Multilevel Optimization: Basics, an Application to the European Gas Market, and an Open Research Problem

Yasmine Beck, Daniel Bienstock, Holger Heitsch, René Henrion, Thomas Kleinert, Martin Schmidt, Johannes Thürauf

@schmaidt

June 23, 2022 — SIAM UKIE National Student Chapter Conference
What is Bilevel Optimization Anyway?

A Real-World Application:
The European Gas Market with Chance Constraints

An Open Problem:
Continuous & Nonconvex Lower Levels
What is Bilevel Optimization Anyway?
Bilevel Optimization in a Nutshell

“Usual” optimization models

- single decision maker
- one set of variables and constraints
- one objective function
Bilevel Optimization in a Nutshell

“Usual” optimization models

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

Bilevel optimization

- two decision makers
- both interact in a hierarchical way
Leader: Alice \( x \)
decides first
anticipates follower (Bob)

Follower: Bob \( y \)
decides second (of course)
Upper-level problem

\[
\min_x F(x, y) \\
\text{s.t. } G(x, y) \geq 0
\]
A Bit More Formal

Upper-level problem

\[
\begin{align*}
\text{“min” } & \quad F(x, y) \\
\text{s.t. } & \quad G(x, y) \geq 0, \quad y \in S(x)
\end{align*}
\]
Upper-level problem

\[
\text{"min}_{x} \ F(x, y) \\
\text{s.t.} \quad G(x, y) \geq 0, \quad y \in S(x)
\]

Lower-level problem

\[
\min_{y} \ f(x, y) \\
\text{s.t.} \quad g(x, y) \geq 0
\]
Upper-level problem

"\( \min_x \) \( F(x, y) \)

s.t. \( G(x, y) \geq 0, \quad y \in S(x) \)

Lower-level problem

\( \min_y f(x, y) \)

s.t. \( g(x, y) \geq 0 \)

Different solution concepts: optimistic vs. pessimistic
The Linear-Linear Case

\[ \min_{x \in \mathbb{R}^n, y \in \mathbb{R}^m} c^T x + d^T y \quad \text{s.t.} \quad Ax + By \geq a, \quad y \in \mathcal{S}(x) \]
The Linear-Linear Case

\[
\begin{align*}
\min_{x \in \mathbb{R}^n, y \in \mathbb{R}^m} & \quad c^T x + d^T y \\
\text{s.t.} & \quad A x + B y \geq a, \quad y \in S(x)
\end{align*}
\]

\(S(x)\) denotes the set of optimal solutions of the \(x\)-parameterized linear problem

\[
\begin{align*}
\min_y & \quad f^T y \\
\text{s.t.} & \quad D y \geq b - C x
\end{align*}
\]
The Linear-Linear Case

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

\(S(x)\) denotes the set of optimal solutions of the \(x\)-parameterized linear problem

\[
\min_y \quad f^T y \quad \text{s.t.} \quad Dy \geq b - Cx
\]

- **Strongly NP-hard** problem in general (Hansen, Jaumard, Savard 1992)
- **Optimistic** variant (Dempe 2002)
How to solve these problems: The KKT reformulation

The lower-level problem is an LP:

$$\min_y f^T y \quad \text{s.t.} \quad Dy \geq b - Cx$$
How to solve these problems: The KKT reformulation

The lower-level problem is an LP:

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

The KKT conditions

$$Cx + Dy \geq b$$
$$\lambda \geq 0, \quad D^T \lambda = f$$
$$\lambda^T (Cx + Dy - b) = 0$$

are both necessary and sufficient
How to solve these problems: The KKT reformulation

The lower-level problem is an LP:

$$\min_y \ f^T y \ \text{s.t.} \ Dy \geq b - Cx$$

The KKT conditions

$$Cx + Dy \geq b$$

$$\lambda \geq 0, \ D^T \lambda = f$$

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

are both necessary and sufficient

Single-level reformulation

$$\min_{x,y,\lambda} \ c^T x + d^T y$$

s.t. \ \ Ax + By \geq a, \ \ Cx + Dy \geq b$$

$$\lambda \in \Omega_D := \{\lambda \geq 0: D^T \lambda = f\}$$

$$\lambda^T (Cx + Dy - b) = 0$$
KKT Reformulation

\[
\begin{align*}
\min_{x,y,\lambda} & \quad c^\top x + d^\top y \\
\text{s.t.} & \quad Ax + By \geq a, \quad Cx + Dy \geq b \\
& \quad \lambda \in \Omega_D := \{\lambda \geq 0 : D^\top \lambda = f\} \\
& \quad \lambda^\top (Cx + Dy - b) = 0
\end{align*}
\]
KKT Reformulation

\[
\min_{x,y,\lambda} \quad c^T x + d^T y \\
\text{s.t.} \quad Ax + By \geq a, \quad Cx + Dy \geq b \\
\lambda \in \Omega_D := \{\lambda \geq 0 : D^T \lambda = f\} \\
\lambda^T (Cx + Dy - b) = 0
\]

- Be careful if the dual multipliers are not unique (Dempe, Dutta 2012)
- Otherwise, all is nice ...
- ... except for the nasty KKT complementarity conditions

\[
\lambda^T (Cx + Dy - b) = 0
\]
How to deal with KKT complementarity conditions

\[ \lambda^T (Cx + Dy - b) = 0 \]
How to deal with KKT complementarity conditions

\[ \lambda^T (C x + D y - b) = 0 \]

That’s a disjunction

\[ \lambda_i = 0 \ \vee \ (C x + D y - b)_i = 0, \ i \in \{1, \ldots, \ell\} \]

Introduce a binary variable and some big-Ms ...

\[ C x + D y - b \leq M_p (1 - u) \]
\[ \lambda \leq M_d u \]
\[ u \in \{0, 1\}^\ell \]
Mixed-Integer Linear Reformulation

\[
\begin{align*}
\min_{x,y,\lambda,u} & \quad c^T x + d^T y \\
\text{s.t.} & \quad Ax + By \geq a, \quad Cx + Dy \geq b \\
& \quad \lambda \in \Omega_D := \{ \lambda \geq 0 : D^T \lambda = f \} \\
& \quad Cx + Dy - b \leq M_p (1 - u) \\
& \quad \lambda \leq M_D u \\
& \quad u \in \{0, 1\}^\ell
\end{align*}
\]
Mixed-Integer Linear Reformulation

\[
\begin{align*}
\min_{x,y,\lambda,u} & \quad c^T x + d^T y \\
\text{s.t.} & \quad Ax + By \geq a, \quad Cx + Dy \geq b \\
& \quad \lambda \in \Omega_D := \{\lambda \geq 0 : D^T \lambda = f\} \\
& \quad Cx + Dy - b \leq M_D (1 - u) \\
& \quad \lambda \leq M_D u \\
& \quad u \in \{0,1\}^\ell
\end{align*}
\]

But how to choose the nasty big-\(M\)s?
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:
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.
A Real-World Application: The European Gas Market with Chance Constraints
The European Entry-Exit Gas Market

Level 1  TSO announces technical capacities and booking price floors
Level 2  Traders book, i.e., sign mid- to long-term capacity contracts
Level 3  Traders nominate at a day-ahead market
Level 4  TSO cost-optimally transports the given nominations
The European Entry-Exit Gas Market

Level 1  TSO announces technical capacities and booking price floors
Level 2  Traders book, i.e., sign mid- to long-term capacity contracts
Level 3  Traders nominate at a day-ahead market
Level 4  TSO cost-optimally transports the given nominations

Grimm, Schewe, S., Zöttl (2019)

- Four-level modeling of the European entry-exit gas market
- Identification of assumptions that allow to simplify the model
- Perfect competition → reduction to a bilevel model
The European Entry-Exit Gas Market
Bilevel Modeling Under Perfect Competition: Upper Level

\[
\begin{align*}
\max_{q_{TC}, \bar{\pi}_{book}, \pi, q} & \quad \varphi^u(q_{nom}, q) = \sum_{t \in T} \left( \sum_{i \in P_-} \int_0^{q_{nom}^t} P_{i,t} (s) \, ds - \sum_{i \in P_+} c_{i,\text{var}} q_{nom}^t \right) - \sum_{t \in T} \sum_{a \in A} c_{\text{trans}} (q_{a,t}) \\
\text{s.t.} & \quad 0 \leq q_{TC}^u, 0 \leq \bar{\pi}_{book}^u \quad \text{for all } u \in V_+ \cup V_- \\
& \quad \sum_{u \in V_+ \cup V_-} \sum_{i \in P_u} \pi_{u,book}^t q_{book}^i = \sum_{t \in T} \sum_{a \in A} c_{\text{trans}} (q_{a,t}) \\
& \quad (\pi, q) \in \mathcal{F}(q_{nom}) \\
& \quad (q_{book}^t, q_{nom}^t) \in \arg\max \{ \text{lower-level problem} \}
\end{align*}
\]
Bilevel Modeling Under Perfect Competition: Lower Level

\[
\begin{align*}
\max_{q_{\text{book}}, q_{\text{nom}}} & \quad \sum_{t \in T} \left( \sum_{i \in P_-} \int_0^{q_{\text{nom},i,t}} p_{i,t}(s) \, ds - \sum_{i \in P_+} c_{i}^{\text{var}} q_{i,t}^{\text{nom}} \right) - \sum_{u \in V_+} \sum_{i \in P_u} \pi_{i}^{\text{book}} q_{i}^{\text{book}} \\
\text{s.t.} & \quad \sum_{i \in P_u} q_{i}^{\text{book}} \leq q_{u}^{\text{TC}} \quad \text{for all } u \in V_+ \cup V_- \\
& \quad 0 \leq q_{i,t}^{\text{nom}} \leq q_{i}^{\text{book}} \quad \text{for all } i \in P_- \cup P_+, \ t \in T \\
& \quad \sum_{i \in P_-} q_{i,t}^{\text{nom}} - \sum_{i \in P_+} q_{i,t}^{\text{nom}} = 0 \quad \text{for all } t \in T
\end{align*}
\]
In reality, exit players $i \in \mathcal{P}_-$ nominate quantities $q_{i,t}^{\text{nom}}$ without exactly knowing the actual load $\xi_{i,t}$.

Load vector $\xi = (\xi_{i,t})_{i \in \mathcal{P}_-, t \in \mathcal{T}}$ with log-concave cumulative distribution function.

In particular: $\xi \sim \mathcal{N}(m, \Sigma)$.

Modeling assumption: the TSO imposes a fee $\mu$ on the exit players $i \in \mathcal{P}_-$ to ensure that the realized loads are covered up to a specified safety level $p \in [0, 1]$.

Joint (over all times and exit players) probabilistic constraint

$$\mathbb{P} \left( \xi_{i,t} \leq q_{i,t}^{\text{nom}} \text{ for all } i \in \mathcal{P}_-, t \in \mathcal{T} \right) \geq p$$

Log-concavity of the Gaussian distribution function implies that the log-transformed probabilistic load coverage constraint

$$h(q_{\text{nom}}^-) := \log p - \log \mathbb{P} \left( \xi_{i,t} \leq q_{i,t}^{\text{nom}} \text{ for all } i \in \mathcal{P}_-, t \in \mathcal{T} \right) \leq 0$$

is convex.
\[
\begin{align*}
\min_{x,y} & \quad F(x,y) \\
\text{s.t.} & \quad G(x,y) \leq 0 \\
& \quad x \in \mathbb{R}^{n_x}, \quad y \in \mathbb{R}^{n_y} \\
& \quad y \in S(x)
\end{align*}
\]
Bilevel Optimization

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

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

\[
S(x) = \arg \min_{y} \{ f(x, y) : g(x, y) \leq 0, \ y \in \mathbb{R}^{n_y} \}.
\]
Black-Box Constraint in the Lower Level

A “small” extension

\[ S(x) = \arg \min_y \{ f(x, y) : g(x, y) \leq 0, \ b(y) \leq 0, \ y \in \mathbb{R}^n \} \]
A “small” extension

\[ S(x) = \arg \min_y \{ f(x, y) : g(x, y) \leq 0, \ b(y) \leq 0, \ y \in \mathbb{R}^n \} \]

Assumption

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

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

• Shared constraint set

\[ \Omega := \{(x, y) : G(x, y) \leq 0, \ g(x, y) \leq 0, \ b(y) \leq 0\} \]
Some Notation & Single-Level Reformulation

• Shared constraint set

\[ \Omega := \{ (x, y) : G(x, y) \leq 0, \ g(x, y) \leq 0, \ b(y) \leq 0 \} \]

• Projection onto the decision space of the leader

\[ \Omega_u := \{ x : \exists y \text{ with } (x, y) \in \Omega \} \]
Some Notation & Single-Level Reformulation

• Shared constraint set
  \[ \Omega := \{ (x, y) : G(x, y) \leq 0, \ g(x, y) \leq 0, \ b(y) \leq 0 \} \]

• Projection onto the decision space of the leader
  \[ \Omega_u := \{ x : \exists y \text{ with } (x, y) \in \Omega \} \]

• Optimal value function of the lower level
  \[ \varphi(x) = \min_y \{ f(x, y) : g(x, y), \ b(y) \leq 0, \ y \in \mathbb{R}^{n_y} \} \]
Some Notation & Single-Level Reformulation

• Shared constraint set

\[ \Omega := \{(x, y) : G(x, y) \leq 0, \ g(x, y) \leq 0, \ b(y) \leq 0\} \]

• Projection onto the decision space of the leader

\[ \Omega_u := \{x : \exists y \text{ with } (x, y) \in \Omega\} \]

• Optimal value function of the lower level

\[ \varphi(x) = \min_y \{f(x, y) : g(x, y), \ b(y) \leq 0, \ y \in \mathbb{R}^{ny}\} \]

• Single-level reformulation

\[
\begin{align*}
\min_{x,y} & \quad F(x, y) \\
\text{s.t.} & \quad G(x, y) \leq 0, \ g(x, y) \leq 0, \ b(y) \leq 0 \\
& \quad f(x, y) \leq \varphi(x) \\
& \quad x \in \mathbb{R}^{nx}, \quad y \in \mathbb{R}^{ny}
\end{align*}
\]
- **Main challenge**: black-box constraint $b(y) \leq 0$
- **Not given explicitly** $\Rightarrow$ optimality conditions (KKT) are not given explicitly as well
Obstacles and Pitfalls

- **Main challenge**: black-box constraint $b(y) \leq 0$
- **Not given explicitly** → optimality conditions (KKT) are not given explicitly as well
- **Possible remedies**
  - Cutting plane techniques (Kelley 1960)
  - Outer approximation (Duran, Grossmann 1986; Fletcher, Leyffer 1994)
Obstacles and Pitfalls

- **Main challenge**: black-box constraint \( b(y) \leq 0 \)
- **Not given explicitly** → optimality conditions (KKT) are not given explicitly as well
- **Possible remedies**
  - Cutting plane techniques (Kelley 1960)
  - Outer approximation (Duran, Grossmann 1986; Fletcher, Leyffer 1994)
- **But**: \( b(y) \leq 0 \) can only by satisfied up to a prescribed tolerance
- **Specifying the quality of solutions via \( \varepsilon-\delta \)-optimality**
  - Global optimization (Locatelli, Schoen 2013)
  - Bilevel optimization (Mitsos, Lemonidis, Barton 2008)
Definition
For $\delta = (\delta_G, \delta_g, \delta_b, \delta_f) \in \mathbb{R}^{m_u + m_\ell + 2}_{\geq 0}$, a point $\overline{x}, \overline{y} \in \mathbb{R}^{n_x} \times \mathbb{R}^{n_y}$ is called $\delta$-feasible for the bilevel problem if $G(\overline{x}, \overline{y}) \leq \delta_G$, $g(\overline{x}, \overline{y}) \leq \delta_g$, $b(y) \leq \delta_b$, and $f(x, y) \leq \varphi(x) + \delta_f$ hold. Moreover, for $\varepsilon \geq 0$, a point $(x^*, y^*) \in \mathbb{R}^{n_x} \times \mathbb{R}^{n_y}$ is called $\varepsilon$-$\delta$-optimal for the bilevel problem, if it is $\delta$-feasible and if $F(x^*, y^*) \leq F^* + \varepsilon$ holds, with $F^*$ denoting the optimal objective function value of the bilevel problem.
**Definition**

For \( \delta = (\delta_G, \delta_g, \delta_b, \delta_f) \in \mathbb{R}_{\geq 0}^{m_u + m \ell + 2} \), a point \((\bar{x}, \bar{y}) \in \mathbb{R}^{n_x} \times \mathbb{R}^{n_y}\) is called \(\delta\)-feasible for the bilevel problem if \(G(\bar{x}, \bar{y}) \leq \delta_G, g(\bar{x}, \bar{y}) \leq \delta_g, b(y) \leq \delta_b\), and \(f(x, y) \leq \varphi(x) + \delta_f\) hold. Moreover, for \(\varepsilon \geq 0\), a point \((x^*, y^*) \in \mathbb{R}^{n_x} \times \mathbb{R}^{n_y}\) is called \(\varepsilon\)-\(\delta\)-optimal for the bilevel problem, if it is \(\delta\)-feasible and if \(F(x^*, y^*) \leq F^* + \varepsilon\) holds, with \(F^*\) denoting the optimal objective function value of the bilevel problem.

- A \(\delta\)-feasible point \((\bar{x}, \bar{y})\) is \(\delta_f\)-(\(\delta_g, \delta_b\))-optimal for the lower level with fixed \(x = \bar{x}\)
- Assume \(f\) and \(g\) pose no challenges \(\rightarrow\) choose \(\delta_f = \delta_g = 0\)
- Assume \(F\) and \(G\) pose no challenges \(\rightarrow\) can we obtain 0-\(\delta\)-optimal solutions with \(\delta = (0, 0, \delta_b, 0)\)?
0-δ-optimal solutions with \( \delta = (0, 0, \delta_b, 0) \)?
0-\(\delta\)-optimal solutions with \(\delta = (0, 0, \delta_b, 0)\)?

- Consider the relaxed lower-level problem

\[
\min_{y \in \mathbb{R}^n_y} f(\bar{x}, y) \quad \text{s.t.} \quad g(\bar{x}, y) \leq 0, \quad b(y) \leq \delta_b
\]

- Denote the optimal value function by \(\varphi(x)\)

- Relaxation property yields \(\underline{\varphi}(x) \leq \varphi(x)\) for all feasible \(x \in \Omega_u\)
Consider the relaxed lower-level problem

\[
\min_{y \in \mathbb{R}^n} \quad f(\bar{x}, y) \quad \text{s.t.} \quad g(\bar{x}, y) \leq 0, \quad b(y) \leq \delta_b
\]

Denote the optimal value function by \( \bar{\varphi}(x) \)

Relaxation property yields \( \bar{\varphi}(x) \leq \varphi(x) \) for all feasible \( x \in \Omega_u \)

First-relax-then-reformulate leads to a single-level problem with \( f(x, y) \leq \bar{\varphi}(x) \)

If \( \bar{\varphi}(x) < \varphi(x) \) holds for any \( x \in \Omega_u \), this single-level reformulation is not a relaxation of the original single-level reformulation
• Consider the relaxed lower-level problem

$$\min_{y \in \mathbb{R}^{ny}} f(\bar{x}, y) \quad \text{s.t.} \quad g(\bar{x}, y) \leq 0, \ b(y) \leq \delta_b$$

• Denote the optimal value function by $\bar{\varphi}(x)$
• Relaxation property yields $\bar{\varphi}(x) \leq \varphi(x)$ for all feasible $x \in \Omega_u$
• First-relax-then-reformulate leads to a single-level problem with $f(x, y) \leq \bar{\varphi}(x)$
• If $\bar{\varphi}(x) < \varphi(x)$ holds for any $x \in \Omega_u$, this single-level reformulation is not a relaxation of the original single-level reformulation
• It is not clear whether and how $\varepsilon$-$\delta$-optimality can be guaranteed
0-δ-optimal solutions with \( \delta = (0, 0, \delta_b, 0) \)?

- Consider the relaxed lower-level problem

\[
\min_{y \in \mathbb{R}^n} f(\bar{x}, y) \quad \text{s.t.} \quad g(\bar{x}, y) \leq 0, \ b(y) \leq \delta_b
\]

- Denote the optimal value function by \( \bar{\varphi}(x) \)

- Relaxation property yields \( \bar{\varphi}(x) \leq \varphi(x) \) for all feasible \( x \in \Omega_u \)

- First-relax-then-reformulate leads to a single-level problem with \( f(x, y) \leq \bar{\varphi}(x) \)

- If \( \bar{\varphi}(x) < \varphi(x) \) holds for any \( x \in \Omega_u \), this single-level reformulation is not a relaxation of the original single-level reformulation

- It is not clear whether and how \( \varepsilon-\delta \)-optimality can be guaranteed

Can we hope for the \( \delta \)-feasible points with \( \delta = (0, 0, \delta_b, 0) \)?
A “First-Relax-Then-Reformulate” Approach

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

$$\{y \in \mathbb{R}^{ny} : b(y) \leq 0\} \subseteq \{y \in \mathbb{R}^{ny} : E^{r+1}y \leq e^{r+1}\} \subseteq \{y \in \mathbb{R}^{ny} : E^ry \leq e^r\}$$
A “First-Relax-Then-Reformulate” Approach

- Block-box constraint $b(y) \geq 0$ is convex
- Construct a sequence of linear outer approximations $(E^r, e^r)_{r \in \mathbb{N}}$ of the black-box constraint $b(y) \leq 0$ with the property
  \[
  \{ y \in \mathbb{R}^{ny} : b(y) \leq 0 \} \subseteq \{ y \in \mathbb{R}^{ny} : E^r y \leq e^r \} \subseteq \{ y \in \mathbb{R}^{ny} : E^r y \leq e^r \}
  \]
- For a given upper-level solution $\bar{x} \in \Omega_u$ and $r \in \mathbb{N}$, the adapted lower-level problem reads
  \[
  \min_{y \in \mathbb{R}^{ny}} f(\bar{x}, y) \quad \text{s.t.} \quad g(\bar{x}, y) \leq 0, \ E^r y \leq e^r
  \]
- This is a relaxation of the original lower-level problem
A “First-Relax-Then-Reformulate” Approach

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

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

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

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

• This is a relaxation of the original lower-level problem
• $\varphi^r(x)$: optimal value function
• Assumption: Slater’s constraint qualification holds
A “First-Relax-Then-Reformulate” Approach

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

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

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

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

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

**Proposition**

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

$$\varphi^r(x) \leq \varphi^{r+1}(x) \leq \varphi(x)$$
A “First-Relax-Then-Reformulate” Approach

Modified variant of the single-level reformulation

\[
\begin{align*}
\min_{x,y} & \quad F(x, y) \\
\text{s.t.} & \quad G(x, y) \leq 0, \quad g(x, y) \leq 0 \\
& \quad E^r y \leq e^r \\
& \quad f(x, y) \leq \varphi^r(x) \\
& \quad x \in \mathbb{R}^{n_x}, \quad y \in \mathbb{R}^{n_y}
\end{align*}
\]
A “First-Relax-Then-Reformulate” Approach

Modified variant of the single-level reformulation

\[
\begin{align*}
\min_{x,y} & \quad F(x, y) \\
\text{s.t.} & \quad G(x, y) \leq 0, \quad g(x, y) \leq 0 \\
& \quad E^r y \leq e^r \\
& \quad f(x, y) \leq \varphi^r(x) \\
& \quad x \in \mathbb{R}^{n_x}, \quad y \in \mathbb{R}^{n_y}
\end{align*}
\]

Feasibility problem

\[
\begin{align*}
\min_{x,y,s} & \quad s \\
\text{s.t.} & \quad G(x, y) \leq 0, \quad g(x, y) \leq 0 \\
& \quad E^r y \leq e^r \\
& \quad f(x, y) \leq \varphi^r(x) + s \\
& \quad x \in \mathbb{R}^{n_x}, \quad y \in \mathbb{R}^{n_y}
\end{align*}
\]
Algorithm “First-Relax-Then-Reformulate”.

1: Choose $\delta_b > 0$, set $r = 0$, $s = 0$, $\chi = \infty$, $E^0 = [0 \ldots 0] \in \mathbb{R}^{1 \times n_y}$, $e^0 = 0 \in \mathbb{R}$.

2: while $\chi > \delta_b$ or $s > 0$ do

3: Construct $E^{r+1}$ and $e^{r+1}$

4: if the modified variant of the single-level reformulation is feasible then

5: Solve this problem to obtain $(x^{r+1}, y^{r+1})$ and set $s = 0$.

6: else if the feasibility problem is feasible then

7: Solve this problem to obtain $(x^{r+1}, y^{r+1}, s)$.

8: else

9: Return “The original problem is infeasible.”

10: end if

11: Set $r \leftarrow r + 1$ and $\chi = b(y^r)$.

12: end while

13: Return $(\bar{x}, \bar{y}) = (x^r, y^r)$.

Theorem: If the algorithm terminates, then $(\bar{x}, \bar{y})$ is $(0, 0, \delta_b, 0)$-feasible for original bilevel problem.
Algorithm “First-Relax-Then-Reformulate”.

1: Choose $\delta_b > 0$, set $r = 0$, $s = 0$, $\chi = \infty$, $E^0 = [0 \ldots 0] \in \mathbb{R}^{1 \times ny}$, $e^0 = 0 \in \mathbb{R}$.
2: while $\chi > \delta_b$ or $s > 0$ do
3: Construct $E^{r+1}$ and $e^{r+1}$
4: if the modified variant of the single-level reformulation is feasible then
5: Solve this problem to obtain $(x^{r+1}, y^{r+1})$ and set $s = 0$.
6: else if the feasibility problem is feasible then
7: Solve this problem to obtain $(x^{r+1}, y^{r+1}, s)$.
8: else
9: Return “The original problem is infeasible.”
10: end if
11: Set $r \leftarrow r + 1$ and $\chi = b(y^r)$.
12: end while
13: Return $(\bar{x}, \bar{y}) = (x^r, y^r)$.

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

Node 1

Node 2

Node 3

Node 4

Node 5

Entry 1

Entry 2

Entry 3

Exit 1

Exit 2

Exit 3

$\Delta L_a = 150$

$\Delta L_a = 180$

$\Delta L_a = 90$

$\Delta L_a = 190$

$\Delta L_a = 130$

$\Delta L_a = 110$

$\Delta L_a = 80$

$\Delta L_a = 210$

var$_1 = 274.8$

var$_2 = 270.4$

var$_3 = 250.3$

var$_4 = -16.60$

var$_5 = -13.80$

var$_6 = -20.70$

Pipe 1

Pipe 2

Pipe 3

Pipe 4

Pipe 5

Pipe 6

Pipe 7

Pipe 8

Pipe 9

Pipe 10

Pipe 11
## Numerical Results

<table>
<thead>
<tr>
<th>( p )</th>
<th>Bisection Runtime</th>
<th>Bisection #Iter.</th>
<th>Bounding Runtime</th>
<th>Bounding #Iter.</th>
<th>( \delta )-Feasibility #Iter.</th>
<th>( \delta )-Feasibility Runtime</th>
<th>Total Runtime</th>
<th>Gap</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.60</td>
<td>12.13</td>
<td>32</td>
<td>36.80</td>
<td>10</td>
<td>28.97</td>
<td>42</td>
<td>77.9</td>
<td>0.001</td>
</tr>
<tr>
<td>0.65</td>
<td>14.15</td>
<td>28</td>
<td>32.00</td>
<td>16</td>
<td>40.71</td>
<td>44</td>
<td>86.86</td>
<td>0.001</td>
</tr>
<tr>
<td>0.70</td>
<td>11.13</td>
<td>26</td>
<td>29.70</td>
<td>13</td>
<td>39.70</td>
<td>39</td>
<td>80.53</td>
<td>0.001</td>
</tr>
<tr>
<td>0.75</td>
<td>9.04</td>
<td>25</td>
<td>28.55</td>
<td>6</td>
<td>14.19</td>
<td>31</td>
<td>51.78</td>
<td>0.002</td>
</tr>
<tr>
<td>0.80</td>
<td>7.98</td>
<td>25</td>
<td>29.06</td>
<td>4</td>
<td>6.26</td>
<td>29</td>
<td>43.3</td>
<td>0.005</td>
</tr>
<tr>
<td>0.85</td>
<td>11.08</td>
<td>21</td>
<td>24.01</td>
<td>3</td>
<td>7.41</td>
<td>24</td>
<td>42.5</td>
<td>0.006</td>
</tr>
<tr>
<td>0.90</td>
<td>11.05</td>
<td>23</td>
<td>26.34</td>
<td>8</td>
<td>27.52</td>
<td>31</td>
<td>64.91</td>
<td>0.017</td>
</tr>
<tr>
<td>0.95</td>
<td>5.96</td>
<td>24</td>
<td>27.99</td>
<td>6</td>
<td>14.14</td>
<td>30</td>
<td>48.09</td>
<td>0.010</td>
</tr>
<tr>
<td>0.96</td>
<td>7.56</td>
<td>22</td>
<td>24.56</td>
<td>3</td>
<td>4.17</td>
<td>25</td>
<td>36.29</td>
<td>0.011</td>
</tr>
<tr>
<td>0.97</td>
<td>6.94</td>
<td>21</td>
<td>23.96</td>
<td>4</td>
<td>9.20</td>
<td>25</td>
<td>40.10</td>
<td>0.015</td>
</tr>
<tr>
<td>0.98</td>
<td>4.63</td>
<td>25</td>
<td>93.68</td>
<td>9</td>
<td>106.31</td>
<td>34</td>
<td>204.62</td>
<td>0.032</td>
</tr>
<tr>
<td>0.99</td>
<td>6.96</td>
<td>26</td>
<td>29.76</td>
<td>10</td>
<td>1250.65</td>
<td>36</td>
<td>1287.37</td>
<td>0.187</td>
</tr>
</tbody>
</table>
Total Welfare and Price of Load Coverage

![Graph of total welfare and price of load coverage](image-url)
On convex lower-level black-box constraints in bilevel optimization with an application to gas market models with chance constraints

Holger Heitsch, René Henrion, Thomas Kleinert & Martin Schmidt

Journal of Global Optimization (2022) | Cite this article

Abstract

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

Upper-level problem

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

Lower-level problem

\[
\min_{y} f(x, y) \\
\text{s.t. } g(x, y) \geq 0
\]
Who can solve this problem?

Upper level

\[
\max_{x \in \mathbb{R}^2} \quad F(x, y) = x_1 - 2y_{n+1} + y_{n+2}
\]

s.t. \( (x_1, x_2) \in [x_1, \bar{x}_1] \times [x_2, \bar{x}_2] \)

\( y \in S(x) \)

\[
\bullet \quad x, \bar{x} \in \mathbb{R}^2 \text{ with } 1 \leq x_i < \bar{x}_i, \ i \in \{1, 2\}
\]

\bullet Upper level is an LP with simple bound constraints

\bullet Upper level has no coupling constraints
Who can solve this problem?

Upper level

\[ \max_{x \in \mathbb{R}^2} F(x, y) = x_1 - 2y_{n+1} + y_{n+2} \]

s.t. \((x_1, x_2) \in [x_1, \bar{x}_1] \times [x_2, \bar{x}_2]\)

\[ y \in S(x) \]

Lower level

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

s.t. \(y_1 + y_n = \frac{1}{2}\)

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

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

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

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

- \(x, \bar{x} \in \mathbb{R}^2\) with \(1 \leq x_i < \bar{x}_i, i \in \{1, 2\}\)
- 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

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

Result #1

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

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

Result #1
For every feasible leader’s decision \((x_1, x_2) \in [x_1, \bar{x}_1] \times [x_2, \bar{x}_2]\), a feasible follower’s decision \(y\) satisfies \(y_n > 0\).

Result #2
For every feasible leader’s decision \((x_1, x_2) \in [x_1, \bar{x}_1] \times [x_2, \bar{x}_2]\), the set of optimal solutions of the lower-level problem is a singleton.
Exact feasibility

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

**Result #1**

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

**Result #2**

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

**Result #3**

The bilevel problem has a **unique** solution given by \(x^* = (x_1, \bar{x}_2)\) with an optimal objective function value of \(F^* = x_1 + \bar{x}_2\).
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 $\varepsilon$-feasible for the problem $\max_{x \in \mathbb{R}^n} \{f(x) : g_i(x) \leq 0\}$ 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 $\varepsilon$-feasible for the problem $\max_{x \in \mathbb{R}^n} \{ f(x) : g(x) \leq 0 \}$ 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 $\varepsilon$-feasible follower’s decision $y$ with $y_n = 0$ for every feasible leader’s decision $(x_1, x_2) \in [x_1, \bar{x}_1] \times [x_2, \bar{x}_2]$. 
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 $\varepsilon$-feasible for the problem $\max_{x \in \mathbb{R}^n} \{f(x) : g_i(x) \leq 0 \}$ 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 $\varepsilon$-feasible follower’s decision $y$ with $y_n = 0$ for every feasible leader’s decision $(x_1, x_2) \in [x_1, \bar{x}_1] \times [x_2, \bar{x}_2]$.

Result #5
Unless $\varepsilon < 2^{-2^{n-1}}$, the set of $\varepsilon$-feasible follower’s solutions is not a singleton for every feasible leader’s decision $(x_1, x_2) \in [x_1, \bar{x}_1] \times [x_2, \bar{x}_2]$. 
Result #3 (revisited)

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

**Result #3 (revisited)**

The bilevel problem has a unique solution given by \(x^* = (x_1, \bar{x}_2)\) with an optimal objective function value of \(F^* = x_1 + \bar{x}_2\).

**Result #6**

Let \(\varepsilon \geq 2^{-2^{n-1}}\) and suppose that we allow for \(\varepsilon\)-feasible follower’s solutions.

Then, the **optimistic optimal solution** of the bilevel problem is given by \(x_0^* = (\bar{x}_1, \bar{x}_2)\) with an optimal objective function value of \(F_0^* = \bar{x}_1 + \bar{x}_2\).

The **pessimistic optimal solution** is given by \(x_p^* = (x_1, x_2)\) with an optimal objective function value of \(F_p^* = -x_1 - x_2\).
\(\varepsilon\)-feasibility

Result #3 (revisited)
The bilevel problem has a unique solution given by \(x^* = (x_1, \bar{x}_2)\) with an optimal objective function value of \(F^* = x_1 + \bar{x}_2\).

Result #6
Let \(\varepsilon \geq 2^{-2^{n-1}}\) and suppose that we allow for \(\varepsilon\)-feasible follower’s solutions.

Then, the optimistic optimal solution of the bilevel problem is given by \(x_0^* = (\bar{x}_1, \bar{x}_2)\) with an optimal objective function value of \(F_0^* = \bar{x}_1 + \bar{x}_2\).

The pessimistic optimal solution is given by \(x_p^* = (x_1, x_2)\) with an optimal objective function value of \(F_p^* = -x_1 - x_2\).

- 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?
Some Related Literature

• A Survey on Mixed-Integer Programming Techniques in Bilevel Optimization
  Thomas Kleinert, Martine Labbé, Ivana Ljubic, Martin Schmidt
  EURO Journal on Computational Optimization

• On Convex Lower-Level Black-Box Constraints in Bilevel Optimization with an Application to Gas Market Models with Chance Constraints
  Holger Heitsch, René Henrion, Thomas Kleinert, Martin Schmidt
  Journal of Global Optimization

• On a Computationally Ill-Behaved Bilevel Problem with a Continuous and Nonconvex Lower Level
  Yasmine Beck, Daniel Bienstock, Johannes Thürauf, Martin Schmidt