Recent algorithmic advances in bilevel optimization

Martin Schmidt

@schmaidt

July 5, 2022 — EURO 2022, Espoo, Finland

The Team

Smart People

Yasmine Beck

Dan Bienstock

Veronika Grimm

Thomas Kleinert

Johannes Thürauf

The Mascot

Overview

What is Bilevel Optimization Anyway?

"Usual" optimization models

- a single decision maker
- one set of variables and constraints
- one objective function

Bilevel Optimization in a Nutshell

“Usual” optimization models

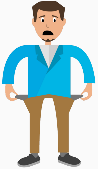
- a single decision maker
- one set of variables and constraints
- one objective function

Bilevel optimization

- two decision makers
- both interact in a hierarchical way

Hierarchical Decision Making

Leader: Alice x
decides first
anticipates follower (Bob)



Follower: Bob y
decides second (of course)

A Bit More Formal

Upper-level problem

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

A Bit More Formal

Upper-level problem

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

A Bit More Formal

Upper-level problem

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

Lower-level problem

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

A Bit More Formal

Upper-level problem

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

Lower-level problem

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

Different solution concepts: optimistic vs. pessimistic

The Linear-Linear Case

$$\min_{x \in \mathbb{R}^n, y \in \mathbb{R}^m} \quad c^\top x + d^\top y \quad \text{s.t.} \quad Ax + By \geq a, \; y \in \mathcal{S}(x)$$

The Linear-Linear Case

$$\min_{x \in \mathbb{R}^n, y \in \mathbb{R}^m} \quad c^\top x + d^\top y \quad \text{s.t.} \quad Ax + By \geq a, \; y \in \mathcal{S}(x)$$

$\mathcal{S}(x)$: set of optimal solutions of the x-parameterized linear problem

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

The Linear-Linear Case

$$\min_{x\in\mathbb{R}^n, y\in\mathbb{R}^m} \quad c^\top x + d^\top y \quad \text{s.t.} \quad Ax + By \geq a, \; y \in \mathcal{S}(x)$$

$\mathcal{S}(x)$: set of optimal solutions of the x-parameterized linear problem

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

- Strongly NP-hard problem (Hansen, Jaumard, Savard 1992)
- Checking local optimality is NP-hard (Vicente et al. 1994)
- Mixed-integer linear bilevel problems are Σ_p^2-hard (Lodi et al. 2014)
- Optimistic variant (Dempe 2002)

How to solve these problems: The KKT reformulation

The lower-level problem is an LP:

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

How to solve these problems: The KKT reformulation

The lower-level problem is an LP:

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

The KKT conditions

$$\begin{aligned} Cx + Dy &\geq b \\ \lambda \geq 0,\ D^\top \lambda &= f \\ \lambda^\top (Cx + Dy - b) &= 0 \end{aligned}$$

are both necessary and sufficient

How to solve these problems: The KKT reformulation

The lower-level problem is an LP:

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

The KKT conditions

$$\begin{aligned} Cx + Dy &\geq b \\ \lambda \geq 0,\ D^\top \lambda &= f \\ \lambda^\top (Cx + Dy - b) &= 0 \end{aligned}$$

are both necessary and sufficient

Single-level reformulation

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

KKT Reformulation

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

KKT Reformulation

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

- All is nice ...
- ... except for the nasty KKT complementarity conditions

$$\lambda^\top (Cx + Dy - b) = 0$$

How to deal with KKT complementarity conditions

$$\lambda^\top (Cx + Dy - b) = 0$$

How to deal with KKT complementarity conditions

$$\lambda^\top (Cx + Dy - b) = 0$$

That's a disjunction

$$\lambda_i = 0 \quad \vee \quad (Cx + Dy - b)_i = 0, \quad i \in \{1, \dots, \ell\}$$

Introduce a binary variable and some big-*M*s ...

$$\begin{aligned} Cx + Dy - b &\leq M_{\mathrm{P}}(1 - u) \\ \lambda &\leq M_{\mathrm{D}} u \\ u &\in \{0, 1\}^\ell \end{aligned}$$

Mixed-Integer Linear Reformulation

$$
\begin{aligned}
\min_{x,y,\lambda,u} \quad & c^\top x + d^\top y \\
\text{s.t.} \quad & Ax + By \geq a, \quad Cx + Dy \geq b \\
& \lambda \geq 0, \quad D^\top \lambda = f \\
& Cx + Dy - b \leq M_{\mathrm{P}}(1 - u) \\
& \lambda \leq M_{\mathrm{D}} u \\
& u \in \{0, 1\}^\ell
\end{aligned}
$$

Mixed-Integer Linear Reformulation

$$
\begin{aligned}
\min_{x,y,\lambda,u} \quad & c^\top x + d^\top y \\
\text{s.t.} \quad & Ax + By \geq a, \quad Cx + Dy \geq b \\
& \lambda \geq 0, \quad D^\top \lambda = f \\
& Cx + Dy - b \leq M_{\mathrm{P}}(1 - u) \\
& \lambda \leq M_{\mathrm{D}} u \\
& u \in \{0, 1\}^\ell
\end{aligned}
$$

But how to choose the nasty big-*M*s?

Pitfall #1: Heuristics do not work!

Solving Linear Bilevel Problems Using Big-Ms: Not All That Glitters Is Gold

Salvador Pineda and Juan Miguel Morales

***Abstract*—The most common procedure to solve a linear bilevel problem in the PES community is, by far, to transform it into an equivalent single-level problem by replacing the lower level with its KKT optimality conditions. Then, the complementarity conditions are reformulated using additional binary variables and large enough constants (big-Ms) to cast the single-level problem as a mixed-integer linear program that can be solved using optimization software. In most cases, such large constants are tuned by trial and error. We show, through a counterexample, that this widely used trial-and-error approach may lead to highly suboptimal solutions. Then, further research is required to properly select big-M values to solve linear bilevel problems.**

***Index Terms*—Bilevel programming, optimality conditions, mathematical program with equilibrium constraints (MPEC).**

in [5]. Dealing with the solution to this variant goes beyond the purposes of this letter and thus, we assume $d_i = 0$. This assumption is common in several applications of linear bilevel programming in the PES technical literature. For example, in long-term planning models formulated as bilevel problems [6], [7], [8], [9], the upper-level problem determines investment decisions to maximize investor's profit, while the lower-level problem yields the dispatch quantities to minimize operating cost. In most cases, upper-level constraints model maximum available capacities to be installed and/or budget limitations, but do not include lower-level dispatch variables.

Since the lower-level optimization problem is linear, it can be replaced with its KKT optimality conditions as follows:

Pitfall #2: It's really hard!

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

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

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

Abstract

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

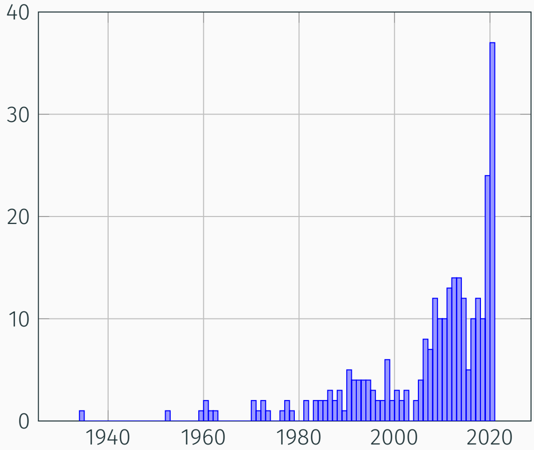
Remedy: It's not even required anymore!

Why there is no need to use a big-M in linear bilevel optimization: A computational study of two ready-to-use approaches

Thomas Kleinert[1,2] and Martin Schmidt[3]

Abstract. Linear bilevel optimization problems have gained increasing attention both in theory as well as in practical applications of Operations Research (OR) during the last years and decades. The latter is mainly due to the ability of this class of problems to model hierarchical decision processes. However, this ability makes bilevel problems also very hard to solve. Since no general-purpose solvers are available, a "best-practice" has developed in the applied OR community, in which not all people want to develop tailored algorithms but "just use" bilevel optimization as a modeling tool for practice. This best-practice is the big-M reformulation of the Karush–Kuhn–Tucker (KKT) conditions of the lower-level problem—an approach that has been shown to be highly problematic by Pineda and Morales (2019). Choosing invalid values for M yields solutions that may be arbitrarily bad. Checking the validity of the big-Ms is however shown to be as hard as solving the original bilevel problem in Kleinert et al. (2019). Nevertheless, due to its appealing simplicity, especially w.r.t. the required implementation effort, this ready-to-use approach still is the most popular method. Until now, there has been a lack of approaches that are competitive both in terms of implementation effort and computational cost.

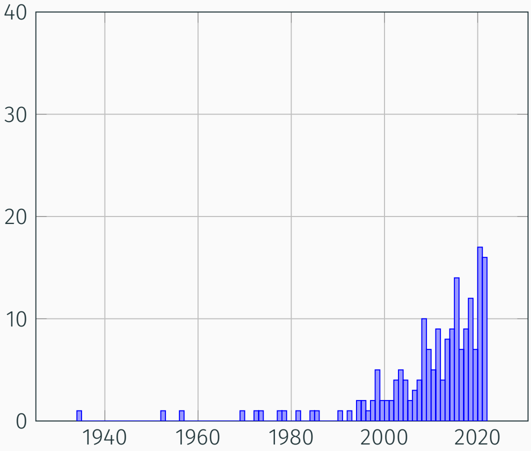
In this note we demonstrate that there is indeed another competitive ready-to-use approach: If the SOS-1 technique is applied to the KKT complementarity conditions, adding the simple additional root-node inequality developed by Kleinert et al. (2020) leads to a competitive performance—without having all the possible theoretical disadvantages of the big-M approach.



Based on the references in

A Survey on Mixed-Integer Programming Techniques in Bilevel Optimization

by Kleinert, Labbé, Ljubić, Schmidt



Based on the references in
A Survey on Bilevel Optimization Under Uncertainty
by Beck, Ljubić, Schmidt

Outer Approximation for MIQP-QP Bilevel Problems

Algorithms

- Branch-and-bound/cut techniques
    - Moore, Bard (1990)
    - Xu, Wang (2014)
    - Tahernejad, Ralphs, DeNegre (2016)
    - Lozano, Smith (2017)
    - Fischetti, Ljubić, Monaci, Sinnl (2017, 2018)
- Many problem-specific Benders-like approaches
    - Grimm et al. (2019)
    - Kleinert, Schmidt (2019)

Algorithms

- Branch-and-bound/cut techniques
  - Moore, Bard (1990)
  - Xu, Wang (2014)
  - Tahernejad, Ralphs, DeNegre (2016)
  - Lozano, Smith (2017)
  - Fischetti, Ljubić, Monaci, Sinnl (2017, 2018)
- Many problem-specific Benders-like approaches
  - Grimm et al. (2019)
  - Kleinert, Schmidt (2019)

Are there other classic approaches in MINLP?

Algorithms

- Branch-and-bound/cut techniques
  - Moore, Bard (1990)
  - Xu, Wang (2014)
  - Tahernejad, Ralphs, DeNegre (2016)
  - Lozano, Smith (2017)
  - Fischetti, Ljubić, Monaci, Sinnl (2017, 2018)
- Many problem-specific Benders-like approaches
  - Grimm et al. (2019)
  - Kleinert, Schmidt (2019)

Are there other classic approaches in MINLP?

Yes! Outer Approximation (Duran, Grossmann 1986; Fletcher, Leyffer 1994)

The MIQP-QP Bilevel Problem

$$
\begin{aligned}
\min_{x,y} \quad & q_u(x,y) = \frac{1}{2}x^\top H_u x + c_u^\top x + \frac{1}{2}y^\top G_u y + d_u^\top y \\
\text{s.t.} \quad & Ax + By \geq a \\
& x_i \in \mathbb{Z} \cap [x_i^-, x_i^+] \quad \text{for all } i \in I := \{1, \ldots, |I|\} \\
& x_i \in \mathbb{R} \quad \text{for all } i \in R := \{|I| + 1, \ldots, n_x\} \\
& y \in \arg\min_{\bar{y}} \left\{ q_l(\bar{y}) = \frac{1}{2}\bar{y}^\top G_l \bar{y} + d_l^\top \bar{y} : Cx_I + D\bar{y} \geq b,\ \bar{y} \in \mathbb{R}^{n_y} \right\}
\end{aligned}
$$

$$\begin{aligned}
\min_{x,y} \quad & q_u(x,y) = \frac{1}{2}x^\top H_u x + c_u^\top x + \frac{1}{2}y^\top G_u y + d_u^\top y \\
\text{s.t.} \quad & Ax + By \geq a \\
& x_i \in \mathbb{Z} \cap [x_i^-, x_i^+] \quad \text{for all } i \in I := \{1, \ldots, |I|\} \\
& x_i \in \mathbb{R} \quad \text{for all } i \in R := \{|I| + 1, \ldots, n_x\} \\
& y \in \arg\min_{\bar{y}} \left\{ q_l(\bar{y}) = \frac{1}{2}\bar{y}^\top G_l \bar{y} + d_l^\top \bar{y} : Cx_I + D\bar{y} \geq b,\ \bar{y} \in \mathbb{R}^{n_y} \right\}
\end{aligned}$$

Assumptions

1. All upper-level integer variables x_I are bounded
2. All linking variables are integer

Dual of the lower level

$$\begin{aligned} \max_{y,\lambda} \quad & \bar{g}(x_l; y, \lambda) = -\frac{1}{2} y^\top G_l y - (Cx_l - b)^\top \lambda \\ \text{s.t.} \quad & G_l y + d_l = D^\top \lambda, \quad \lambda \geq 0 \end{aligned}$$

Weak duality $q_l(y) \geq \bar{g}(x_l; y, \lambda)$ holds for all primal-dual feasible points

Strong duality can be ensured by

$$c(x_l, y, \lambda) := q_l(y) - \bar{g}(x_l; y, \lambda) = y^\top G_l y + d_l^\top y - b^\top \lambda + \lambda^\top C x_l \leq 0$$

Dual of the lower level

$$\begin{aligned} \max_{y,\lambda} \quad & \bar{g}(x_I; y, \lambda) = -\frac{1}{2} y^\top G_l y - (C x_I - b)^\top \lambda \\ \text{s.t.} \quad & G_l y + d_l = D^\top \lambda, \quad \lambda \geq 0 \end{aligned}$$

Weak duality $q_l(y) \geq \bar{g}(x_I; y, \lambda)$ holds for all primal-dual feasible points

Strong duality can be ensured by

$$c(x_I, y, \lambda) := q_l(y) - \bar{g}(x_I; y, \lambda) = y^\top G_l y + d_l^\top y - b^\top \lambda + \lambda^\top C x_I \leq 0$$

Equivalent MIQCQP

$$\begin{aligned} \min_{x,y,\lambda} \quad & q_u(x, y) \\ \text{s.t.} \quad & (x, y) \in P \\ & G_l y + d_l = D^\top \lambda, \quad \lambda \geq 0 \\ & c(x_I, y, \lambda) \leq 0 \end{aligned}$$

Convexification of the Single-Level MIQCQP

$$c(x_l, y, \lambda) := q_l(y) - \bar{g}(x_l; y, \lambda) = y^\top G_l y + d_l^\top y - b^\top \lambda + \lambda^\top C x_l \leq 0$$

Convexification of the Single-Level MIQCQP

$$c(x_l, y, \lambda) := q_l(y) - \bar{g}(x_l; y, \lambda) = y^\top G_l y + d_l^\top y - b^\top \lambda + \lambda^\top C x_l \leq 0$$

- All linking variables x_l are bounded integers
- Standard linearization of products of continuous and integer variables (Zare et al. 2019)

Convexification of the Single-Level MIQCQP

$$c(x_l, y, \lambda) := q_l(y) - \bar{g}(x_l; y, \lambda) = y^\top G_l y + d_l^\top y - b^\top \lambda + \lambda^\top C x_l \leq 0$$

- All linking variables x_l are bounded integers
- Standard linearization of products of continuous and integer variables (Zare et al. 2019)

Equivalent convex MIQCQP

$$\begin{aligned}
\min_{x,y,\lambda,w,s} \quad & q_u(x, y) \\
\text{s.t.} \quad & (x, y) \in P \\
& G_l y + d_l = D^\top \lambda, \quad \lambda \geq 0 \\
& \lambda^\top C x_l \text{ linearization} \\
& \text{Strong duality: } \hat{c}(y, \lambda, w) \leq 0
\end{aligned}$$

Let's denote the feasible set by $\mathcal{F}$

$$
\begin{aligned}
\min_{x,y,\lambda,w,s} \quad & q_u(x,y) \\
\text{s.t.} \quad & (x,y) \in P \\
& G_l y + d_l = D^\top \lambda, \quad \lambda \geq 0 \\
& \text{linearization of } \lambda^\top C x_I \\
& \bar{c}(\bar{y}^l; y, \lambda, w) \leq 0, \quad l = 0, \ldots, p-1
\end{aligned}
\tag{M^p}
$$

Add linear outer approximation cut $\bar{c}(\bar{y}^l; y, \lambda, w) \leq 0$ to the master problem after every iteration

$$
\bar{c}(\bar{y}; y, \lambda, w) := 2\bar{y}^\top G_l y + d_l^\top y - b^\top y + \sum_{j \in I} \sum_{r=1}^{\bar{r}_j} 2^{r-1} w_{jr} - \bar{y}^\top G_l \bar{y}
$$

$$
\begin{aligned}
\min_{x,y,\lambda,w,s} \quad & q_u(x, y) \\
\text{s.t.} \quad & (x, y) \in P \\
& G_l y + d_l = D^\top \lambda, \quad \lambda \geq 0 \\
& \text{linearization of } \lambda^\top C x_l \\
& \bar{c}(\bar{y}^l; y, \lambda, w) \leq 0, \quad l = 0, \dots, p-1
\end{aligned} \tag{M^p}
$$

Add linear outer approximation cut $\bar{c}(\bar{y}^l; y, \lambda, w) \leq 0$ to the master problem after every iteration

$$
\bar{c}(\bar{y}; y, \lambda, w) := 2\bar{y}^\top G_l y + d_l^\top y - b^\top y + \sum_{j \in I} \sum_{r=1}^{\bar{r}_j} 2^{r-1} w_{jr} - \bar{y}^\top G_l \bar{y}
$$

Lemma

For every iteration $p \geq 1$, $\mathcal{F} \subseteq \mathcal{M}^p \subseteq \mathcal{M}^{p-1}$ holds.

The Subproblem

- How to choose the linearization points $\bar{y}$?

The Subproblem

- How to choose the linearization points $\bar{y}$?
- Fix $x_I = x_I^p$ and $s = s^p$ and solve the convex QCQP

$$
\begin{aligned}
\min_{x_R, y, \lambda, w} \quad & q_u(x_I^p, x_R, y) \\
\text{s.t.} \quad & (x_I^p, x_R, y) \in P \\
& G_l y + d_l = D^\top \lambda, \quad \lambda \geq 0 \\
& w_{jr} = s_{jr}^p \sum_{i=1}^{m_l} c_{ij} \lambda_i, \quad j \in I, \; r \in [\bar{r}_j] \\
& \hat{c}(y, \lambda, w) \leq 0
\end{aligned}
\tag{S^p}
$$

The Subproblem

- How to choose the linearization points $\bar{y}$?
- Fix $x_I = x_I^p$ and $s = s^p$ and solve the convex QCQP

$$
\begin{aligned}
\min_{x_R, y, \lambda, w} \quad & q_u(x_I^p, x_R, y) \\
\text{s.t.} \quad & (x_I^p, x_R, y) \in P \\
& G_l y + d_l = D^\top \lambda, \quad \lambda \geq 0 \\
& w_{jr} = s_{jr}^p \sum_{i=1}^{m_l} c_{ij} \lambda_i, \quad j \in I, \ r \in [\bar{r}_j] \\
& \hat{c}(y, \lambda, w) \leq 0
\end{aligned}
\tag{S^p}
$$

Assumption

For every feasible subproblem (S^p), the Abadie CQ holds at the solution $(\bar{x}_R^p, \bar{y}^p, \bar{\lambda}^p, \bar{w}^p)$.

Why?

Lemma

Let $z^p = (x_I^p, x_R^p, y^p, \lambda^p, w^p, s^p)$ *be an optimal solution of the master problem* (M^p) *and assume that the subproblem* (S^p) *is feasible and has the optimal solution* $(\bar{x}_R^p, \bar{y}^p, \bar{\lambda}^p, \bar{w}^p)$*. Suppose further that the ACQ holds and consider the new master problem that is obtained by adding the outer-approximation cut* $\bar{c}(\bar{y}^p; y, \lambda, w) \leq 0$ *to* (M^p)*. Then, for any feasible point of the form* $z = (x_I^p, x_R, y, \lambda, w, s^p)$ *of this problem the following holds:*

$$q_u(x_I^p, x_R, y) \geq q_u(x_I^p, \bar{x}_R^p, \bar{y}^p).$$

Why?

Lemma

Let $z^p = (x_I^p, x_R^p, y^p, \lambda^p, w^p, s^p)$ *be an optimal solution of the master problem* (M^p) *and assume that the subproblem* (S^p) *is feasible and has the optimal solution* $(\bar{x}_R^p, \bar{y}^p, \bar{\lambda}^p, \bar{w}^p)$*. Suppose further that the ACQ holds and consider the new master problem that is obtained by adding the outer-approximation cut* $\bar{c}(\bar{y}^p; y, \lambda, w) \leq 0$ *to* (M^p)*. Then, for any feasible point of the form* $z = (x_I^p, x_R, y, \lambda, w, s^p)$ *of this problem the following holds:*

$$q_u(x_I^p, x_R, y) \geq q_u(x_I^p, \bar{x}_R^p, \bar{y}^p).$$

Difference to Fletcher, Leyffer (1994):
Using the solution of (S^p) does not explicitly cut off the integer solution x_I^p

The Infeasible Case

- z^p solves (M^p) $\rightarrow (y^p, \lambda^p)$ is primal-dual feasible for the lower level with fixed x_I^p
  $\rightarrow$ lower level has an optimal solution (satisfying strong duality)
- Infeasibility of (S^p): z^p is feasible for (S^p) without the strong-duality inequality

The Infeasible Case

- z^p solves $(\mathrm{M}^p) \rightarrow (y^p, \lambda^p)$ is primal-dual feasible for the lower level with fixed x_I^p
  $\rightarrow$ lower level has an optimal solution (satisfying strong duality)
- Infeasibility of (S^p): z^p is feasible for (S^p) without the strong-duality inequality

$$
\begin{aligned}
\min_{x_R, y, \lambda, w} \quad & \hat{c}(y, \lambda, w) \\
\text{s.t.} \quad & (x_I^p, x_R, y) \in P \\
& G_l y + d_l = D^\top \lambda, \quad \lambda \geq 0 \\
& w_{jr} = s_{jr}^p \sum_{i=1}^{m_l} c_{ij} \lambda_i, \quad j \in I,\ r \in [\bar{r}_j]
\end{aligned}
\tag{F^p}
$$

The Infeasible Case

- z^p solves $(\mathrm{M}^p) \to (y^p, \lambda^p)$ is primal-dual feasible for the lower level with fixed x_I^p
  $\to$ lower level has an optimal solution (satisfying strong duality)
- Infeasibility of (S^p): z^p is feasible for (S^p) without the strong-duality inequality

$$\begin{aligned} \min_{x_R, y, \lambda, w} \quad & \hat{c}(y, \lambda, w) \\ \text{s.t.} \quad & (x_I^p, x_R, y) \in P \\ & G_l y + d_l = D^\top \lambda, \quad \lambda \geq 0 \\ & w_{jr} = s_{jr}^p \sum_{i=1}^{m_l} c_{ij} \lambda_i, \quad j \in I, \; r \in [\bar{r}_j] \end{aligned} \tag{F^p}$$

Lemma

Let z^p be a solution of the master problem (M^p), let the subproblem (S^p) be infeasible, and let $(\bar{x}_R^p, \bar{y}^p, \bar{\lambda}, \bar{w})$ be a solution of the feasibility problem (F^p). Then, $\hat{c}(\bar{y}^p, \bar{\lambda}^p, \bar{w}^p) > 0$ and every $z = (x_I^p, x_R, y, \lambda, w, s^p) \in \mathcal{M}^p$ is infeasible for the constraint

$$\bar{c}(\bar{y}^p; y, \lambda, w) \leq 0.$$

Initialize $\phi = -\infty$, $\Phi = \infty$, and $p = 0$.
while $\phi < \Phi$ **do**
  Solve the master problem (M^p).
  if (M^p) is infeasible **then**
    Return "The bilevel problem is infeasible."
  else
    Let z^p be the optimal solution of (M^p) and set $\phi = q_u(x^p, y^p)$.
  end if
  Solve the subproblem (S^p), or the feasibility problem (F^p) if (S^p) is infeasible, and obtain $(\bar{x}_R^p, \bar{y}^p, \bar{\lambda}^p, \bar{w}^p)$.
  if (S^p) is feasible and $q_u(x_I^p, \bar{x}_R^p, \bar{y}^p) < \Phi$ **then**
    Set $z^* = (x_I^p, \bar{x}_R^p, \bar{y}^p, \bar{\lambda}^p, \bar{w}^p, s^p)$ and $\Phi = q_u(x_I^p, \bar{x}_R^p, \bar{y}^p)$.
  end if
  Add the outer approximation cut $\bar{c}(\bar{y}^p; y, \lambda, w) \leq 0$ to (M^p).
  Set $p \leftarrow p + 1$.
end while
Return z^*.

Theorem

The multitree outer approximation algorithm terminates after a finite number of iterations at an optimal solution of the original bilevel problem or with an indication that the problem is infeasible.

Theorem

The multitree outer approximation algorithm terminates after a finite number of iterations at an optimal solution of the original bilevel problem or with an indication that the problem is infeasible.

Algorithmic Enhancements

- Additional outer-approximation cuts
- Early termination of the master problem
- Warmstarting the master problem

Proposition

Let z^p be a solution of the master problem (M^p). Further, let $q_l^*(x_I^p)$ be the optimal objective value of the parametric lower-level problem for $x_I = x_I^p$ and let $(\tilde{x}_R^p, \tilde{y}^p)$ be the solution of the QCQP

$$\begin{aligned} \min_{x_R, y} \quad & q_u(x_I^p, x_R, y) \\ \text{s.t.} \quad & (x_I^p, x_R, y) \in P, \\ & q_l(y) \leq q_l^*(x_I^p). \end{aligned} \tag{$\star$}$$

Then, the following properties hold:

1. $(x_I^p, \tilde{x}_R^p, \tilde{y}^p) \in \mathcal{F}$, i.e., it is a bilevel feasible point of the original bilevel problem in the optimistic sense.
2. The solution $(\tilde{x}_R^p, \tilde{y}^p)$ of the subproblem (S^p) and the solution of Problem ($\star$) coincide in the sense that $(\bar{x}_R^p, \bar{y}^p) = (\tilde{x}_R^p, \tilde{y}^p)$.

Proposition

Let z^p be a solution of the master problem (M^p). Further, let $q_l^*(x_I^p)$ be the optimal objective value of the parametric lower-level problem for $x_I = x_I^p$ and let $(\tilde{x}_R^p, \tilde{y}^p)$ be the solution of the QCQP

$$\begin{aligned} \min_{x_R, y} \quad & q_u(x_I^p, x_R, y) \\ \text{s.t.} \quad & (x_I^p, x_R, y) \in P, \\ & q_l(y) \le q_l^*(x_I^p). \end{aligned} \tag{$\star$}$$

Then, the following properties hold:

1. $(x_I^p, \tilde{x}_R^p, \tilde{y}^p) \in \mathcal{F}$, i.e., it is a bilevel feasible point of the original bilevel problem in the optimistic sense.
2. The solution $(\tilde{x}_R^p, \tilde{y}^p)$ of the subproblem (S^p) and the solution of Problem ($\star$) coincide in the sense that $(\bar{x}_R^p, \bar{y}^p) = (\tilde{x}_R^p, \tilde{y}^p)$.

Solving (S^p) can be replaced with subsequently solving the lower level and the QCQP ($\star$).

A Single-Tree Variant

- LP/NLP based B&B by Quesada, Grossmann (1992): avoid multiple search trees
- Implementations: FilMINT by Abhishek et al. (2010), Bonmin by Bonami et al. (2008)

- LP/NLP based B&B by Quesada, Grossmann (1992): avoid multiple search trees
- Implementations: FilMINT by Abhishek et al. (2010), Bonmin by Bonami et al. (2008)

$$
\begin{aligned}
\min_{x,y,\lambda,w,s} \quad & q_u(x,y) \\
\text{s.t.} \quad & Ax + By \geq a, \quad Cx_I + Dy \geq b \\
& l \leq x_I \leq u \\
& G_l y + d_l = D^\top \lambda, \quad \lambda \geq 0 \\
& \lambda^\top C x_I \text{ linearization} \\
& \bar{c}(\bar{y}^l; y, \lambda, w) \leq 0, \quad l = 0, \ldots, p-1
\end{aligned}
\qquad (\mathrm{N}^p(l,u))
$$

Single-Tree Outer Approximation for MIQP-QP Bilevel Problems

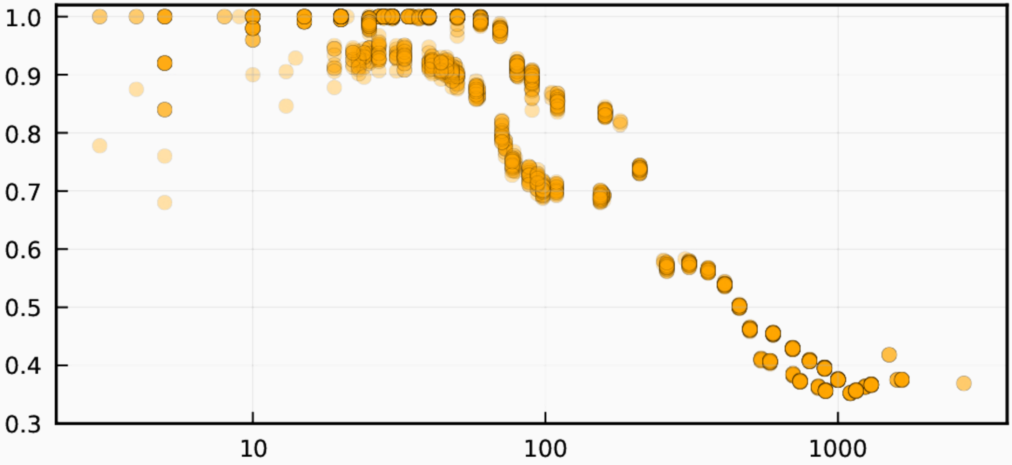
Initialize $\Phi = \infty$, $p = 0$, $l = x^-$, $u = x^+$, and $z^* =$ **none**.
Initialize the set of open node problems $\mathcal{O} := \{(\mathrm{N}^p(l,u))\}$
while $\mathcal{O} \neq \emptyset$ **do**
    Remove a QP $(\mathrm{N}^p(l,u))$ from $\mathcal{O}$ and solve it to obtain a solution $z^{l,u}$.
    if $(\mathrm{N}^p(l,u))$ is infeasible or $q_u(x^{l,u}, y^{l,u}) \geq \Phi$ **then**
        Subtree can be pruned. Continue.
    else if $z^{l,u}$ is integer feasible and $q_u(x^{l,u}, y^{l,u}) < \Phi$ **then**
        Set $x_I^p = x_I^{l,u}$ and $s^p = s^{l,u}$ and solve the subproblem (S^p) or the feasibility problem (F^p) to obtain $(\bar{x}_R^p, \bar{y}^p, \bar{\lambda}^p, \bar{w}^p)$.
        if (S^p) is feasible and $q_u(x_I^p, \bar{x}_R^p, \bar{y}^p) < \Phi$ **then**
            Set $z^* = (x_I^p, \bar{x}_R^p, \bar{y}^p, \bar{\lambda}^p, \bar{w}^p, s^p)$ and $\Phi = q_u(x_I^p, \bar{x}_R^p, \bar{y}^p)$.
        end if
        Re-add the problem: $\mathcal{O} \leftarrow \mathcal{O} \cup \{(\mathrm{N}^p(l,u))\}$.
        Add the outer-approximation cut $\bar{c}(\bar{y}^p; y, \lambda, w) \leq 0$ to all problems in $\mathcal{O}$.
        Set $p \leftarrow p + 1$.
    else
        Branch on a fractional x_i^p, $i \in I$, to obtain new bounds l^1, u^1 and l^2, u^2.
        Update $\mathcal{O} \leftarrow \mathcal{O} \cup \{\mathrm{M}^p(l^1, u^1), \mathrm{M}^p(l^2, u^2)\}$.
    end if
end while
Return z^* or, if z^* is **none**, return "The bilevel problem is infeasible".

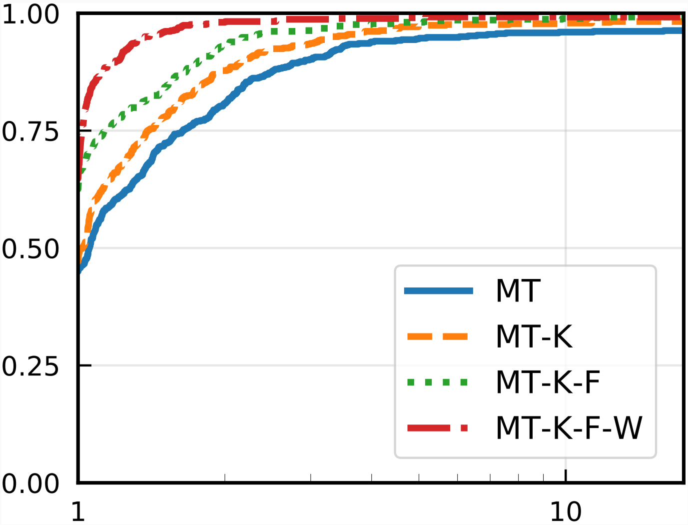
Software & Hardware Setup

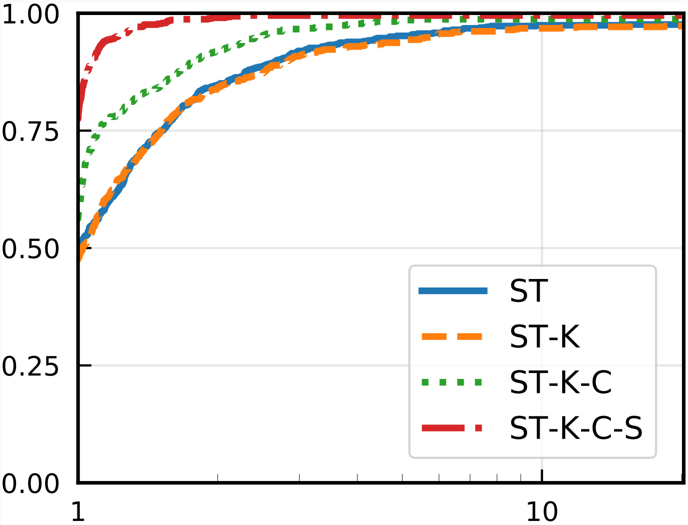
- C++ implementation
- Gurobi 9.0.1 for solving all (MI)(QC)QPs
- Increased NumericFocus (3)
- Tightened integer feasibility tolerance (10^{-9})
- Xeon E3-1240 v6 CPUs with 4 cores, 3.7 GHz, and 32 GB RAM
- Time limit of 1 h
- Implementation of the single-tree approach using callbacks and lazy constraints

- Test set from Kleinert, Schmidt (2019)
- Based on MIP-MIP instances from the literature
- Excluded instances that are too hard or too easy
- Filtered set of 423 instances

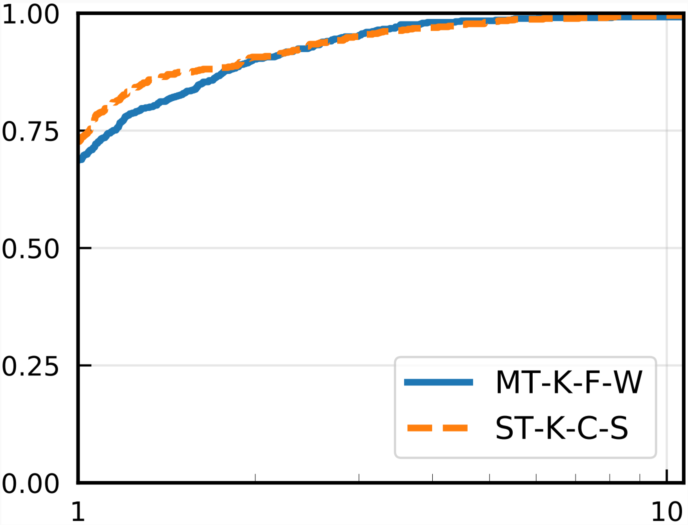
Test Set: Density vs. Size

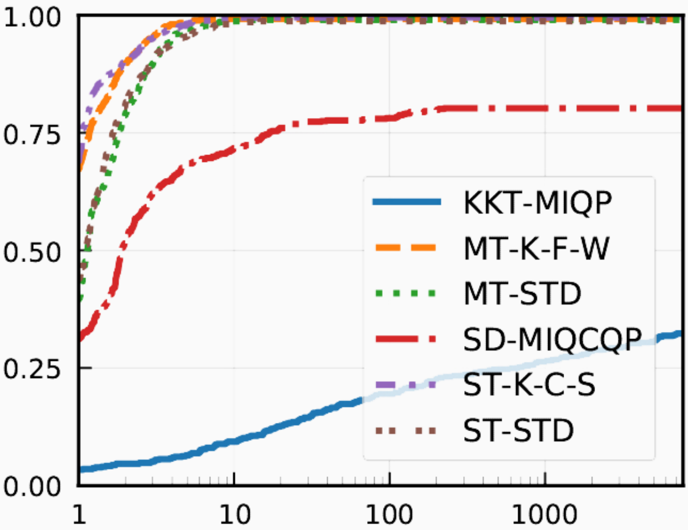



1.00
0.75
0.50
0.25
0.00
1
10
MT
MT-K
MT-K-F
MT-K-F-W


1.00
0.75
0.50
0.25
0.00
1
10
ST
ST-K
ST-K-C
ST-K-C-S

Single-Tree vs. Multi-Tree




1.00
0.75
0.50
0.25
0.00
1
10
100
1000
KKT-MIQP
MT-K-F-W
MT-STD
SD-MIQCQP
ST-K-C-S
ST-STD

An Open Problem: Continuous & Nonconvex Lower Levels

Upper-level problem

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

Lower-level problem

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

Who can solve this problem?

Upper level

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

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

Who can solve this problem?

Upper level

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

Lower level

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

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

Exact feasibility

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

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

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

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

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

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

Result #1

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

Exact feasibility

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

Result #1

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

Result #2

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

Exact feasibility

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

Result #1

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

Result #2

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

Result #3

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

Definition

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

Definition

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

Result #4

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

ε-feasibility

Definition

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

Result #4

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

Result #5

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

Result #3 (revisited)

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

ε-feasibility

Result #3 (revisited)

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

Result #6

Let $\varepsilon \geq 2^{-2^{n-1}}$ and suppose that we allow for ε-feasible follower's solutions.

Then, the optimistic optimal solution of the bilevel problem is given by $x_o^* = (\bar{x}_1, \bar{x}_2)$ with an optimal objective function value of $F_o^* = \bar{x}_1 + \bar{x}_2$.

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

ε-feasibility

Result #3 (revisited)

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

Result #6

Let $\varepsilon \geq 2^{-2^{n-1}}$ and suppose that we allow for ε-feasible follower's solutions.

Then, the optimistic optimal solution of the bilevel problem is given by $x_o^* = (\bar{x}_1, \bar{x}_2)$ with an optimal objective function value of $F_o^* = \bar{x}_1 + \bar{x}_2$.

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

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

Well ... and now?

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

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