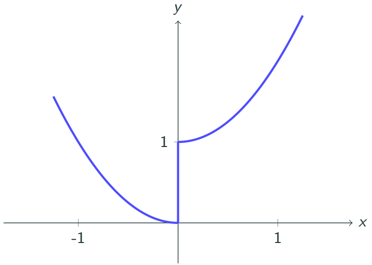
$$\begin{aligned} \min_{x \in X} \quad & F(x) \\ \text{s.t.} \quad & G(x, y) \geq 0 \quad \text{for all } y \in S(x), \end{aligned}$$

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

Some remarks

- The chosen solution concept (optimistic vs. pessimistic) is very important

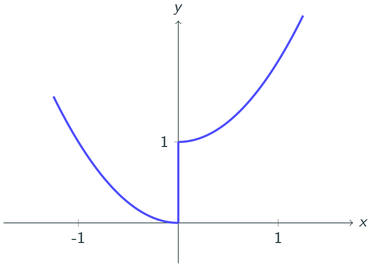
Mapping $x \mapsto F(x, S(x))$



Some remarks

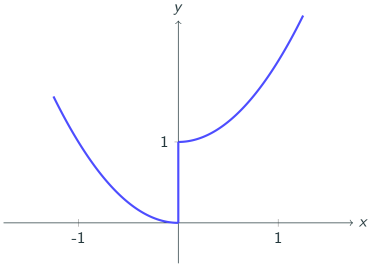
- The chosen solution concept (optimistic vs. pessimistic) is very important
- It even changes whether a solution exists or not

Mapping $x \mapsto F(x, S(x))$



Some remarks

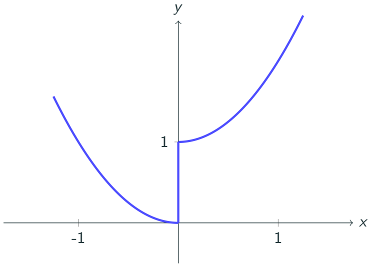
Mapping $x \mapsto F(x, S(x))$



- The chosen solution concept (optimistic vs. pessimistic) is very important
- It even changes whether a solution exists or not
- Example: the optimal solution $(x, y) = (0, 0)$ with objective function value 0 is attained in the last example if one considers the optimistic bilevel problem

Some remarks

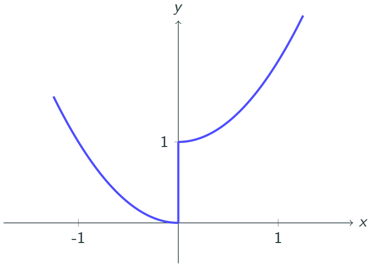
Mapping $x \mapsto F(x, S(x))$



- The chosen solution concept (optimistic vs. pessimistic) is very important
- It even changes whether a solution exists or not
- Example: the optimal solution $(x, y) = (0, 0)$ with objective function value 0 is attained in the last example if one considers the optimistic bilevel problem
- For all other choices of $y \in S(0) = [0, 1]$, the bilevel problem is not solvable since the infimum 0 of the upper-level's objective function is not attained anymore

Some remarks

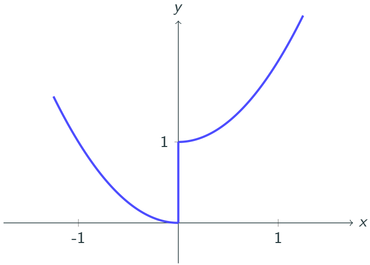
Mapping $x \mapsto F(x, S(x))$



- The chosen solution concept (optimistic vs. pessimistic) is very important
- It even changes whether a solution exists or not
- Example: the optimal solution $(x, y) = (0, 0)$ with objective function value 0 is attained in the last example if one considers the optimistic bilevel problem
- For all other choices of $y \in S(0) = [0, 1]$, the bilevel problem is not solvable since the infimum 0 of the upper-level's objective function is not attained anymore
- This, in particular, also applies to the pessimistic bilevel problem in this example

Some remarks

Mapping $x \mapsto F(x, S(x))$



- The chosen solution concept (optimistic vs. pessimistic) is very important
- It even changes whether a solution exists or not
- Example: the optimal solution $(x, y) = (0, 0)$ with objective function value 0 is attained in the last example if one considers the optimistic bilevel problem
- For all other choices of $y \in S(0) = [0, 1]$, the bilevel problem is not solvable since the infimum 0 of the upper-level's objective function is not attained anymore
- This, in particular, also applies to the pessimistic bilevel problem in this example
- If the lower-level solution is unique for all $x \in \Omega_x$, both the pessimistic and the optimistic variants of the bilevel problem coincide

Definition (Graph of the solution set mapping)

The set

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

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

Definition (Local and global optimal solution)

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

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

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

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

Definition (Local and global optimal solution)

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

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

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

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

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

J. Opl Res. Soc. Vol. 32, pp. 783 to 792, 1981

0160-5682/81/090783-10$02.00/0

A Representation and Economic Interpretation of a Two-Level Programming Problem

JOSÉ FORTUNY-AMAT and BRUCE McCARL

Graduate School of Administration, University of California, Riverside, California, U.S.A. and Purdue University, West Lafayette, Indiana, U.S.A.

This paper first presents a formulation for a class of hierarchial problems that show a two-stage decision making process; this formulation is termed multilevel programming and could be defined, in general, as a mathematical programming problem (master) containing other multilevel programs in the constraints (subproblems). A two-level problem is analyzed in detail, and we develop a solution procedure that replaces the subproblem by its Kuhn–Tucker conditions and then further transforms it into a mixed integer quadratic programming problem by exploiting the disjunctive nature of the complementary slackness conditions.

An example problem is solved and the economic implications of the formulation and its solution are reviewed.

Single-Level Reformulations

- Fortuny-Amat and McCarl (1981): The beginning of reformulating bilevel problems as single-level problems
- Always the same main idea: replace the lower-level problem with its optimality condition
- Afterward, solve the “ordinary” single-level problem

Single-Level Reformulations

- Fortuny-Amat and McCarl (1981): The beginning of reformulating bilevel problems as single-level problems
- Always the same main idea: replace the lower-level problem with its optimality condition
- Afterward, solve the "ordinary" single-level problem

Three main techniques

1. Use the optimal value function of the lower-level problem
2. Use the KKT conditions of the lower-level problem
3. Use the strong-duality theorem for the lower-level problem

Single-Level Reformulation using the Optimal Value Function

Consider the general optimistic bilevel problem

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

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

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

Single-Level Reformulation using the Optimal Value Function

By using the optimal value function

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

Single-Level Reformulation using the Optimal Value Function

By using the optimal value function

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

we can equivalently re-write the problem as

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

Single-Level Reformulation using the Optimal Value Function

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

- Looks like a usual single-level problem

Single-Level Reformulation using the Optimal Value Function

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

- Looks like a usual single-level problem
- However, the problem is the optimal value function $\varphi : \mathbb{R}^{n_x} \to \mathbb{R}$

Single-Level Reformulation using the Optimal Value Function

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

- Looks like a usual single-level problem
- However, the problem is the optimal value function $\varphi : \mathbb{R}^{n_x} \to \mathbb{R}$
- Evaluation: solve the lower-level problem for a given x

Single-Level Reformulation using the Optimal Value Function

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

- Looks like a usual single-level problem
- However, the problem is the optimal value function $\varphi : \mathbb{R}^{n_x} \to \mathbb{R}$
- Evaluation: solve the lower-level problem for a given x
- In most cases: optimal value function is not known in algebraic, i.e., in closed, form

Single-Level Reformulation using the Optimal Value Function

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

- Looks like a usual single-level problem
- However, the problem is the optimal value function $\varphi : \mathbb{R}^{n_x} \to \mathbb{R}$
- Evaluation: solve the lower-level problem for a given x
- In most cases: optimal value function is not known in algebraic, i.e., in closed, form
- It is usually nonsmooth (even under strong assumptions)

KKT Reformulation for LP-LP Bilevel Problems

Most classic approach to obtain a single-level reformulation:

Exploit optimality conditions for the lower-level problem

KKT Reformulation for LP-LP Bilevel Problems

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
  - later more on CQs

Reminder: NLP basics

Consider the general nonlinear optimization problem (NLP)

$$\begin{aligned}\min_{x\in\mathbb{R}^n} \quad & f(x) \\ \text{s.t.} \quad & g_i(x) \geq 0, \quad i \in I = \{1,\ldots,m\}, \\ & h_j(x) = 0, \quad j \in J = \{1,\ldots,p\}.\end{aligned}$$

Reminder: NLP basics

Consider the general nonlinear optimization problem (NLP)

$$
\begin{aligned}
\min_{x \in \mathbb{R}^n} \quad & f(x) \\
\text{s.t.} \quad & g_i(x) \geq 0, \quad i \in I = \{1, \ldots, m\}, \\
& h_j(x) = 0, \quad j \in J = \{1, \ldots, p\}.
\end{aligned}
$$

Assumptions

- Objective function $f : \mathbb{R}^n \to \mathbb{R}$ is continuously differentiable
- Constraint functions $g_i : \mathbb{R}^n \to \mathbb{R}$, $i \in I$, and $h_j : \mathbb{R}^n \to \mathbb{R}$, $j \in J$, are continuously differentiable

Reminder: NLP basics

Consider the general nonlinear optimization problem (NLP)

$$\begin{aligned}\min_{x\in\mathbb{R}^n} \quad & f(x)\\ \text{s.t.} \quad & g_i(x) \geq 0, \quad i \in I = \{1,\dots,m\},\\ & h_j(x) = 0, \quad j \in J = \{1,\dots,p\}.\end{aligned}$$

Assumptions

- Objective function $f : \mathbb{R}^n \to \mathbb{R}$ is continuously differentiable
- Constraint functions $g_i : \mathbb{R}^n \to \mathbb{R}$, $i \in I$, and $h_j : \mathbb{R}^n \to \mathbb{R}$, $j \in J$, are continuously differentiable

Notation

- Feasible set is denoted by $\mathcal{F}$.

Let $B_\varepsilon(x^*) = \{x \in \mathbb{R}^n \colon \|x - x^*\| < \varepsilon\}$ be the open ε-ball around x^*.

Let $B_\varepsilon(x^*) = \{x \in \mathbb{R}^n \colon \|x - x^*\| < \varepsilon\}$ be the open ε-ball around x^*.

Definition ((Strict) Local Minimizer)

A point $x^* \in \mathbb{R}^n$ is called a local minimizer if x^* is feasible and if an $\varepsilon > 0$ exists such that $f(x) \geq f(x^*)$ for all $x \in \mathcal{F} \cap B_\varepsilon(x^*)$. The point is called a strict local minimizer if $f(x) > f(x^*)$ holds for all $x \in (\mathcal{F} \cap B_\varepsilon(x^*)) \setminus \{x^*\}$.

Reminder: NLP basics

Let $B_\varepsilon(x^*) = \{x \in \mathbb{R}^n \colon \|x - x^*\| < \varepsilon\}$ be the open ε-ball around x^*.

Definition ((Strict) Local Minimizer)

A point $x^* \in \mathbb{R}^n$ is called a local minimizer if x^* is feasible and if an $\varepsilon > 0$ exists such that $f(x) \geq f(x^*)$ for all $x \in \mathcal{F} \cap B_\varepsilon(x^*)$. The point is called a strict local minimizer if $f(x) > f(x^*)$ holds for all $x \in (\mathcal{F} \cap B_\varepsilon(x^*)) \setminus \{x^*\}$.

Besides local minimizers we will also consider global minimizers.

Definition ((Strict) Global Minimizers)

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

Definition (Active Inequality Constraints)

Let $x \in \mathcal{F}$ be a feasible point. Then, the set

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

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

Definition (Active Inequality Constraints)

Let $x \in \mathcal{F}$ be a feasible point. Then, the set

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

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

Definition (Abadie Constraint Qualification)

We say that a feasible point $x \in \mathcal{F}$ satisfies the Abadie constraint qualification (ACQ) if $T_X(x) = T_{\text{lin}}(x)$ holds.

Reminder: NLP basics

Definition (Active Inequality Constraints)

Let $x \in \mathcal{F}$ be a feasible point. Then, the set

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

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

Definition (Abadie Constraint Qualification)

We say that a feasible point $x \in \mathcal{F}$ satisfies the Abadie constraint qualification (ACQ) if $T_X(x) = T_{\text{lin}}(x)$ holds.

Definition (Lagrangian Function)

The function

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

is called Lagrangian function of the NLP.

Reminder: KKT Conditions, KKT Point, Lagrangian Multipliers

Consider the general NLP with continuously differentiable functions f, g, and h.

1. The conditions

$$\begin{aligned} \nabla_x \mathcal{L}(x, \lambda, \mu) &= 0, \\ h(x) &= 0, \\ \lambda \geq 0,\ g(x) \geq 0,\ \lambda^\top g(x) &= 0 \end{aligned}$$

   are called Karush–Kuhn–Tucker (or KKT) conditions of the NLP.

Reminder: KKT Conditions, KKT Point, Lagrangian Multipliers

Consider the general NLP with continuously differentiable functions f, g, and h.

1. The conditions

$$\begin{aligned} \nabla_x \mathcal{L}(x, \lambda, \mu) &= 0, \\ h(x) &= 0, \\ \lambda \geq 0,\ g(x) \geq 0,\ \lambda^\top g(x) &= 0 \end{aligned}$$

are called Karush–Kuhn–Tucker (or KKT) conditions of the NLP. Here and in what follows,

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

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

Reminder: KKT Conditions, KKT Point, Lagrangian Multipliers

Consider the general NLP with continuously differentiable functions f, g, and h.

1. The conditions

$$\begin{aligned} \nabla_x \mathcal{L}(x, \lambda, \mu) &= 0, \\ h(x) &= 0, \\ \lambda \geq 0, \ g(x) \geq 0, \ \lambda^\top g(x) &= 0 \end{aligned}$$

   are called Karush–Kuhn–Tucker (or KKT) conditions of the NLP. Here and in what follows,

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

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

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

Theorem (KKT Theorem under the Abadie CQ)

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

Definition (Linear Independence Constraint Qualification)

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

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

are linearly independent.

Theorem (KKT Theorem under the LICQ)

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

Back to bilevel problems

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

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

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

KKT reformulation of LP-LP bilevel problems

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

KKT reformulation of LP-LP bilevel problems

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

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

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

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

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

- 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

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

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

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

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?

Idea

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

How to solve the KKT reformulation?

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.

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

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

Reminder: An LP and its dual

Consider the linear optimization problem

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

with $c \in \mathbb{R}^n$, $b \in \mathbb{R}^m$, and $A \in \mathbb{R}^{m \times n}$.

Reminder: An LP and its dual

Consider the linear optimization problem

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

with $c \in \mathbb{R}^n$, $b \in \mathbb{R}^m$, and $A \in \mathbb{R}^{m\times n}$.

The dual problem of the above LP is the LP

$$\begin{aligned} \max_{\lambda\in\mathbb{R}^m} \quad & b^\top \lambda \\ \text{s.t.} \quad & A^\top \lambda \leq c. \end{aligned}$$

Reminder: the weak-duality theorem

Theorem

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

$$b^\top \lambda \leq c^\top x$$

holds.

Reminder: the strong-duality theorem

Theorem

Consider the pair of primal and dual LPs. Then, the following statements are equivalent:

1. *The primal and the dual problem both are feasible.*
2. *The primal and the dual problem both have optimal solutions* $x^* \in \mathbb{R}^n$ *and* $\lambda^* \in \mathbb{R}^m$ *and*

$$c^\top x^* = b^\top \lambda^*$$

   holds.
3. *The primal and the dual problem both have a finite optimal objective value.*

Duality for the lower level problem

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

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

Duality for the lower level problem

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

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

The dual problem of this x-parameterized lower-level problem is given by

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

Weak and strong duality

Weak duality

For a given decision x of the leader, weak duality of linear optimization states that

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

holds for every primal and dual feasible pair y and λ.

Weak and strong duality

Weak duality

For a given decision x of the leader, weak duality of linear optimization states that

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

holds for every primal and dual feasible pair y and λ.

Strong duality

By strong duality, we know that every such feasible pair is a pair of optimal solutions if

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

holds.

Consequently, we can reformulate the bilevel problem as

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

KKT vs. strong duality

The KKT reformulation and the strong-duality based reformulation are equivalent:

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

Excursus

How to Really Solve a Mixed-Integer Linear Problem?

Nonlinear but convex lower-level problems

We now consider the bilevel problem

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

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

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

"Convexity" assumptions

- $y \mapsto f(x, y)$ is a convex function
- $y \mapsto g(x, y)$ is a concave function for all $x \in X$, i.e., for all feasible leader's decisions
- Y is a "simple" convex set (e.g., variable bounds)

Definition (Slater's constraint qualification for the lower level)

For a given upper-level feasible point $x \in X$ of the bilevel problem we say that Slater's constraint qualification holds for the lower-level problem

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

if there exists a (so-called Slater) point $\hat{y}(x)$ with $g_i(x, \hat{y}(x)) > 0$ for all $i = 1, \ldots, \ell$.

Slater's condition for the lower level

Definition (Slater's constraint qualification for the lower level)

For a given upper-level feasible point $x \in X$ of the bilevel problem we say that Slater's constraint qualification holds for the lower-level problem

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

if there exists a (so-called Slater) point $\hat{y}(x)$ with $g_i(x, \hat{y}(x)) > 0$ for all $i = 1, \ldots, \ell$.

One further assumption

We assume that all constraint functions g_i, $i = 1, \ldots, \ell$, are nonlinear.

Slater's condition for the lower level

Definition (Slater's constraint qualification for the lower level)

For a given upper-level feasible point $x \in X$ of the bilevel problem we say that Slater's constraint qualification holds for the lower-level problem

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

if there exists a (so-called Slater) point $\hat{y}(x)$ with $g_i(x, \hat{y}(x)) > 0$ for all $i = 1, \dots, \ell$.

One further assumption

We assume that all constraint functions g_i, $i = 1, \dots, \ell$, are nonlinear.

One further remark

If the lower-level problem has equality constraints $h(x, y) = 0$, a Slater point only has to be feasible w.r.t. these constraints, i.e., $h(x, \hat{y}) = 0$ has to hold.

KKT Reformulation for Parametric Convex Lower-Level

- Let Slater's constraint qualification hold for all upper-level feasible x
- Re-write the bilevel problem using the KKT conditions of the lower-level problem

KKT Reformulation for Parametric Convex Lower-Level

- Let Slater's constraint qualification hold for all upper-level feasible x
- Re-write the bilevel problem using the KKT conditions of the lower-level problem

$$
\begin{aligned}
\min_{x,y,\lambda} \quad & F(x,y) \\
\text{s.t.} \quad & x \in X, \\
& \nabla_y \mathcal{L}(x,y,\lambda) = \nabla_y f(x,y) - \sum_{i=1}^{\ell} \lambda_i \nabla_y g_i(x,y) = 0, \\
& g(x,y) \geq 0, \\
& \lambda \geq 0, \\
& \lambda^\top g(x,y) = 0.
\end{aligned}
$$

One part of the "equivalence"

Theorem (Dempe and Dutta (2012))

Let (x^*, y^*) be a global optimal solution of the bilevel problem and assume that the lower-level problem is a convex optimization problem that satisfies Slater's constraint qualification for x^*. Then, the point (x^*, y^*, λ^*) is a global optimal solution of the single-level KKT reformulation for every

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

Theorem (Dempe and Dutta (2012))

Let (x^*, y^*) be a global optimal solution of the bilevel problem and assume that the lower-level problem is a convex optimization problem that satisfies Slater's constraint qualification for x^*. Then, the point (x^*, y^*, λ^*) is a global optimal solution of the single-level KKT reformulation for every

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

Proof.

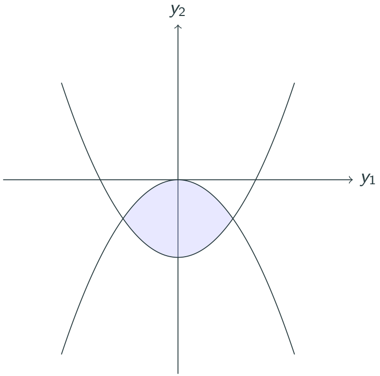
Since the x^*-parameterized lower-level problem is convex and since this parametric convex problem satisfies Slater's constraint qualification for the given x^*, the KKT theorem for convex problems implies that $\lambda^* \in \Lambda(x^*, y^*)$ holds if and only if $(x^*, y^*) \in \operatorname{gph}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

What if Slater's condition is violated (Dempe and Dutta 2012)

The feasible region of the x-parameterized convex lower-level problem for $x = 1$.

x-parameterized convex lower-level problem:

$$\min_{y_1, y_2} \quad y_1 \quad \text{s.t.} \quad y_1^2 - y_2 \le x, \; y_1^2 + y_2 \le 0$$



What if Slater's condition is violated (Dempe and Dutta 2012)

$$\min_{y_1,y_2} \quad y_1 \quad \text{s.t.} \quad y_1^2 - y_2 \le x,\ y_1^2 + y_2 \le 0.$$

- If $x = 0$, the only feasible point of this lower-level problem is $y = (y_1, y_2) = (0, 0)$

What if Slater's condition is violated (Dempe and Dutta 2012)

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

- If $x = 0$, the only feasible point of this lower-level problem is $y = (y_1, y_2) = (0, 0)$
- Thus, Slater's constraint qualification is violated

What if Slater's condition is violated (Dempe and Dutta 2012)

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

- If $x = 0$, the only feasible point of this lower-level problem is $y = (y_1, y_2) = (0, 0)$
- Thus, Slater's constraint qualification is violated
- For $x \geq 0$ (this will be our upper-level constraint later on), the lower-level's optimal solutions are given by

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

What if Slater's condition is violated (Dempe and Dutta 2012)

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

- If $x = 0$, the only feasible point of this lower-level problem is $y = (y_1, y_2) = (0, 0)$
- Thus, Slater's constraint qualification is violated
- For $x \geq 0$ (this will be our upper-level constraint later on), the lower-level's optimal solutions are given by

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

- For $x > 0$, the Lagrangian multipliers are given by

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

What if Slater's condition is violated (Dempe and Dutta 2012)

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

- If $x = 0$, the only feasible point of this lower-level problem is $y = (y_1, y_2) = (0, 0)$
- Thus, Slater's constraint qualification is violated
- For $x \geq 0$ (this will be our upper-level constraint later on), the lower-level's optimal solutions are given by

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

- For $x > 0$, the Lagrangian multipliers are given by

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

- If $x = 0$, the problem does not satisfy Slater's constraint qualification and the KKT conditions are not satisfied.

What if Slater's condition is violated (Dempe and Dutta 2012)

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

- If $x = 0$, the only feasible point of this lower-level problem is $y = (y_1, y_2) = (0, 0)$
- Thus, Slater's constraint qualification is violated
- For $x \geq 0$ (this will be our upper-level constraint later on), the lower-level's optimal solutions are given by

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

- For $x > 0$, the Lagrangian multipliers are given by

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

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

What if Slater's condition is violated (Dempe and Dutta 2012)

Consider now the upper-level problem

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

where $S(x)$ is again the solution-set mapping of the previously discussed lower-level problem.

What if Slater's condition is violated (Dempe and Dutta 2012)

Consider now the upper-level problem

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

where $S(x)$ is again the solution-set mapping of the previously discussed lower-level problem.

- Unique global optimal solution
$$x = 0, \quad y = (0, 0)$$
with objective function value 0.
- Moreover: no local optimal solutions

What if Slater's condition is violated (Dempe and Dutta 2012)

The lower level revisited

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

What if Slater's condition is violated (Dempe and Dutta 2012)

The lower level revisited

$$\min_{y_1, y_2} \quad y_1 \quad \text{s.t.} \quad y_1^2 - y_2 \le x, \; y_1^2 + y_2 \le 0$$

The Lagrangian of the lower-level problem reads

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

What if Slater's condition is violated (Dempe and Dutta 2012)

The lower level revisited

$$\min_{y_1, y_2} \quad y_1 \quad \text{s.t.} \quad y_1^2 - y_2 \le x, \; y_1^2 + y_2 \le 0$$

The Lagrangian of the lower-level problem reads

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

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

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

What if Slater's condition is violated (Dempe and Dutta 2012)

The KKT reformulation thus reads

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

What if Slater's condition is violated (Dempe and Dutta 2012)

KKT reformulation

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

What if Slater's condition is violated (Dempe and Dutta 2012)

KKT reformulation

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

- $(x, y(x), \lambda(x))$ is, by construction, feasible for the MPCC for $x > 0$

What if Slater's condition is violated (Dempe and Dutta 2012)

KKT reformulation

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

- $(x, y(x), \lambda(x))$ is, by construction, feasible for the MPCC for $x > 0$
- The corresponding objective function value of the bilevel problem converges to 0 for $x \to 0$.

What if Slater's condition is violated (Dempe and Dutta 2012)

KKT reformulation

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

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

What if Slater's condition is violated (Dempe and Dutta 2012)

KKT reformulation

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

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

Take-home message

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

The other part of the “equivalence”

Theorem (Dempe and Dutta (2012))

Let (x^*, y^*, λ^*) be a global optimal solution of the KKT reformulation and let the lower-level problem be convex. Moreover, suppose that Slater’s constraint qualification is satisfied for the lower-level problem for every $x \in X$. Then, (x^*, y^*) is a global optimal solution of the bilevel problem.

Theorem (Dempe and Dutta (2012))

Let (x^*, y^*, λ^*) *be a global optimal solution of the KKT reformulation and let the lower-level problem be convex. Moreover, suppose that Slater's constraint qualification is satisfied for the lower-level problem for every* $x \in X$*. Then,* (x^*, y^*) *is a global optimal solution of the bilevel problem.*

Remark

We will soon see in the proof that we really need that Slater's condition holds for all $x \in X$ and not only for $x = x^*$.

The other part of the "equivalence": proof

Proof.

Suppose that (x^*, y^*, λ^*) is a global optimal solution of the KKT reformulation.

The other part of the "equivalence": proof

Proof.

Suppose that (x^*, y^*, λ^*) is a global optimal solution of the KKT reformulation.

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

Proof.

Suppose that (x^*, y^*, λ^*) is a global optimal solution of the KKT reformulation.

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

Assume now that (x^*, y^*) is not a global optimal solution of the original bilevel problem.

Proof.

Suppose that (x^*, y^*, λ^*) is a global optimal solution of the KKT reformulation.

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

Assume now that (x^*, y^*) is not a global optimal solution of the original bilevel problem.

Then, there exists a point (x, y) with $x \in X$ and $y \in S(x)$ such that

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

holds.

The other part of the "equivalence": proof

Proof.

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

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

holds.

The other part of the “equivalence”: proof

Proof.

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

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

holds.

Consequently, (x, y, λ) is a feasible point for the KKT reformulation that has a better objective function value as (x^*, y^*, λ^*).

Proof.

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

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

holds.

Consequently, (x, y, λ) is a feasible point for the KKT reformulation that has a better objective function value as (x^*, y^*, λ^*).

This is a contradiction to the global optimality of (x^*, y^*, λ^*) and the claim follows. □

What if Slater's conditions is missing (Dempe and Dutta 2012)

We consider the bilevel problem

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

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

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

What if Slater's conditions is missing (Dempe and Dutta 2012)

We consider the bilevel problem

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

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

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

- $y = 0$ is the only feasible solution
- Thus, $y = 0$ is the uniquely determined global optimal solution of the lower-level problem (independent of the leader's decision x).

What if Slater's conditions is missing (Dempe and Dutta 2012)

We consider the bilevel problem

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

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

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

- $y = 0$ is the only feasible solution
- Thus, $y = 0$ is the uniquely determined global optimal solution of the lower-level problem (independent of the leader's decision x).
- This means that there exists no x for which Slater's constraint qualification holds for the lower-level problem.

We consider the bilevel problem

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

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

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

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

What if Slater's conditions is missing (Dempe and Dutta 2012)

KKT reformulation

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

What if Slater's conditions is missing (Dempe and Dutta 2012)

KKT reformulation

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

- All feasible solutions of this MPCC are of the form $(0, 0, \lambda)$ with $\lambda \geq 0$.

What if Slater's conditions is missing (Dempe and Dutta 2012)

KKT reformulation

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

- All feasible solutions of this MPCC are of the form $(0, 0, \lambda)$ with $\lambda \geq 0$.
- Since the objective function does not depend on λ,
  all these points are also global optimal solutions of the MPCC.

What if Slater's conditions is missing (Dempe and Dutta 2012)

KKT reformulation

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

- All feasible solutions of this MPCC are of the form $(0, 0, \lambda)$ with $\lambda \geq 0$.
- Since the objective function does not depend on λ,
  all these points are also global optimal solutions of the MPCC.
- None of them correspond to the optimal solution $(1, 0)$ of the bilevel problem.

What about local solutions?

Attention!

- One needs to be very careful when solving the KKT reformulation only to local optimality
- There exist problems for which the KKT reformulation has local minima that do not correspond to local minima of the bilevel problem
- Thus: On the level of local minima, the bilevel problem and its KKT reformulation are not equivalent
- Details: Dempe and Dutta (2012)

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

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 J. F. Bard (1998).

- 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 J. F. Bard (1998).

Theorem

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

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.

Since $S(x)$ is a singleton for all $x \in \Omega_x$, the optimal value function $\varphi(x)$ is a well-defined function. By using the strong-duality theorem, 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 $\lambda^1, \ldots, \lambda^s$ be the set of all the dual polyhedron's vertices, i.e., the set of vertices of the polyhedron defined by

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

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

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

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.

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.

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

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

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

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.

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, 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*.*

Solutions appear at vertices of the HPR: proof

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

Solutions appear at vertices of the HPR: proof

Let $(x^1, y^1), \ldots, (x^r, y^r)$ 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

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

Solutions appear at vertices of the HPR: proof

Let $(x^1, y^1), \ldots, (x^r, y^r)$ 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

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

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 $i \in \{1, \dots, r\}$ 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.

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.

Setting

- One of the first proposed algorithms to solve LP-LP bilevel problems
- Bialas and Karwan (1984)

Setting

- One of the first proposed algorithms to solve LP-LP bilevel problems
- Bialas and Karwan (1984)

Consider the LP-LP bilevel problem

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

- One of the first proposed algorithms to solve LP-LP bilevel problems
- Bialas and Karwan (1984)

Consider the LP-LP bilevel problem

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

Assumptions

The bilevel-feasible set is non-empty and bounded and $S(x)$ is a singleton for all $x \in \Omega_x$.

Main idea

- Use that a bilevel-optimal solution is attained at one of the vertices of the shared constraint set Ω
- We carry out a search over the vertices of Ω to find a solution
  - Similar to the simplex method for LPs

Main idea

- Use that a bilevel-optimal solution is attained at one of the vertices of the shared constraint set Ω
- We carry out a search over the vertices of Ω to find a solution
    - Similar to the simplex method for LPs

Consider the high-point relaxation

$$\begin{aligned} \min_{x,y} \quad & c_x^\top x + c_y^\top y \\ \text{s.t.} \quad & Ax \geq a, \\ & Cx + Dy \geq b. \end{aligned}$$

The vertices of the HPR

Let us denote with

$$(x^1, y^1), (x^2, y^2), \ldots, (x^r, y^r)$$

the ordered set of vertices of Ω, i.e., of basic feasible solutions of the high-point relaxation.

The vertices of the HPR

Let us denote with

$$(x^1, y^1), (x^2, y^2), \dots, (x^r, y^r)$$

the ordered set of vertices of Ω, i.e., of basic feasible solutions of the high-point relaxation.

The ordering is chosen so that

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

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

The vertices of the HPR

Let us denote with

$$(x^1, y^1), (x^2, y^2), \dots, (x^r, y^r)$$

the ordered set of vertices of Ω, i.e., of basic feasible solutions of the high-point relaxation.

The ordering is chosen so that

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

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

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

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

The vertices of the HPR

Let us denote with

$$(x^1, y^1), (x^2, y^2), \ldots, (x^r, y^r)$$

the ordered set of vertices of Ω, i.e., of basic feasible solutions of the high-point relaxation.

The ordering is chosen so that

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

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

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

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

In other words:

- Find the first vertex in the ordered list whose y-component is an optimal solution of the follower's problem.
- Then, (x^{K^*}, y^{K^*}) is a global optimal solution of the LP-LP bilevel problem.

The Kth best algorithm

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

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

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

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

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

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

5: **end if**

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

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

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

7: Set $i \leftarrow i + 1$ and choose (x^i, y^i) with $c_x^\top x^i + c_y^\top y^i = \min_{x,y} \left\{c_x^\top x + c_y^\top y \colon (x, y) \in W\right\}$.
Go to Step 2.

Some remarks

- Uniqueness of the follower's problem is required in Step 2 and 3, where we check if the current vertex is bilevel feasible.
- A crucial and costly part of the algorithm (that we do not discuss here) is the computation of all adjacent extreme points in Step 6.
- For more details; see J. F. Bard (1998).

The KKT reformulation revisited

We know: the general LP-LP bilevel problem

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

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

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

The basic idea behind LP-LP bilevel branch-and-bound

Start with solving the problem

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

The basic idea behind LP-LP bilevel branch-and-bound

Start with solving the problem

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

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

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

The basic idea behind LP-LP bilevel branch-and-bound

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

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

holds.

The basic idea behind LP-LP bilevel branch-and-bound

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

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

holds.

Take such an i and construct two new sub-problems:

The basic idea behind LP-LP bilevel branch-and-bound

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

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

holds.

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

$$\lambda_i = 0$$

is added

The basic idea behind LP-LP bilevel branch-and-bound

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

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

holds.

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

$$\lambda_i = 0$$

is added and one in which the constraint

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

is added.

The basic idea behind LP-LP bilevel branch-and-bound

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

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

holds.

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

$$\lambda_i = 0$$

is added and one in which the constraint

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

is added.

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

A bit of notation

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

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

Branch-and-bound for LP-LP bilevel problems

1: $u \leftarrow +\infty$ and $Q \leftarrow \{(\emptyset, \emptyset)\}$.

2: **while** $Q \neq \emptyset$ **do**

3: Choose any $(P, D) \in Q$ and set $Q \leftarrow Q \setminus \{(P, D)\}$.

4: Solve the node problem for P and D.

5: **if** the node problem for P and D is infeasible **then** go to Step 2.

6: Let $(\bar{x}, \bar{y}, \bar{\lambda})$ denote the solution of node's problem for P and D.

7: **if** $c_x^\top \bar{x} + c_y^\top \bar{y} \geq u$ **then** go to Step 2.

8: **if** $(\bar{x}, \bar{y}, \bar{\lambda})$ satisfies $\lambda_i(C_{i\cdot} x + D_{i\cdot} y - b_i) = 0$ for all $i \in \{1, \dots, \ell\}$ **then**

9: Set $(x^*, y^*, \lambda^*) \leftarrow (\bar{x}, \bar{y}, \bar{\lambda})$ as well as $u \leftarrow c_x^\top x^* + c_y^\top y^*$ and go to Step 2.

10: **end if**

11: Choose any $i \in \{1, \dots, \ell\}$ with $\lambda_i(C_{i\cdot} x + D_{i\cdot} y - b_i) > 0$. Set $Q \leftarrow Q \cup \{(P \cup \{i\}, D), (P, D \cup \{i\})\}$.

12: **end while**

13: **if** $u < +\infty$ **then**

14: Return the optimal solution (x^*, y^*, λ^*).

15: **else**

16: Return the statement "The given LP-LP bilevel problem is infeasible."

17: **end if**

Definition (Relaxation)

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

Definition (Relaxation)

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

- The easiest way to obtain a relaxation is to simply delete constraints from a given set of constraints.

Definition (Relaxation)

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

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

Correctness

The problem

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

for given P and D

Correctness

The problem

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

for given P and D is a relaxation of the problem

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

Lemma (Bounding Lemma)

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

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

Furthermore, the infeasibility of the relaxation (R) *implies the infeasibility of Problem* (N)*.*

Lemma (Bounding Lemma)

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

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

Furthermore, the infeasibility of the relaxation (R) *implies the infeasibility of Problem* (N)*.*

Proof.

Both statements immediately follow from the definition of a relaxation. □

Lemma (Branching Lemma)

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

Theorem (Correctness Theorem)

Suppose that the root-node relaxation of the KKT reformulation is bounded. Then, the branch-and-bound algorithm terminates after a finite number of visited nodes with a global optimal solution of of the KKT reformulation or with the correct indication of infeasibility.

Theorem (Correctness Theorem)

Suppose that the root-node relaxation of the KKT reformulation is bounded. Then, the branch-and-bound algorithm terminates after a finite number of visited nodes with a global optimal solution of of the KKT reformulation or with the correct indication of infeasibility.

Proof.

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

How to implement this method?

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

How to implement this method?

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

A set of non-negative variables $x_1, \ldots, x_n$ is called a special ordered set of type 1 if there exists at most one index $i \in \{1, \ldots, n\}$ with $x_i > 0$ and $x_j = 0$ for all $j \neq i$.

How to implement this method?

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

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

We denote this property of the set of variables $x_1, \dots, x_n$ in the following via

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

How to implement this method?

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

A set of non-negative variables $x_1, \ldots, x_n$ is called a special ordered set of type 1 if there exists at most one index $i \in \{1, \ldots, n\}$ with $x_i > 0$ and $x_j = 0$ for all $j \neq i$.

We denote this property of the set of variables $x_1, \ldots, x_n$ in the following via

$$\mathrm{SOS1}(x_1, \ldots, x_n).$$

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

How to implement this method?

If we now introduce the non-negative auxiliary variables

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

How to implement this method?

If we now introduce the non-negative auxiliary variables

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

we can state the complementarity conditions

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

How to implement this method?

If we now introduce the non-negative auxiliary variables

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

we can state the complementarity conditions

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

equivalently as

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

How to implement this method?

If we now introduce the non-negative auxiliary variables

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

we can state the complementarity conditions

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

equivalently as

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

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

Consider now the general bilevel mixed-integer linear problem

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

where the vectors c_x, c_y, d, a, b and matrices A, C, D are defined as before.

Mixed-Integer Linear Bilevel Problems

Consider now the general bilevel mixed-integer linear problem

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

where the vectors c_x, c_y, d, a, b and matrices A, C, D are defined as before.

The sets X and Y specify integrality constraints on a subset of x- and y-variables.

Hardness

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

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

Hardness

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

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

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

Hardness

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

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

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

This means,

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

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

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

Hardness

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

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

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

This means,

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

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

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

The optimal value function $\varphi(x)$ thus corresponds to a parametric MILP in this case.

Hence, it is nonconvex, not continuous, and in general very difficult to describe.

Hardness

- It is now NP-hard to check whether a given point (x, y) is a feasible solution of the bilevel MILP.

- It is now NP-hard to check whether a given point (x, y) is a feasible solution of the bilevel MILP.
- Jeroslow (1985) showed that k-level discrete optimization problems are Σ_k^p-hard, even when the variables are binary and all constraints are linear.

Hardness

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

Attainability issues

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

(i) only upper-level variables are discrete,

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

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

Attainability issues

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

(i) only upper-level variables are discrete,

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

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

- Assumption: all discrete variables are bounded and the bilevel-feasible set is non-empty
- For Case (i) and (ii), an optimal solution always exists and Case (i) can be reduced to a mixed-integer linear program
- Case (ii) can be "reduced" to a linear trilevel problem
- However, for Case (iii), Moore and J. F. Bard (1990) and also Vicente, G. Savard, and J. Júdice (1996) provide examples that demonstrate that the bilevel feasible region may not be closed and, hence, the optimal solution may not be attainable.

The example by Köppe, Queyranne, and Ryan (2010)

Consider

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

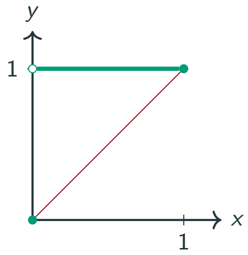
The example by Köppe, Queyranne, and Ryan (2010)

Consider

$$\inf_{0 \le x \le 1,\, y} \left\{ x - y \colon y \in \operatorname*{arg\,min}_{\bar{y} \in \mathbb{Z}} \{\bar{y} \colon \bar{y} \ge x,\ 0 \le \bar{y} \le 1\} \right\},$$

This is equivalent to

$$\inf_x \{x - \lceil x \rceil \colon 0 \le x \le 1\}.$$



The example by Köppe, Queyranne, and Ryan (2010)

Consider

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

This is equivalent to

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

The example by Köppe, Queyranne, and Ryan (2010)

Consider

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

This is equivalent to

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

- The infimum is -1, which is never attained

The example by Köppe, Queyranne, and Ryan (2010)

Consider

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

This is equivalent to

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

- The infimum is -1, which is never attained
- Frequent assumption in the literature: all linking variables are discrete

The example by Köppe, Queyranne, and Ryan (2010)

Consider

$$\inf_{0\le x\le 1, y} \left\{ x - y \colon y \in \operatorname*{arg\,min}_{\bar{y}\in\mathbb{Z}} \{\bar{y} \colon \bar{y} \ge x,\ 0 \le \bar{y} \le 1\} \right\}.$$

This is equivalent to

$$\inf_x \{x - \lceil x \rceil \colon 0 \le x \le 1\}.$$

- The infimum is -1, which is never attained
- Frequent assumption in the literature: all linking variables are discrete
- Non-linking upper-level variables can be moved to the lower level (Bolusani and Ralphs 2020; Tahernejad, Ralphs, and DeNegre 2020), which effectively translates the latter assumption into "all upper-level variables are discrete".

The example by Moore and Bard

We consider the discrete bilevel problem

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

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

$$\begin{aligned} -25x + 20\bar{y} \le 30, \quad x + 2\bar{y} \le 10, \\ 2x - \bar{y} \le 15, \quad 2x + 10\bar{y} \ge 15. \end{aligned}$$

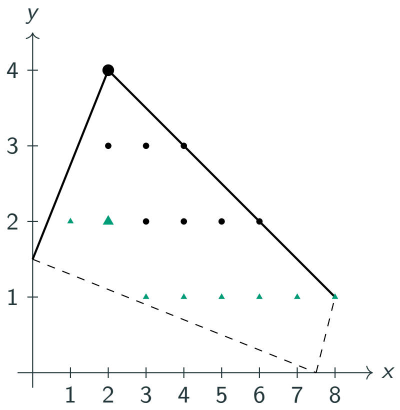
The example by Moore and Bard

We consider the discrete bilevel problem

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

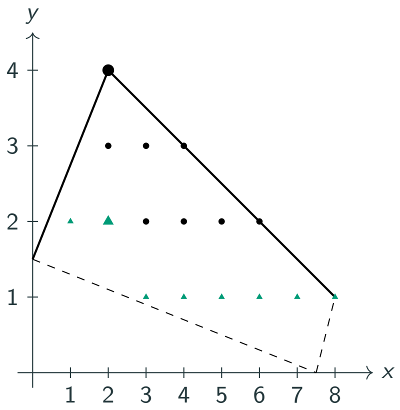
where $\mathcal{P}$ is a polytope defined by

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



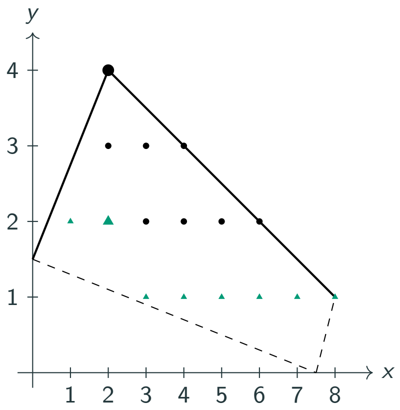
The example by Moore and Bard

- Discrete points are feasible for the high-point relaxation.



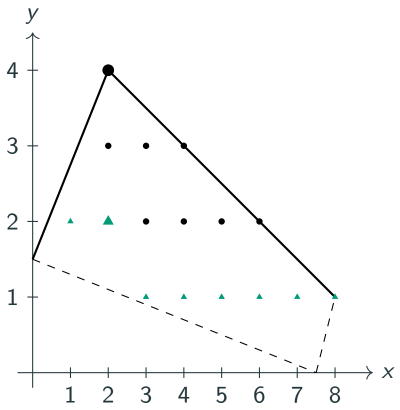
The example by Moore and Bard

- Discrete points are feasible for the high-point relaxation.
- The point $(2, 4)$ is the optimal solution of the high-point relaxation



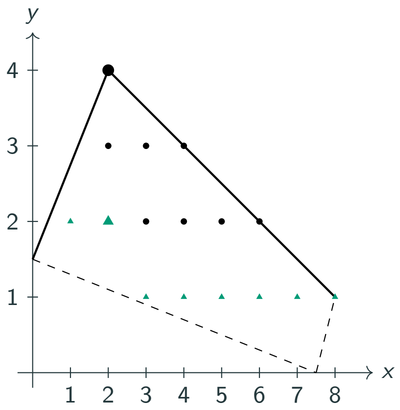
The example by Moore and Bard

- Discrete points are feasible for the high-point relaxation.
- The point $(2, 4)$ is the optimal solution of the high-point relaxation
- The point $(2, 2)$ is the optimal solution of the bilevel MILP.



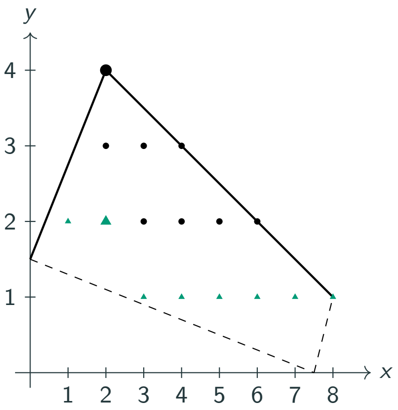
The example by Moore and Bard

- Discrete points are feasible for the high-point relaxation.
- The point $(2, 4)$ is the optimal solution of the high-point relaxation
- The point $(2, 2)$ is the optimal solution of the bilevel MILP.
- Triangles represent bilevel-feasible solutions



The example by Moore and Bard

- Discrete points are feasible for the high-point relaxation.
- The point $(2, 4)$ is the optimal solution of the high-point relaxation
- The point $(2, 2)$ is the optimal solution of the bilevel MILP.
- Triangles represent bilevel-feasible solutions
- Dashed lines represent the feasible region of the bilevel LP in which the integrality constraints on the upper- and lower-level variables are "relaxed"



Setting

Consider the mixed-integer linear bilevel problem

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

Setting

Consider the mixed-integer linear bilevel problem

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

- The variables x and y are split in $x = (x_{C_x}, x_{I_x})$ and $y = (y_{C_y}, y_{I_y})$.

Setting

Consider the mixed-integer linear bilevel problem

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

- The variables x and y are split in $x = (x_{C_x}, x_{I_x})$ and $y = (y_{C_y}, y_{I_y})$.
- x_{C_x} and y_{C_y} are the upper- as well as lower-level variables that are continuous-valued

Setting

Consider the mixed-integer linear bilevel problem

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

- The variables x and y are split in $x = (x_{C_x}, x_{I_x})$ and $y = (y_{C_y}, y_{I_y})$.
- x_{C_x} and y_{C_y} are the upper- as well as lower-level variables that are continuous-valued
- x_{I_x} and y_{I_y} are upper- as well as lower-level variables that are integer-valued

Setting

Consider the mixed-integer linear bilevel problem

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

- The variables x and y are split in $x = (x_{C_x}, x_{I_x})$ and $y = (y_{C_y}, y_{I_y})$.
- x_{C_x} and y_{C_y} are the upper- as well as lower-level variables that are continuous-valued
- x_{I_x} and y_{I_y} are upper- as well as lower-level variables that are integer-valued

Assumption

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

Setting

Integrality is encoded by using the sets X and Y via

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

- n_{x_I} and n_{y_I}: number of integer variables in the upper- as well as the lower-level problem

Fathoming rules

Goal: design a branch-and-bound method for bilevel MILPs.

Fathoming rules

Goal: design a branch-and-bound method for bilevel MILPs.

Let us first recap the main fathoming rules that we use in the classic branch-and-bound method for linear bilevel problems.

Fathoming rules

Goal: design a branch-and-bound method for bilevel MILPs.

Let us first recap the main fathoming rules that we use in the classic branch-and-bound method for linear bilevel problems.

There, we fathomed nodes according to the following three rules:

Rule 1 The problem at the current node is infeasible.

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

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

Fathoming rules

Goal: design a branch-and-bound method for bilevel MILPs.

Let us first recap the main fathoming rules that we use in the classic branch-and-bound method for linear bilevel problems.

There, we fathomed nodes according to the following three rules:

Rule 1 The problem at the current node is infeasible.

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

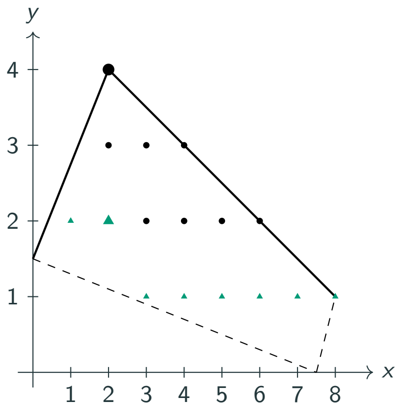
Rule 3 The problem at the current node is feasible w.r.t. all complementarity constraints.

Since we branch on integers again, Rule 3 translates into ...

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

The example by Moore and Bard—revisited

- Bilevel solution: $(x^*, y^*) = (2, 2)$
- Optimal objective function value $F(x^*, y^*) = -22$.
- Optimal solution of the problem in which we "relax" all integrality conditions is the point $(x, y) = (8, 1)$.
- This point is even integer- and bilevel-feasible.
- The corresponding objective function value, however, is $F(x, y) = -18$.
- This is worse than the optimal objective function value.



Two crucial observations

Observation

The solution of the continuous "relaxation" of the mixed-integer linear bilevel problem does not provide a valid lower bound on the solution of the original problem.

Two crucial observations

Observation

The solution of the continuous "relaxation" of the mixed-integer linear bilevel problem does not provide a valid lower bound on the solution of the original problem.

Observation

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

Let's do it anyway ...

These two observations already render Rule 2 and Rule 3 invalid in general.

Let's do it anyway . . .

These two observations already render Rule 2 and Rule 3 invalid in general.

The following example (also taken from Moore and J. F. Bard (1990)) shows what goes wrong if Rule 3 is applied although it is invalid.

Let's do it anyway ...

These two observations already render Rule 2 and Rule 3 invalid in general.

The following example (also taken from Moore and J. F. Bard (1990)) shows what goes wrong if Rule 3 is applied although it is invalid.

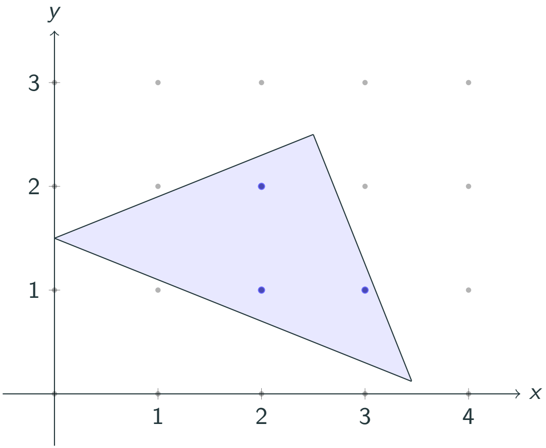
We consider the integer linear bilevel problem

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

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

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

Let's do it anyway . . .



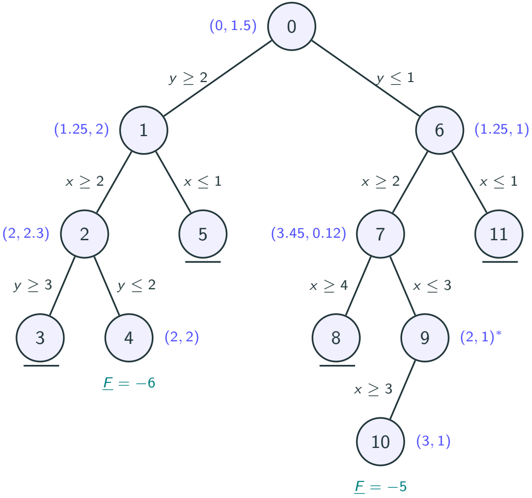
- Shared constraint set contains three integer-feasible points: $(2, 1)$, $(2, 2)$, and $(3, 1)$.
- If the leader chooses $x = 2$, the follower chooses $y = 2$, leading to $F = -6$.
- If the leader decides for $x = 3$, the follower optimally reacts with $y = 1$, leading to an objective function value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

Let's do it anyway . . .

Let us now consider what a classic depth-first branch-and-bound method would look like if we (as usual) branch on fractional integer variables and if "relaxations" are obtained by relaxing integrality restrictions.

Let's do it anyway . . .

Let us now consider what a classic depth-first branch-and-bound method would look like if we (as usual) branch on fractional integer variables and if "relaxations" are obtained by relaxing integrality restrictions.



Another observation

Thus, we can make the third main observation.

Observation

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

Some notation

- I_x and I_y: index sets of integer variables of the leader and the follower

Some notation

- I_x and I_y: index sets of integer variables of the leader and the follower
- U^x and U^y: $|I_x|$- as well as $|I_y|$-dimensional vectors of upper bounds for the integer variables of the leader and of the follower

Some notation

- I_x and I_y: index sets of integer variables of the leader and the follower
- U^x and U^y: $|I_x|$- as well as $|I_y|$-dimensional vectors of upper bounds for the integer variables of the leader and of the follower
- If an integer variable is not bounded from above in the original problem, the corresponding entry in U^x or U^y is set to ∞.

Some notation

- I_x and I_y: index sets of integer variables of the leader and the follower
- U^x and U^y: $|I_x|$- as well as $|I_y|$-dimensional vectors of upper bounds for the integer variables of the leader and of the follower
- If an integer variable is not bounded from above in the original problem, the corresponding entry in U^x or U^y is set to ∞.
- Assumption: all initial lower bounds of all integers variables are 0

Some notation

- I_x and I_y: index sets of integer variables of the leader and the follower
- U^x and U^y: $|I_x|$- as well as $|I_y|$-dimensional vectors of upper bounds for the integer variables of the leader and of the follower
- If an integer variable is not bounded from above in the original problem, the corresponding entry in U^x or U^y is set to ∞.
- Assumption: all initial lower bounds of all integers variables are 0
- Can be encoded using the sets X and Y of the original problem formulation.

Some notation

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

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

Some notation

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

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

The notation Y_0 is used to indicate that no other bounds than the original ones are imposed on the follower's integer variables.

Some notation

Note further that for a node k along the path from the root to node l, the problem associated to node l is derived from the problem of the node k by additionally imposing bounds on the integer variables, i.e.,

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

holds ...

Note further that for a node k along the path from the root to node l, the problem associated to node l is derived from the problem of the node k by additionally imposing bounds on the integer variables, i.e.,

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

holds ... which means that

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

as well as

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

holds.

Some notation

The sets

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

and

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

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

Some notation

For later reference we define the problem

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

with the lower-level problem

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

as the bilevel problem at node k in which the integrality constraints are omitted.

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

Some notation

The continuous high-point relaxation is given by

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

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

Bounding theorem 1

Theorem (Moore and J. F. Bard (1990))

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

Theorem (Moore and J. F. Bard (1990))

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

Proof.

Consider any successor node l of node k in the branch-and-bound tree, i.e., $X_l \subseteq X_k$ and $Y_l \subseteq Y_0$ holds.

Theorem (Moore and J. F. Bard (1990))

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

Proof.

Consider any successor node l of node k in the branch-and-bound tree, i.e., $X_l \subseteq X_k$ and $Y_l \subseteq Y_0$ holds. Let (x^l, y^l) be a global optimal solution of the mixed-integer linear bilevel problem associated with node l.

Theorem (Moore and J. F. Bard (1990))

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

Proof.

Consider any successor node l of node k in the branch-and-bound tree, i.e., $X_l \subseteq X_k$ and $Y_l \subseteq Y_0$ holds. Let (x^l, y^l) be a global optimal solution of the mixed-integer linear bilevel problem associated with node l. Assume now that $F(x^l, y^l) < F_k^{\text{hpr}}$ holds.

Theorem (Moore and J. F. Bard (1990))

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

Proof.

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

Bounding theorem 2

Theorem

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

Theorem

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

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

Bounding theorem 2: proof

Let again (x^l, y^l) be the solution of the mixed-integer linear bilevel problem associated with node l, which is a successor node of node k, i.e., $X_l \subseteq X_k$ and $Y_l \subseteq Y_k$ holds.

Bounding theorem 2: proof

Let again (x^l, y^l) be the solution of the mixed-integer linear bilevel problem associated with node l, which is a successor node of node k, i.e., $X_l \subseteq X_k$ and $Y_l \subseteq Y_k$ holds.

Assume again that $F(x^l, y^l) < F_k^{\text{hpr}}$ holds. This directly implies that (x^l, y^l) cannot be feasible for the high-point relaxation of the continuous relaxation of the mixed-integer linear bilevel problem at node k.

Bounding theorem 2: proof

Let again (x^l, y^l) be the solution of the mixed-integer linear bilevel problem associated with node l, which is a successor node of node k, i.e., $X_l \subseteq X_k$ and $Y_l \subseteq Y_k$ holds.

Assume again that $F(x^l, y^l) < F_k^{\text{hpr}}$ holds. This directly implies that (x^l, y^l) cannot be feasible for the high-point relaxation of the continuous relaxation of the mixed-integer linear bilevel problem at node k.

We consider the points (x', y') of the convex combination of (x^l, y^l) and (x^k, y^k), i.e.,

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

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

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

since F is linear.

Bounding theorem 2: proof

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

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

for $\lambda > 0$.

Bounding theorem 2: proof

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

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

for $\lambda > 0$.

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

Bounding corollary

Corollary

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

Bounding corollary

Corollary

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

Proof.

Relaxing all restrictions in Y_k is equivalent to replacing Y_k with Y_0. Thus, the first theorem applies. □

1: Set $k = 0$ and initialize X_k and Y_k with the bounds of the original mixed-integer linear bilevel problem. Set $R_k^x = \emptyset$, $R_k^y = \emptyset$, and $F^* = \infty$.

2: Solve (C-HPR). If this problem is infeasible go to Step 7. Otherwise, let F_k^{hpr} be the optimal objective function value. If $F_k^{\text{hpr}} \geq F^*$ holds, go to Step 7 as well.

3: Solve (C-BLP). If this problem is infeasible, go to Step 7. Otherwise, denote the solution as (x^k, y^k).

4: If (x^k, y^k) is integer-feasible, go to Step 5. Otherwise, select a fractional leader variable index $j \in I_x$ or a fractional follower variable index $j \in I_y$ and place a new bound on the selected variable. Set $k \leftarrow k + 1$ and update X_k or Y_k as well as R_k^x or R_k^y accordingly. Go to Step 2.

5: Fix $x = x^k$ and solve the follower's problem to obtain the overall bilevel feasible point $(x^k, \hat{y}^k)$. Compute $F(x^k, \hat{y}^k)$ and update $F^* = \min\{F^*, F(x^k, \hat{y}^k)\}$.

6: If $\underline{x}_j^k = \bar{x}_j^k$ for all $j \in I_x$ and if $\underline{y}_j^k = \bar{y}_j^k$ for all $j \in I_y$ holds, go to Step 7. Otherwise, select an integer variable $j \in I_x$ with $\underline{x}_j^k < \bar{x}_j^k$ or a $j \in I_y$ with $\underline{y}_j^k < \bar{y}_j^k$ and place a new bound on it. Set $k \leftarrow k + 1$ and update X_k or Y_k as well as R_k^x or R_k^y accordingly. Go to Step 2.

7: If no open node exists, go to Step 8. Otherwise, branch on the lastly added open node, set $k \leftarrow k + 1$, and update X_k or Y_k as well as R_k^x or R_k^y accordingly.

8: If $F^* = \infty$, the original mixed-integer linear bilevel problem is infeasible. Otherwise, F^* is the global optimal objective function value.

Proposition

If all follower variables are integer, the branch-and-bound algorithm finds the global optimal solution of the mixed-integer linear bilevel problem.

Proposition

If all follower variables are integer, the branch-and-bound algorithm finds the global optimal solution of the mixed-integer linear bilevel problem.

Proposition

Assume that an optimum exists for the mixed-integer linear bilevel problem and that all follower variables are continuous. If the fathoming rules 2 and 3 are used, the branch-and-bound algorithm always terminates with the global optimal solution.

7. Outlook

What you should have learned today

You should have learned . . .

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

I hope you had fun!

What we have not looked at

- Further branch-and-bound/branch-and-cut methods for bilevel MILPs
  - Jonathan F. Bard and Moore (1992), DeNegre and Ralphs (2009), Fischetti et al. (2017, 2018), Tahernejad, Ralphs, and DeNegre (2017), and Xu and Wang (2014), . . .

What we have not looked at

- Further branch-and-bound/branch-and-cut methods for bilevel MILPs
  - Jonathan F. Bard and Moore (1992), DeNegre and Ralphs (2009), Fischetti et al. (2017, 2018), Tahernejad, Ralphs, and DeNegre (2017), and Xu and Wang (2014), ...
- Pessimistic bilevel optimization
  - Wiesemann et al. (2013); Liu, Fan, et al. (2018); Liu, Fan, et al. (2020) ...

What we have not looked at

- Further branch-and-bound/branch-and-cut methods for bilevel MILPs
    - Jonathan F. Bard and Moore (1992), DeNegre and Ralphs (2009), Fischetti et al. (2017, 2018), Tahernejad, Ralphs, and DeNegre (2017), and Xu and Wang (2014), . . .
- Pessimistic bilevel optimization
    - Wiesemann et al. (2013); Liu, Fan, et al. (2018); Liu, Fan, et al. (2020) . . .
- Continuous and nonlinear bilevel optimization
    - Dempe (2002) and the very many references therein

What we have not looked at

- Further branch-and-bound/branch-and-cut methods for bilevel MILPs
  - Jonathan F. Bard and Moore (1992), DeNegre and Ralphs (2009), Fischetti et al. (2017, 2018), Tahernejad, Ralphs, and DeNegre (2017), and Xu and Wang (2014), ...
- Pessimistic bilevel optimization
  - Wiesemann et al. (2013); Liu, Fan, et al. (2018); Liu, Fan, et al. (2020) ...
- Continuous and nonlinear bilevel optimization
  - Dempe (2002) and the very many references therein
- Bilevel optimization under uncertainty
  - Besançon et al. (2019, 2020); Burtscheidt and Claus (2020); Burtscheidt, Claus, and Dempe (2020); Dempe, Ivanov, et al. (2017); Ivanov (2018); Jain, Ordonez, et al. (2008); Pita, Jain, Tambe, et al. (2010); Yanikoglu and Kuhn (2018)

What we have not looked at

- Further branch-and-bound/branch-and-cut methods for bilevel MILPs
  - Jonathan F. Bard and Moore (1992), DeNegre and Ralphs (2009), Fischetti et al. (2017, 2018), Tahernejad, Ralphs, and DeNegre (2017), and Xu and Wang (2014), . . .
- Pessimistic bilevel optimization
  - Wiesemann et al. (2013); Liu, Fan, et al. (2018); Liu, Fan, et al. (2020) . . .
- Continuous and nonlinear bilevel optimization
  - Dempe (2002) and the very many references therein
- Bilevel optimization under uncertainty
  - Besançon et al. (2019, 2020); Burtscheidt and Claus (2020); Burtscheidt, Claus, and Dempe (2020); Dempe, Ivanov, et al. (2017); Ivanov (2018); Jain, Ordonez, et al. (2008); Pita, Jain, Tambe, et al. (2010); Yanikoglu and Kuhn (2018)
- . . . and many more topics

What can you do in your PhD thesis on bilevel optimization?

- What about algorithms for bilevel problems with continuous linking variables?
- What about further cutting planes?
- What about presolve methods?
- What about computational pessimistic bilevel optimization?
- What about bilevel optimization under uncertainty?

- The community needs well-curated bilevel instance sets
- The community needs open-source software

A little advertising

Martin Schmidt, Yasmine Beck:

A Gentle and Incomplete Introduction to Bilevel Optimization

http://www.optimization-online.org/DB_FILE/2021/06/8450.pdf

SCAN ME

Enjoy the Germany match tonight!