Capacity planning with competitive decision-makers: Trilevel MILP formulation and solution approaches

Carlos Florensa Campo[a], Pablo Garcia-Herreros[a], Pratik Misra[b], Erdem Arslan[b], Sanjay Mehta[b], Ignacio E. Grossmann[a,*]

[a] *Department of Chemical Engineering, Carnegie Mellon University, Pittsburgh, USA*
[b] *Air Products and Chemicals, Inc. Allentown, USA*

Abstract

Capacity planning addresses the decision problem of an industrial producer investing on infrastructure to satisfy future demand at the highest profit. Traditional formulations neglect the rational behavior of external decision-makers by assuming static competition and captive markets. We propose a mathematical programming formulation with three levels of decision-makers to fully capture the dynamic of competitive markets. The trilevel model is transformed into a bilevel optimization problem with mixed-integer variables in both levels by replacing the third-level linear program with its optimality conditions. We introduce new definitions required for the analysis of degeneracy in multilevel problems, and develop two novel algorithms to solve these challenging problems. Each algorithm is shown to converge to a different type of degenerate solution. The computational experiments for capacity expansion in industrial gas markets show that no algorithm is strictly better in terms of performance.

Keywords: Multilevel programming, degeneracy, capacity expansion, competitive markets.

1. Introduction

Industrial gas companies rely on capacity expansion models to plan the investments that allow them to satisfy future demands. In this industry, the proximity of producers to customers is a key competitive advantage that increases supply reliability and reduces transportation costs. This feature makes capacity planning a major strategic decision that impacts the market share that can be obtained in an environment with rational customers.

1.1. Capacity planning and bilevel programming

Capacity planning is a widely studied problem in areas requiring the development of long-term infrastructure, like in electrical power supply [20], and communication networks [5]. In the process systems engineering community, the capacity planning problem has been extended to consider aspects of process design [22] and product development [16]. The formulation can be applied

*Corresponding author
Email address: grossmann@cmu.edu (Ignacio E. Grossmann)

to problems with large capital investments whose feasibility, effectiveness, and profitability can only be assessed in a long time horizon. Therefore, capacity planning is critical for many industries [15] and it is recognized as a key topic in Enterprise-Wide Optimization [12].

Nevertheless, the capacity planning problem in a fully competitive environment has not been formulated before in a mathematical programming framework. Previous works have focused on different market models, assuming static competitors. Murphy and Smeers [20] proposed models with oligopolistic and perfect markets for the expansion of electric infrastructure. Recently, Garcia-Herreros et al. [10] presented a more realistic model proposing a bilevel formulation. It models the expansion decisions of a company maximizing its profit, subject to markets selecting their suppliers to satisfy their demands at minimum cost.

A bilevel optimization problem is a mathematical program with a second mathematical program in its constraints. It can be interpreted as a game with an upper-level player, called the *leader*, that first decides its strategy with perfect information of the criterion ruling the behavior of the lower-level player, called *follower*. Once the follower observes the leader's decision, it reacts according to its own interests. The potential to coordinate decision-making in decentralized systems has already been recognized [1]. Interesting multilevel programming models have been developed for traffic planning [17], optimal taxation of biofuels [3], parameter estimation [18], and product introduction [21].

However, a major extension of the bilevel formulations is needed to model a fully competitive environment. Three nested optimization problems are required instead of two because decisions are made sequentially by three players with conflicting interests. The company planning its capacity is the first to decide on its expansion strategy. Then, the competition observe the leader's plan and optimize their own capacity expansions. After all expansions are fixed, the market selects providers to minimize the cost of satisfying its demands. The logical decisions of the first two levels require the use of integer variables, whereas the third level can be formulated as a linear program (LP).

There has been little work on multilevel optimization involving more that two players with discrete variables. The electrical network defense is the only problem for which a trilevel mixed-integer linear programming (TMILP) model has already been proposed [24]. However, the solution procedures for this formulation are problem specific and there is scarce theoretical study of the general properties of trilevel optimization problems. Our research presents a novel framework for capacity planning, discusses new ideas about how degeneracy affects multilevel problems, and proposes two solution algorithms.

1.2. Solution approaches for multilevel programming

The first step for both algorithms is to reformulate the trilevel problem as a bilevel problem, replacing the third-level by its optimality conditions. The reformulation is based on strong duality of the lower-level LP, which offers documented advantages over the standard KKT reformulation because it avoids the addition of discrete variables [10]. In the bilevel reformulation, the second level models the capacity expansion of the competitors and enforces optimality of the third-level problem. The resulting formulation is a Bilevel Mixed-Integer Linear Program (BMILP) with discrete variables in both levels; these type of problems is still considered an open question in Operations Research [6].

The numerical solution of BMILPs has been receiving increasing attention during the past years, but the existing literature has only considered academic examples with a few discrete variables. The first Branch-&-Bound algorithm was developed by Moore and Bard [19]; it was based exclusively on the solution of LPs. Later, the same authors proposed a binary search tree algorithm that obtains the rational reaction of the lower level by solving a MILP after fixing the decision of the leader [2]; in the worst case, both algorithms conduct an exhaustive exploration of the leader's decision space. DeNegre and Ralphs [8] derived a locally valid cut that can be added to the Branch-&-Bound procedure proposed by Bard and Moore [2]; however, these cuts tend to be weak in problems with parameters of different magnitudes or with non-integer coefficients.

The framework proposed by Gümüş and Floudas [13] is based on replacing the lower-level MILP by the equivalent LP over the convex hull of the feasible region. This strategy allows using the reformulation techniques developed for LPs, but it comes at the expense of introducing an exponential number of new variables and constraints. Faísca et al. [9] have used multi-parametric programming to obtain a function that characterizes the optimal lower-level response for any potential decision of the leader. This procedure can be extremely involved, but is interesting from a theoretical point of view because the multi-parametric solution explicitly describes the feasible region of the bilevel problem.

Recently, there have been two relevant contributions for our research. Xu and Wang [23] proposed a general spatial Branch-&-Bound search that splits the variables of the leader in polyhedral sets called stability regions; stability regions characterize the decisions of the leader that share the same optimal reaction of the follower. Zeng and An [25] developed a reformulation-decomposition approach that iteratively approximates the rational reaction of the follower based on linear inequalities in the space of the leader decision variables. Both contributions have been important for the development of our algorithms.

We present two algorithms that leverage and expand the relaxation obtained by eliminating the objective function of the lower level, known as *high-point* (HP) problem. The first algorithm is a constraint-directed exploration; it eliminates decisions of the leader that have been explored, as well as all other decisions that induce the same reaction of the other players. The second algorithm is a decomposition solution strategy involving a master problem and a subproblem. The main idea is to incorporate in the master problem the reactions of the competitors that are iteratively observed; this procedure shows an interesting speed-up in instances with few rational alternatives for the competition.

The rest of the paper is structured as follows. In Section 2, we describe the capacity planning problem in a competitive environment. In Section 3, we propose the mathematical formulation for the capacity planning in a competitive environment. Section 4 explores the implications of degeneracy in trilevel optimization problems. In Section 5, we elaborate on the properties of the capacity planning model that are useful for the development of two novel algorithms. The algorithms are described in Sections 6 and 7. In Section 8, we illustrate the implementation of the algorithms on two instances of the capacity planning problem. Finally, Section 9 reviews the novelty of this work and indicates directions for future research.

2. Problem statement

The capacity planning problem in a competitive environment considers three players with independent decision criteria: the first-level industrial producer (leader) planning its expansion strategy, competitors that are allowed to expand, and costumers that select their providers according to the available supply. The objective of the capacity planning problem is to establish the expansion plan that maximizes the Net Present Value (NPV) obtained by the leader during a finite time horizon. Sales are bounded by deterministic demands during the entire time horizon and depend on the actions taken by other decision-makers. Initially, the leader is given a set of plants with finite production capacity and a set of candidate locations where new plants can be built. Capacity of the plants can be expanded in discrete increments by paying a fixed cost; only discrete capacity expansions are considered to model the installation of new production lines. The attractiveness of expanding plants depends on the possibility of satisfying a higher demand or saving in production and transportation costs.

The competitors are other industrial producers that observe the decisions of the leader and select expansion plans with the objective of maximizing their own NPV. All production plants that are not controlled by the leader are assumed to behave as a rational decision-maker with a centralized planner. The competition controls a set of open plants with given initial capacity; they are also allowed to open new plants in a set of candidate locations. Expansions at open plants are modeled with discrete increments in capacity and fixed costs. We assume that the total initial capacity in the plants controlled by the leader and the competitors is enough to satisfy all demands throughout the horizon, which is modeled by including a plant that has a large capacity an offers products at a high cost.

Demand assignments in each time period are decided after observing the available capacities in the plants controlled by the leader and the competition. The market behaves as a centralized decision-maker that minimizes the total cost paid by all customers; demand assignments are based exclusively on availability and price. Industrial producers can only influence market decisions by changing their production capacity since prices are fixed parameters. Furthermore, the prices are constructed following two common assumptions in an industrial environment, detailed in the following Eqns. (1)-(2).

- *The leader offers homogeneous prices:* the price $P_{t,i,j}$ offered to a certain customer (j) is the same regardless of the plant $i \in I^{\mathcal{L}}$.

$$P_{t,i,j} = P_{t,j} \qquad \forall\ t \in T,\ i \in I^{\mathcal{L}},\ j \in J \tag{1}$$

- *Competitors offer site-dependent prices:* the price $P_{t,i,j}$ offered to customers depend on a raw price ($P_{t,i}^{raw}$) and the transportation cost ($G_{t,i,j}$) from that plant to the customers (j) .

$$P_{t,i,j} = P_{t,i}^{raw} + G_{t,i,j} \qquad \forall\ t \in T,\ i \in I^{\mathcal{C}},\ j \in J \tag{2}$$

The decision process takes place sequentially, and perfect information is assumed for all players. The perfect information assumption implies that higher level decision-makers are aware of the decision criteria of the lower levels, and lower-level decision-makers observe the actions of the higher levels before choosing their response. The optimal solution of the problem characterizes the expansion plan that optimizes the objective function of the leader, considering a rational reaction of the competitors and the market.

3. Capacity expansion with competitive decision-makers: trilevel formulation

According to the problem statement, we define the objective function of the industrial producers as the maximization of their NPV over a finite time horizon. The objective function presented in Eqn. (3) is the decision criterion of the leader,

$$\begin{aligned}\mathrm{NPV}^{\mathcal{L}} =& \sum_{t\in T}\sum_{i\in I^{\mathcal{L}}}\sum_{j\in J} P_{t,i,j}y_{t,i,j} \\ &- \sum_{t\in T}\sum_{i\in I^{\mathcal{L}}} \left(A_{t,i}v_{t,i} + B_{t,i}w_{t,i} + E_{t,i}x_{t,i}\right) \\ &- \sum_{t\in T}\sum_{i\in I^{\mathcal{L}}}\sum_{j\in J} \left(F_{t,i} + G_{t,i,j}\right) y_{t,i,j} \end{aligned} \tag{3}$$

where T, $I^{\mathcal{L}}$, and J are respectively the index sets for time periods (t), plants controlled by the leader ($i \in I^{\mathcal{L}}$), and customers (j). The first term represents the income obtained from sales. Variables $y_{t,i,j}$ indicate the quantities sold from plant i to customer j at time t; coefficients $P_{t,i,j}$ are the selling prices. In the second term, $v_{t,i}$ is a binary variable indicating if a new plant is built at location i during time t; its fixed cost is given by coefficients $A_{t,i}$. Maintenance costs are modeled with binary variables $w_{t,i}$ that indicate which plants are open; the maintenance cost per time period is given by $B_{t,i}$. Capacity expansion decisions are modeled with binary variables $x_{t,i}$; the fixed cost of the expansions is given by $E_{t,i}$. The third term models the operating costs by associating sales ($y_{t,i,j}$) with the unit costs of production ($F_{t,i}$) and transportation ($G_{t,i,j}$).

It is worth noticing that the objective presented in Eqn. (3) is not only a function of the decision variables representing the planning decisions; it also depends on demand assignment variables ($y_{t,i,j}$) that are controlled by the customers. The competing industrial producers are assumed to maximize their own NPV with the same income and cost structure of the leader. In this framework, the trilevel formulation can be modeled with Eqns. (4)-(16).

$$\max_{v^{\mathcal{L}},w^{\mathcal{L}},x^{\mathcal{L}},c^{\mathcal{L}}} NPV^{\mathcal{L}}\left(v^{\mathcal{L}}, w^{\mathcal{L}}, x^{\mathcal{L}}, c^{\mathcal{L}}, y\right) \tag{4}$$

$$\text{s.t.} \quad w_{t,i} = V_{0,i} + \sum_{t'\in T_t^-} v_{t',i} \qquad \forall\ t\in T, i\in I^{\mathcal{L}} \tag{5}$$

$$x_{t,i} \le w_{t,i} \qquad \forall\ t\in T, i\in I^{\mathcal{L}} \tag{6}$$

$$c_{t,i} = C_{0,i} + \sum_{t'\in T_t^-} H_i x_{t',i} \qquad \forall\ t\in T, i\in I^{\mathcal{L}} \tag{7}$$

$$\max_{v^{\mathcal{C}},w^{\mathcal{C}},x^{\mathcal{C}},c^{\mathcal{C}}} NPV^{\mathcal{C}}\left(v^{\mathcal{C}}, w^{\mathcal{C}}, x^{\mathcal{C}}, c^{\mathcal{C}}, y\right) \tag{8}$$

$$
\begin{aligned}
\text{s.t.} \quad & w_{t,i} = V_{0,i} + \sum_{t' \in T_t^-} v_{t',i} && \forall \;\; t \in T, i \in I^{\mathcal{C}} && (9)\\
& x_{t,i} \le w_{t,i} && \forall \;\; t \in T, i \in I^{\mathcal{C}} && (10)\\
& c_{t,i} = C_{0,i} + \sum_{t' \in T_t^-} H_i x_{t',i} && \forall \;\; t \in T, i \in I^{\mathcal{C}} && (11)\\
& y = \underset{y \in Y(c^{\mathcal{L}}, c^{\mathcal{C}})}{\arg\min} \left\{ \sum_{t \in T} \sum_{i \in I} \sum_{j \in J} P_{t,i,j} y_{t,i,j} \right\} && && (12)\\
& c_{t,i} \in \mathbb{R}^+ && \forall \;\; t \in T, i \in I^{\mathcal{C}}, j \in J && (13)\\
& v_{t,i}, \; w_{t,i}, \; x_{t,i} \in \{0,1\} && \forall \;\; t \in T, i \in I^{\mathcal{C}} && (14)\\
c_{t,i} \in \mathbb{R}^+ & && \forall \;\; t \in T, i \in I^{\mathcal{L}}, j \in J && (15)\\
v_{t,i}, \; w_{t,i}, \; x_{t,i} \in \{0,1\} & && \forall \;\; t \in T, i \in I^{\mathcal{L}} && (16)
\end{aligned}
$$

where the set of production plants I is divided in two subsets denoting the plants controlled by the leader ($I^{\mathcal{L}}$) and the plants controlled by the competitors ($I^{\mathcal{C}}$). The superscript $\mathcal{L}$ identifies the variables controlled by the leader and the superscript $\mathcal{C}$ the plants controlled by the competitors. The constraints modeling the feasible investment strategies for the leader are presented in Eqns. (5)-(7). We define T_t^- as the subset of time periods before time t; formally, $T_t^- = \{t' : t' \in T, t' \le t\}$. Eqn. (5) enforces maintenance for open plants; the parameter $V_{0,i}$ indicates if plant i is initially open. Eqn. (6) restricts expansions to the open plants; only one expansion per time period is allowed in each plant. Eqn. (7) models capacity of plants ($c_{t,i}$) according to their initial capacity ($C_{0,i}$) and discrete increments of size H_i. The corresponding feasible expansion plans for the competitors are presented in Eqns. (9)-(11). Domains for the decisions of the competitors and the leader are expressed by Eqns. (13)-(14) and Eqns. (15)-(16), respectively.

The rational response of the customers is modeled with Eqn. (12). The market minimizes its total discounted cost by controlling the demand assignment variables $y_{t,i,j}$ on the polyhedral set $Y(c^{\mathcal{L}}, c^{\mathcal{C}})$, defined by the followin feasibility Eqns. (18)-(20); the decision space of the assignments depends on the capacity planning strategies chosen by the leader and the competitors. The complete third-level optimization problem that decides demand assignments is presented in Eqns. (17)-(20),

$$
\begin{aligned}
\min_{y} \quad & \sum_{t \in T} \sum_{i \in I} \sum_{j \in J} P_{t,i,j} y_{t,i,j} && && (17)\\
\text{s.t.} \quad & \sum_{j \in J} y_{t,i,j} \le c_{t,i} && \forall \; t \in T, i \in I && (18)\\
& \sum_{i \in I} y_{t,i,j} = D_{t,j} && \forall \; t \in T, j \in J && (19)\\
& y_{t,i,j} \ge 0 && \forall \; t \in T, j \in J, i \in I && (20)
\end{aligned}
$$

where $D_{t,j}$ is the demand of customer j in time period t.

The trilevel formulation for the capacity planning in a competitive environment is a difficult mathematical problem. There are no standard techniques to solve this kind of problems and most of the available literature in multilevel programming focuses on bilevel problems. Therefore, the first step to address it is to reformulate the two lower levels as a single-level optimization problem. Eqns. (8)-(12) is a bilevel formulation modeling the problem of the competitors and the market; the upper level only has discrete variables and the lower level is an LP. Hence, the problem can be transformed into a single-level formulation by replacing the lower level with its optimality conditions.

The most common approach to reformulate a bilevel optimization leverages convexity of the lower level to characterize the set of optimal lower-level solutions using the Karush-Kuhn-Tucker (KKT) optimality conditions. However, in lower-level problems with inequality constraints the KKT approach might be ineffective because it requires the addition of many complementarity constraints. The duality-based approach described by Garcia-Herreros et al. [10] is better suited to model the capacity expansion problem with rational markets because it can be reformulated as a single level MILP without additional discrete variables. The idea is to replace the lower-level LP by constraints guaranteeing primal feasibility, dual feasibility, and strong duality. Hence the set of optimal solutions to the problem (17)-(20) is described by the Eqn. (21)-(25)

$$\sum_{t\in T}\sum_{i\in I}\sum_{j\in J} P_{t,i,j} y_{t,i,j} = \sum_{t\in T}\left[\sum_{j\in J} D_{t,j}\lambda_{t,j} - \sum_{i\in I} c_{t,i}\mu_{t,i}\right] \tag{21}$$

$$\sum_{j\in J} y_{t,i,j} \le c_{t,i} \qquad \forall\ t\in T, i\in I \tag{22}$$

$$\sum_{i\in I} y_{t,i,j} = D_{t,j} \qquad \forall\ t\in T, j\in J \tag{23}$$

$$\lambda_{t,j} - \mu_{t,i} \le P_{t,i,j} \qquad \forall\ t\in T, i\in I, j\in J \tag{24}$$

$$y_{t,i,j};\quad \mu_{t,i}\in\mathbb{R}^+;\quad \lambda_{t,j}\in\mathbb{R} \qquad \forall\ t\in T, i\in I, j\in J \tag{25}$$

where Eqn. (21) enforces strong dualigy and Eqn. (24) are the dual constraints corresponding to primal variables $y_{t,i,j}$. Dual variables associated to Eqn. (18) are denoted by $\mu_{t,i}\in\mathbb{R}^+$ and dual variables associated to Eqn. (19) are denoted by $\lambda_{t,j,k}\in\mathbb{R}$.

It is important to note that Eqn. (21) contains bilinear terms in the product of upper-level variables $c_{t,i}$ and dual variables $\mu_{t,i}$. Bilinear terms are nonconvex; however, we can apply an exact linearization procedure [11] because variables $c_{t,i}$ only take discrete values. The non-linearity is avoided by describing capacities ($c_{t,i}$) in terms of the expansion decisions, according to Eqn. (7) and Eqn. (11). Additionally, new variables ($u_{t,t',i}$) defined for the product of dual variables $\mu_{t,i}$ and expansion variables $x_{t',i}$ are introduced in the formulation. Hence, (21) can be replaced by the Eqns. (26)-(28).

$$\sum_{t\in T}\sum_{i\in I}\sum_{j\in J} P_{t,i,j} y_{t,i,j} = \sum_{t\in T}\left(\sum_{j\in J} D_{t,j}\lambda_{t,j} - \sum_{i\in I} C_{0,i}\mu_{t,i} - \sum_{i\in I}\sum_{t'\in T_t^-} H_i u_{t,t',i}\right) \tag{26}$$

$$u_{t,t',i} \geq \mu_{t,i} - M(1 - x_{t',i}) \qquad \forall\, t \in T, t' \in T_t^-, i \in I \tag{27}$$

$$u_{t,t',i} \in \mathbb{R}^+ \qquad \forall\, t \in T, t' \in T_t^-, i \in I \tag{28}$$

In this case, the exact linearization of the bilinear terms in Eqn. (21) is achieved with only two linear inequalities (Eqns. (27)-(28)) because they are enough to bound variables $u_{t,t',i}$ in the improving direction of the objective function.

The bilevel reformulation of the capacity planning problem in a competitive environment is obtained by replacing Eqn. (12) in the trilevel formulation with the constraints modeling the rational behavior of the market. The optimal response of the market is characterized with the primal feasibility constraints presented in Eqns. (22)-(23), the dual feasiblity constraints presented in Eqn. (24), the linearized version of the strong duality constraint presented in Eqns. (26)-(28), and the domains presented in Eqn. (25).

4. Multilevel programming and degeneracy

The optimal solution of a multilevel program might not be strictly defined if a lower-level problem has several optimal responses to the decisions of the higher levels. Degeneracy gives rise to ambiguity in the lower-level decision criterion because the same optimal objective values can be obtained from a set of responses that might produce different effects in the higher levels. The interpretations of degeneracy have been studied for bilevel programs, but it has not been addressed before in multilevel programming. We first offer some background on bilevel optimization in order to present the definitions needed for the trilevel case and our algorithms.

4.1. Review on bilevel optimization

We consider the general bilevel optimization problem presented in Eqn. (29)-(32). The set of variables is partitioned such that $x \in X$ is the vector variables controlled by the upper level, $y \in Y$ is the vector of variables controlled by the lower level, and u is the vector of dual variables associated with constraints (32).

$$\max_{x \in X} \quad c_1^T x + c_2^T y \tag{29}$$

$$\text{s.t.} \quad A_1 x \leq a \tag{30}$$

$$\max_{y \in Y} \quad d^T y \tag{31}$$

$$\text{s.t.} \quad B_1 x + B_2 y \leq b \tag{32}$$

Our focus is on Bilevel Integer Linear Programs (BILP) and Bilevel Mixed-Integer Linear Programs (BMILP) with discrete variables in both levels, for which we introduce Definition 1.

Definition 1. Given the bilevel program presented in Eqns. (29)-(32), let

- Ω be the *Bilevel Constraint Region*:

$$\Omega = \{(x, y) \in X \times Y : A_1 x \leq a, B_1 x + B_2 y \leq b\} \tag{33}$$

- $\Psi(x)$ be the *Rational Reaction set* for a fixed $x \in X$:

$$\Psi(x) = \arg\max_{y \in Y} \left\{ d^T y : B_1 x + B_2 y \leq b \right\} \tag{34}$$

- IR be the *Inducible Region*:

$$IR = \{(x, y) : x \in X, y \in \Psi(x)\} \tag{35}$$

- $R(y)$ be the *Stability Region* of $y \in Y$:

$$R(y^*) := \{x \in X : y^* \in \Psi(x)\} \tag{36}$$

Using Definition 1 we can establish that the inducible region (IR) characterizes the feasible upper-level decisions and their corresponding rational response in the lower level problem. Therefore, IR is the feasible region of the bilevel optimization problem. An explicit description of the IR would allow reformulating the bilevel problem as a single-level model. Unfortunately this region is usually non-convex, non-connected, and in general very hard to describe; consequently, bilevel problems, even when the lower level is an LP, are known to be $\mathcal{NP}$-hard [14]. An important property of bilevel LPs is that their optimal solutions lie at a vertex of region Ω, which allows using enumerative or reformulation techniques [4]. However, this property does not hold when the lower-level problem has discrete variables.

An interesting property of BILPs and BMILPs is that relaxing integrality conditions of the lower-level variables does not yield a relaxation of the bilevel problem [19]. Therefore, a special type of relaxation is required to develop iterative solution methods for these problems. The *High-Point* (HP) relaxation for bilevel problems is introduced in Proposition 1.

Proposition 1. *Let the High-Point (HP) be the problem obtained by removing the lower-level objective function from the bilevel formulation. The resulting single-level problem, presented in Eqn. 37, is a relaxation of the original bilevel problem.*

$$\max_{(x,y) \in \Omega} c_1^T x + c_2^T y \tag{37}$$

The proof to Proposition 1 is presented by Moore and Bard [19]. The proof demonstrates that the feasible region of the bilevel problem is augmented when the optimality condition of the second-level decisions is removed. We can expect the HP problem to provide a weak upper bound on the bilevel problem because in this relaxation all decision variables are controlled by the upper level.

We illustrate the properties of a BMILP and the regions presented in Definition 1 using Example 1. The 3D plot representing Example 1 is shown in Fig. 1.

Example 1.

$$\max_{(x_1, x_2) \in \mathbb{Z}^+} \quad F = x_1 - 1.5x_2 - 10y \tag{38}$$

$$\text{s.t.} \quad x_1 \leq 6 \tag{39}$$

$$x_2 \leq 5 \tag{40}$$

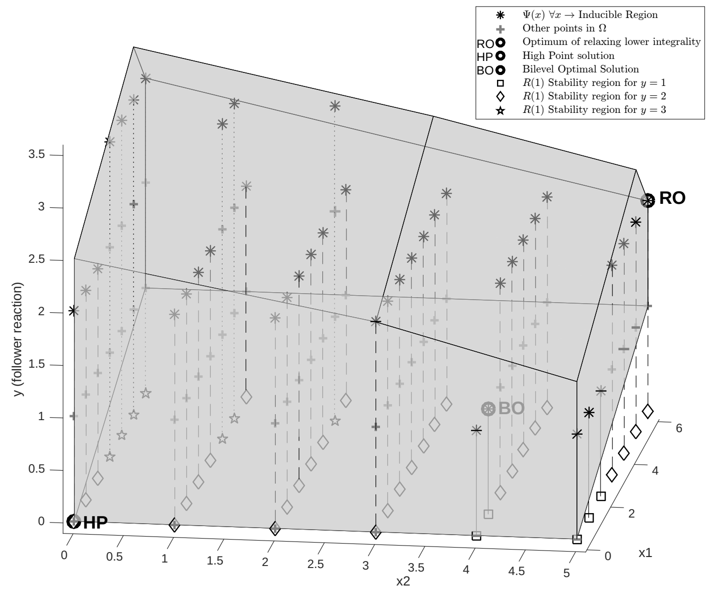

Figure 1: 3D plot of the BMILP presented in Example 1

$$\max_{y \in \mathbb{Z}^+} \quad y \tag{41}$$

$$\text{s.t.} \quad x_1 - 6y \leq 0 \tag{42}$$

$$2.5x_1 + x_2 + 5y \leq 30 \tag{43}$$

$$-3x_1 + 2.5x_2 + 15y \leq 37.5 \tag{44}$$

$$-2x_1 + 2.5x_2 + 10y \leq 27.5 \tag{45}$$

Fig. 1 shows that the optimal solution for bilevel problems with discrete variables do not necessarily lie on a vertex of the convex hull of Ω, which hinders the application of cutting planes methods used in convex optimization. In Example 1, the bilevel optimal solution is $(x_1 = 2, x_2 = 4, y = 1)$ with objective value $F = -15$, whereas the optimal solution to the HP relaxation lies at $(x_1 = 0, x_2 = 0, y = 0)$ with objective value $F = 0$. The optimal solution of the bilevel problem with relaxed integrality in the lower level occurs at $(x_1 = 6, x_2 = 5, y = 2)$, which coincidently is part of the inducible region; however, this solution is not bilevel optimal. Example 1 demonstrates some of the counter-intuitive properties of bilevel programs.

4.2. Degeneracy in bilevel programming

An additional complication of bilevel optimization problems is related to the characterization of optimal solutions when the lower-level problem has multiple optima. Definition 2 describes a modeling approach of a bilevel problem in the presence of degeneracy.

Definition 2. The solution to a bilevel program is considered *optimistic* if any degeneracy of the lower level is resolved in favor of the leader. The rational reaction of the lower-level problem in the *optimistic approach* is formally defined by Eqn. (46).

$$\Psi^{\mathcal{L}}(x) = \underset{y \in \Psi(x)}{\arg\max} \left\{ c_1^T x + c_2^T y \right\} \tag{46}$$

The *optimistic approach* is a common assumption to deal with degeneracy in bilevel programming, mainly because *optimistic solutions* can be found using reformulation techniques. Furthermore an *optimistic* solution can be interpreted as imposing a minimum of collaboration between levels or allowing side-payements from the leader to induce the best reaction among the optimal ones for the follower. However, there is an increasing interest on extending the treatment of degeneracies, studying the alternative formulation. The *pessimistic approach* can be defined as the model in which the lower level selects the response that is most detrimental to the leader in case of degeneracy [7]. These alternative models are considered harder to solve than the *optimistic* approach.

4.3. Degeneracy in trilevel programming

Hierarchical optimization problem with three levels might exhibit new types of solutions. In order to comply with the perfect information assumption, the decision criteria must be completely specified in the case of degeneracy, such that decision-makers that are hierarchically higher can calculate the response of the lower levels. In the following, we propose definitions to clear out ambiguity in our trilevel formulation when the second and third levels have multiple optima.

Definitions for the *Trilevel Constraint Region* (Ω) and the *High-Point* (HP) problem can be extended directly to a trilevel optimization problem. Similarly, the *Inducible Region* (IR) and the *Stability Regions* ($R(y)$ and $R(x^{\mathcal{C}}, y)$) follow the same intuition presented in Definition 1, but their interpretation depend on a new definition of the *Rational Reaction sets.* For notational convenience, we denote by $x^{\mathcal{L}}$ and $x^{\mathcal{C}}$ the first- and second-level decisions, respectively; all other first- and second-level variables can be easily related to them in the capacity planning problem.

Definition 3. The following *Rational Reaction* sets can be identified in a trilevel program.

- The rational reaction set of the third level:

$$\Psi_y(x^{\mathcal{L}}, x^{\mathcal{C}}) = \underset{y \in Y(x^{\mathcal{L}}, x^{\mathcal{C}})}{\arg\min} \left\{ \sum_{t \in T} \sum_{i \in I} \sum_{j \in J} P_{t,i,j} y_{t,i,j} \right\} \tag{47}$$

- The *Sequentially* optimistic reaction set of the third level:

$$\Psi_y^{\mathcal{C}}(x^{\mathcal{L}}, x^{\mathcal{C}}) = \underset{y \in \Psi_y(x^{\mathcal{L}}, x^{\mathcal{C}})}{\arg\max} \left\{ NPV^{\mathcal{C}}(x^{\mathcal{L}}, x^{\mathcal{C}}, y) \right\} \tag{48}$$

- The *Hierarchically* optimistic reaction set of the third level:

$$\Psi_y^{\mathcal{L}}(x^{\mathcal{L}}, x^{\mathcal{C}}) = \underset{y \in \Psi_y(x^{\mathcal{L}}, x^{\mathcal{C}})}{\arg\max} \left\{ NPV^{\mathcal{L}}(x^{\mathcal{L}}, x^{\mathcal{C}}, y) \right\} \tag{49}$$

- The *Sequentially* optimistic reaction set of the second level:

$$\Psi_{x^{\mathcal{C}}}^{Seq}(x^{\mathcal{L}}) = \underset{x^{\mathcal{C}} \in X^{\mathcal{C}}(x^{\mathcal{L}})}{\arg\max} \left\{ NPV^{\mathcal{C}}(x^{\mathcal{L}}, x^{\mathcal{C}}, y) : y \in \Psi_y(x^{\mathcal{L}}, x^{\mathcal{C}}) \right\} \tag{50}$$

- The *Hierarchically* optimistic reaction set of the second level:

$$\Psi^{Hie}_{x^{\mathcal{C}}}(x^{\mathcal{L}}) = \underset{x^{\mathcal{C}} \in X^{\mathcal{C}}(x^{\mathcal{L}})}{\arg\max} \left\{ NPV^{\mathcal{C}}(x^{\mathcal{L}}, x^{\mathcal{C}}, y) : y \in \Psi^{\mathcal{L}}_{y}(x^{\mathcal{L}}, x^{\mathcal{C}}) \right\} \tag{51}$$

The *Rational Reaction* sets presented in Definition 3 suggest a variety of interpretations for degenerate solutions in trilevel programs. We classify the approaches to resolve degeneracy in trilevel programming according to the order in which the upper-level objective functions are favored.

Definition 4. The optimal solution to a trilevel program is considered *Sequentially Optimistic* if degeneracy in the third level is resolved in favor of the second level, and degeneracy in the second level is resolved in favor of the first level. A *Sequentially Optimistic* optimal solution is characterized according to Eqn. (52).

$$\left(\hat{x}^{\mathcal{L}}, \hat{x}^{\mathcal{C}}, \hat{y}\right) = \arg\max \Big\{ NPV^{\mathcal{L}}(x^{\mathcal{L}}, x^{\mathcal{C}}, y) : \; x^{\mathcal{L}} \in X, \; x^{\mathcal{C}} \in \Psi^{Seq}_{x^{\mathcal{C}}}(x^{\mathcal{L}}), \\ y \in \Psi^{\mathcal{C}}_{y}(x^{\mathcal{L}}, x^{\mathcal{C}}) \Big\} \tag{52}$$

Definition 5. The optimal solution to a trilevel program is considered *Hierarchically Optimistic* if degeneracy in the third level is resolved in favor of the first level, and degeneracy in the second level is also resolved in favor of the first level. A *Hierarchically Optimistic* optimal solution is characterized according to Eqn. (53).

$$\left(\hat{x}^{\mathcal{L}}, \hat{x}^{\mathcal{C}}, \hat{y}\right) = \arg\max \big\{ NPV^{\mathcal{L}}(x^{\mathcal{L}}, x^{\mathcal{C}}, y) : \; x^{\mathcal{L}} \in X, \; x^{\mathcal{C}} \in \Psi^{Hie}_{x^{\mathcal{C}}}(x^{\mathcal{L}}), \\ y \in \Psi^{\mathcal{L}}_{y}(x^{\mathcal{L}}, x^{\mathcal{C}}) \big\} \tag{53}$$

Surprisingly, the *Hierarchically Optimistic* model for resolving degeneracy does not guarantee the best possible objective for the first-level decision-maker. Therefore, we present a third optimistic approach to degeneracy.

Definition 6. The optimal solution to a trilevel program is considered *Strategically Optimistic* if degeneracy in the second level is resolved in favor of the first level, and degeneracy in the third level is resolved such that the best first-level solution is obtained. In order to define the *Strategically Optimistic* optimal solution, we characterize the second-level pessimistic reaction set of the third level ($\Upsilon^{\mathcal{C}}_{y}$) according to Eqn. (54).

$$\Upsilon^{\mathcal{C}}_{y}(x^{\mathcal{L}}, x^{\mathcal{C}}) = \underset{y \in \Psi_{y}(x^{\mathcal{L}}, x^{\mathcal{C}})}{\arg\min} \left\{ NPV^{\mathcal{C}}(x^{\mathcal{L}}, x^{\mathcal{C}}, y) \right\} \tag{54}$$

The idea behind the *Strategically Optimistic* model is that the second-level decision-maker accepts any resolution of degeneracy yielding at least the objective value of the second-level pessimistic model. First, let us define in Eqn. (55) the rational reaction set for the second level in the pessimistic framework.

$$\Psi^{\Upsilon}_{x^{\mathcal{C}}}(x^{\mathcal{L}}) = \underset{x^{\mathcal{C}} \in X^{\mathcal{C}}(x^{\mathcal{L}})}{\arg\max} \left\{ NPV^{\mathcal{C}}(x^{\mathcal{L}}, x^{\mathcal{C}}, y) : y \in \Upsilon^{\mathcal{C}}_{y}(x^{\mathcal{L}}, x^{\mathcal{C}}) \right\} \tag{55}$$

Now, the strategic reaction set is defined as the tuple of second- and third-level decisions that belong to the rational reaction set of the third level and yield a better solution to the second level than the second-level pessimistic model. The strategic rational reaction set is defined in Eqn. (56).

$$\Psi^{Str}_{(x^{\mathcal{C}},y)}(x^{\mathcal{L}}) = \big\{ \big(x^{\mathcal{C}}, y\big) : NPV^{\mathcal{C}}(x^{\mathcal{L}}, x^{\mathcal{C}}, y) \geq NPV^{\mathcal{C}}(x^{\mathcal{L}}, \tilde{x}^{\mathcal{C}}, \tilde{y}), \\ \tilde{x}^{\mathcal{C}} \in \Psi^{\Upsilon}_{x^{\mathcal{C}}}(x^{\mathcal{L}}),\ \tilde{y} \in \Upsilon^{\mathcal{C}}_{y}(x^{\mathcal{L}}, \tilde{x}^{\mathcal{C}}),\ y \in \Psi_y \big\} \tag{56}$$

Finally, a *Strategically Optimistic* optimal solution is characterized according to Eqn. (57).

$$\big(\hat{x}^{\mathcal{L}}, \hat{x}^{\mathcal{C}}, \hat{y}\big) = \arg\max \Big\{ NPV^{\mathcal{L}}(x^{\mathcal{L}}, x^{\mathcal{C}}, y) :\ x^{\mathcal{L}} \in X,\ (x^{\mathcal{C}}, y) \in \Psi^{Str}_{(x^{\mathcal{C}},y)}(x^{\mathcal{L}}) \Big\} \tag{57}$$

The difference between the three models for resolving degeneracy is illustrated in Example 2.

Example 2. Fig. 2 describes a case in which different degeneracy resolution models produce different solutions for a given first-level decision ($x^{\mathcal{L}}$). Here, the second level has two rational reactions $x^{\mathcal{C}}_1$ and $x^{\mathcal{C}}_2$, corresponding to different interpretations of third-level degeneracy. If $x^{\mathcal{C}}_1$ is selected, the third level has two alternative optimal reactions (rational reaction set contains outcomes A and B). With $x^{\mathcal{C}}_2$, the third-level is also degenerated (with outcomes C and D).

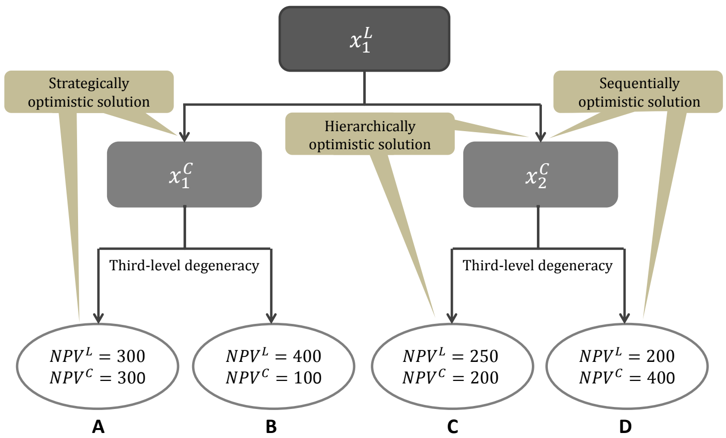

Figure 2: Different types of optimistic solution for a fixed leader decision $x^{\mathcal{L}}_1$

Outcome D is the optimal solution under the *Sequentially Optimistic* model because degeneracy in the third level favors the objective of the competititors ($NPV^{\mathcal{L}} = 200$, $NPV^{\mathcal{C}} = 400$). Under the *Hierarchically Optimistic* model, the optimal solution of the problem is given by outcome C ($NPV^{\mathcal{L}} = 250$, $NPV^{\mathcal{C}} = 200$). This result is counter-intuitive because resolution of third-level degeneracy locally favors the first level, but it forces the second-level to avoid $x^{\mathcal{C}}_1$ and select instead $x^{\mathcal{C}}_2$, which is detrimental for the leader. Also note that even though the *Sequentially Optimistic* model will always give the highest competitors profit for any fixed leader decision, the *Hierarchically Optimistic* model might give a higher competitors profit when solving the full problem. The

Strategically Optimistic solution is outcome A ($NPV^{\mathcal{L}} = 300$, $NPV^{\mathcal{C}} = 300$) and it is always the best a leader can achieve in a trilevel setup. Outcome B is not in the inducible region under any degeneracy resolution model because the competitors will never accept it (will rather have any of the outcomes associated with $x_2^{\mathcal{C}}$).

We have only presented degeneracy resolution models that characterize *optimistic* approaches. However, models for *pessimistic* resolution or mixed resolution (e.g. *optimistic-pessimistic*) can be easily extended from our definitions.

5. Properties of the trilevel capacity planning formulation

In this section we describe the most relevant properties for the algorithms we propose, and indicate the how to exploit them.

5.1. Stability regions

We study the regions $R_{x^{\mathcal{L}}}(\hat{x}^{\mathcal{C}}, \hat{y})$ of the first-level decision space that aggregate points $x^{\mathcal{L}}$ producing the same rational reaction of the lower levels $(\hat{x}^{\mathcal{C}}, \hat{y})$. We can expect the trilevel capacity planning formulation to have large stability regions because the second and third levels are indifferent to the distribution of capacities in the plants controlled by the first level. From the point of view of the market, only the total capacity of the leader in a given time period ($\mathscr{C}_t$) is relevant since all its plants offer the same price.

Definition 7. A *Stability Region* $R_{x^{\mathcal{L}}}(\hat{x}^{\mathcal{C}}, \hat{y})$ in the trilevel capacity planning formulation is the set of first-level decisions that produce the same rational reactions in the second and third levels. Formally, the stability region for some fixed second and third level reactions $(\hat{x}^{\mathcal{C}}, \hat{y})$ is characterized according to Eqn. (58).

$$R_{x^{\mathcal{L}}}(\hat{x}^{\mathcal{C}}, \hat{y}) = \left\{ x^{\mathcal{L}} : \ x^{\mathcal{L}} \in X^{\mathcal{L}}, \ \hat{x}^{\mathcal{C}} \in \Psi^0_{x^{\mathcal{C}}}(x^{\mathcal{L}}), \ \hat{y} \in \Psi^0_y(x^{\mathcal{L}}, \hat{x}^{\mathcal{C}}) \right\} \tag{58}$$

where $\Psi^0_{x^{\mathcal{C}}}(x^{\mathcal{L}})$ and $\Psi^0_y(x^{\mathcal{L}}, \hat{x}^{\mathcal{C}})$ refer to one of the degeneracy resolution models described in Section 4.3.

Another property of the trilevel planning formulation that implies large stability regions can be derived from the intuition that expanding plants with slack capacity does not change the rational response of the second and third levels. Proposition 2 gives the mathematical description of this property.

Proposition 2. *Let $(\hat{Q})$ be the bilevel problem obtained after fixing the first-level decisions to $\hat{x}^{\mathcal{L}}$ in the second- and third-level problems presented in Eqns.* (8)-(14)*. We denote by $(\hat{x}^{\mathcal{C}}, \hat{y})$ a corresponding optimal bilevel solution and by $\hat{\mu}$ the optimal multipliers associated with capacity constraints* (18)*. Then,*

$$x^{\mathcal{L}} \in \Bigg\{ (c_{1,1}, .., c_{|T|,|I^{\mathcal{L}}|}) : \\ \left[\sum_{i \in I^{\mathcal{L}}} c_{t,i} = \sum_{i \in I^{\mathcal{L}}} \hat{c}_{t,i} \right] \vee \left[\begin{array}{c} \sum_{i \in I^{\mathcal{L}}} c_{t,i} \geq \sum_{i \in I^{\mathcal{L}}} \hat{c}_{t,i} \\ \hat{\mu}_{t',i} = 0 \quad \forall \ t' \in T_t^+, \ i \in I^{\mathcal{L}} \end{array} \right] \quad \forall \ t \in T \Bigg\} \tag{59}$$

$$\implies x^{\mathcal{L}} \in R_{x^{\mathcal{L}}}(\hat{x}^{\mathcal{C}}, \hat{y}) \tag{60}$$

where T_t^+ is the subset of periods after time t: $T_t^+ = \{t' : t' \in T, t' \geq t\}$.

Proof. We want to prove that a first-level decision $x^{\mathcal{L}}$ satisfying conditions (59) produces the same rational reaction $(\hat{x}^{\mathcal{C}}, \hat{y})$ as $\hat{x}^{\mathcal{L}}$. We divide the proof of Proposition 2 in three steps.

Step 1. *In the third-level market problem, increasing the capacity of one plant cannot increase the demand assigned to any other plant.*

Let us denote by $(\hat{P})$ the third-level problem with capacities equal to $c_{t,i}$, and by $(\tilde{P})$ the problem in which plant i' increases its capacity by $\Delta C_{i'}$. We want to show that the optimal demand assignments $(y_{t,i,j})$ corresponding to problems $(\hat{P})$ and $(\tilde{P})$ satisfy the conditions presented in Eqn. (61).

$$\sum_{j \in J} \tilde{y}_{t,i,j} \leq \sum_{j \in J} \hat{y}_{t,i,j} \qquad \forall\, t \in T,\; i \in I \backslash \{i'\} \tag{61}$$

First, we notice that fully utilized plants $(\sum_{j \in J} \hat{y}_{t,i,j} = c_{t,i})$ in problem $(\hat{P})$ cannot increase the demand assigned to them. For all other plants with slack capacity $(\sum_{j \in J} \hat{y}_{t,i,j} + \hat{\epsilon}_{t,i} = c_{t,i})$, the Lagrange multiplier $(\hat{\mu}_{t,i})$ associated with the capacity constraint (18) must be zero according to complementary slackness of the third-level LP.

It was proved by Garcia-Herreros et al. [10] that increasing the capacity of one plant cannot produce an increase in the optimal Lagrange multipliers associated with any capacity constraint (18). In this case, the Lagrange multipliers $(\hat{\mu}_{t,i})$ of plants that had slack capacity in problem $(\hat{P})$ remain at zero in the optimum of problem $(\tilde{P})$, as expressed in Eqn. (62)

$$0 \leq \tilde{\mu}_{t,i} \leq \hat{\mu}_{t,i} = 0 \qquad \forall\, (t,i) \in \{(t,i):\; t \in T, i \in I \backslash \{i'\}, \hat{\epsilon}_{t,i} > 0\} \tag{62}$$

Since the slack $(\hat{\epsilon}_{t,i})$ in plants that are not fully utilized in problem $(\hat{P})$ can be arbitrarily small, we conclude that the condition in Eqn. (61) must be satisfied. Otherwise, an increase in the demand assignments would produce a positive value in the Lagrange multipliers $(\tilde{\mu}_{t,i} > 0)$ associated to capacity constraints.

Step 2. *The optimal objective value of the second level cannot improve if the total capacity of the leader increases $\left(\sum_{i \in I^{\mathcal{L}}} \tilde{x}_{t,i} \geq \sum_{i \in I^{\mathcal{L}}} \hat{x}_{t,i}\ \forall\ \ t \in T\right)$ and capacities of the competitors remain constant.*

Recall that the prices offered by competitors are defined in (2). Rewriting the objective function of the competition as in Eqn. (63), it is easy to observe that the margin obtained from every unit sold only dependents on the production cost $(F_{t,i})$ and the raw price $(P_{t,i}^{raw})$ of each plant.

$$\begin{aligned} \text{NPV}^{\mathcal{C}} = &\sum_{t \in T} \sum_{i \in I^{\mathcal{C}}} \sum_{j \in J} \left(P_{t,i}^{raw} - F_{t,i}\right) y_{t,i,j} \\ &- \sum_{t \in T} \sum_{i \in I^{\mathcal{C}}} \left(A_{t,i} v_{t,i} + B_{t,i} w_{t,i} + E_{t,i} x_{t,i}\right) \end{aligned} \tag{63}$$

Now remember that the prices are set for the leader according to Eqn. (1), which implies that with respect to the rest of players, the leader can be seen as being a single plant. Thus any expansion $\left(\sum_{i\in I^{\mathcal{L}}} \tilde{x}_{t,i} \geq \sum_{i\in I^{\mathcal{L}}} \hat{x}_{t,i} \ \forall \ \ t \in T\right)$ fits exactly the case of Step 1. Therefore, the condition presented in Eqn. (61) also implies that the objective function of the second level cannot improve from problem $(\hat{P})$ to $(\tilde{P})$. This condition is formalized in Eqn. (64).

$$NPV^{\mathcal{C}}(\tilde{x}^{\mathcal{L}}, x^{\mathcal{C}}, \tilde{y}) \leq NPV^{\mathcal{C}}(\hat{x}^{\mathcal{L}}, x^{\mathcal{C}}, \hat{y}) \quad \forall\ x^{\mathcal{C}} \in X^{\mathcal{C}}, \quad \begin{array}{l} \tilde{y} \in \Psi_y^0(\hat{x}^{\mathcal{L}}, x^{\mathcal{C}}) \\ \hat{y} \in \Psi_y^0(\hat{x}^{\mathcal{L}}, x^{\mathcal{C}}) \end{array} \tag{64}$$

Step 3. *If the expansion strategy of the leader ($\tilde{x}_{t,i}$) satisfies the conditions presented in Eqn.* (59)*, the bilevel problems ($\hat{Q}$) and ($\tilde{Q}$), resulting from fixing the variables of the leader to $\hat{x}^{\mathcal{L}}$ and $\tilde{x}^{\mathcal{L}}$ respectively, have the same rational reactions.*

First, we use the duality-based reformulation presented in Eqns. (21)-(25) to verify that optimal solutions $(\hat{y})$ to problem $(\hat{P})$ are feasible in $(\tilde{P})$. This is the case because capacity constraints (22) are relaxed with the additional expansions of the leader, and the dual objective function (right-hand side of Eqn. (21)) only changes in coefficients $(c_{t,i})$ for which the optimal Lagrange multipliers $(\hat{\mu}_{t,i})$ are equal to zero. Then, the optimal solution $(\hat{x}^{\mathcal{C}}, \hat{y})$ of the bilevel problem $(\hat{Q})$ resulting from fixing the first-level decisions to $\hat{x}^{\mathcal{L}}$ in Eqns. (8)-(14), is feasible in the bilevel problem $(\tilde{Q})$ since second-level constraints (9)-(11) are not affected by first-level decisions. Therefore, because the objective function of (Q) only depends on $(x^{\mathcal{C}}, y)$ and not $x^{\mathcal{L}}$, the optimal value of problem $(\tilde{Q})$ must be at least as large as the optimal value of problem $(\hat{Q})$; this condition is formalized in Eqn. (65).

$$NPV^{\mathcal{C}}(\hat{x}^{\mathcal{L}}, \hat{x}^{\mathcal{C}}, \hat{y}) = NPV^{\mathcal{C}}(\tilde{x}^{\mathcal{L}}, \hat{x}^{\mathcal{C}}, \hat{y}) \leq NPV^{\mathcal{C}}(\tilde{x}^{\mathcal{L}}, \tilde{x}^{\mathcal{C}}, \tilde{y}) \tag{65}$$

where $\left(\tilde{x}^{\mathcal{C}}, \tilde{y}\right)$ is the optimal solution of $(\tilde{Q})$. Furthermore, we can establish the inequalities given in Eqn. (66),

$$NPV^{\mathcal{C}}(\tilde{x}^{\mathcal{L}}, \tilde{x}^{\mathcal{C}}, \tilde{y}) \leq NPV^{\mathcal{C}}(\hat{x}^{\mathcal{L}}, \tilde{x}^{\mathcal{C}}, \hat{\tilde{y}}) \leq NPV^{\mathcal{C}}(\hat{x}^{\mathcal{L}}, \hat{x}^{\mathcal{C}}, \hat{y}) \tag{66}$$

where $\hat{\tilde{y}}$ is the optimum of $\hat{Q}$ with $x^{\mathcal{C}}$ fixed to $\tilde{x}^{\mathcal{C}}$. The inequality on the left is derived from Eqn. (64), and the inequality on the right follows from optimality of $\left(\hat{x}^{\mathcal{C}}, \hat{y}\right)$ in problem $(\hat{Q})$. Eqns. (65) and (66) together demonstrate that problems $(\hat{Q})$ and $(\tilde{Q})$ have the same optimal objective value that can be given by the optimal solution $\left(\hat{x}^{\mathcal{C}}, \hat{y}\right)$. We conclude that $\hat{x}^{\mathcal{L}}$ and $\tilde{x}^{\mathcal{L}}$ are in the same stability region $R_{x^{\mathcal{L}}}(\hat{x}^{\mathcal{C}}, \hat{y})$, which proves Proposition 2.

5.2. A cut to eliminate $R_{x^{\mathcal{L}}}(\hat{x}^{\mathcal{C}}, \hat{y})$

The capacity planning problem with competitive decision-makers is likely to have large stability regions as a consequence of Proposition 2. All the $x^{\mathcal{L}} \in X^{\mathcal{L}}$ in the same stability region than a certain $\hat{x}^{\mathcal{L}}$ can be characterized from the dual variables $\hat{\mu}^*$ of the optimal solution of the bilevel problem $(\hat{Q})$. As will be exploited in the algorithm of section 6, we don't need to solve again the lower level bilevel problem for any $x^{\mathcal{L}}$ in this stability region and hence here we will deduce a cut that eliminates every point in the stability regions we compute. To do so, we introduce the following parameters and sets.

Definition 8. Given the optimal bilevel solution $(x^{\mathcal{C},k}, y^k, \mu^k, \lambda^k)$ corresponding to problem (Q^k) with first-level decisions fixed to $(x^{\mathcal{L},k})$, we define:

- The total capacity of the leader in time period t:

$$\mathscr{C}_t^k = \sum_{i \in I^{\mathcal{L}}} C_{0,i} + H_i \sum_{t' \in T_t^-} \sum_{i \in I^{\mathcal{L}}} x_{t,i}^k \tag{67}$$

- The subset of time periods in which all plants controlled by the leader do not expand but expansions could change demand assignments:

$$\Gamma_{x0}^k = \left\{ t \in T : \sum_{i \in I^{\mathcal{L}}} x_{t,i}^k = 0, \quad \sum_{t' \in T_t^+} \mu_{t',i}^k > 0 \right\} \tag{68}$$

- The subset of time periods in which the leader expands and further expansions could change demand assignments:

$$\Gamma_{\mu+}^k = \left\{ t \in T : \sum_{i \in I^{\mathcal{L}}} x_{t,i}^k > 0, \quad \sum_{t' \in T_t^+} \mu_{t',i}^k > 0 \right\} \tag{69}$$

- The subset of time periods in which the leader expands but further expansions would not change demand assignments:

$$\Gamma_{\mu0}^k = \left\{ t \in T : \sum_{i \in I^{\mathcal{L}}} x_{t,i}^k > 0, \quad \sum_{t' \in T_t^+} \mu_{t',i}^k = 0 \right\} \tag{70}$$

We characterize the *Stability Region* (R^k) where the first-level decisions $(v^{\mathcal{L},k}, w^{\mathcal{L},k}, x^{\mathcal{L},k}, c^{\mathcal{L},k})$ lies by identifying if an alternative first-level solution satisfies the conditions presented in Proposition 2. In Eqn. 71, we introduce binary variables $z_{0,t}^k$ and $z_{1,t}^k$ to indicate if alternative solutions offer more, less, or the same capacity for the leader with respect to $\mathscr{C}_t^k$.

$$\begin{bmatrix} z_{0,t}^k = 1 \\ \sum\limits_{i \in I^{\mathcal{L}}} c_{t,i} = \mathscr{C}_t^k \end{bmatrix} \vee \begin{bmatrix} z_{1,t}^k = 1 \\ \sum\limits_{i \in I^{\mathcal{L}}} c_{t,i} < \mathscr{C}_t^k \end{bmatrix} \vee \begin{bmatrix} z_{0,t}^k + z_{1,t}^k = 0 \\ \sum\limits_{i \in I^{\mathcal{L}}} c_{t,i} > \mathscr{C}_t^k \end{bmatrix} \quad \forall\, t \in T \tag{71}$$

Based on the variables that compare capacities in alternative solutions to capacities in the solution of problem (Q^k), we can characterize the rest of the *Stability Region* of the solution $(v^{\mathcal{L},k}, w^{\mathcal{L},k}, x^{\mathcal{L},k}, c^{\mathcal{L},k})$ with Eqn. 72.

$$\sum_{i \in I^{\mathcal{L}}} \sum_{t \in \Gamma_{x0}^k} x_{t,i} + \sum_{t \in \Gamma_{\mu0}^k} z_{1,t}^k + \sum_{t \in \Gamma_{\mu+}^k} (1 - z_{0,t}^k) = 0 \tag{72}$$

A *no-good* cut to exclude all solutions that belong to this region $(x^{\mathcal{L}} \in R^k)$ is obtained by forcing the left-hand side of Eqn. (72) to be greater or equal than one (≥ 1).

5.3. Equations to tighten the HP relaxation

The *High-Point* (HP) problem usually yields a weak lower bound to the bilevel optimization problem because it gives control of all variables to the first level. The column-and-constraint generation method developed by Zeng and An [25] proposes a strategy to tighten the HP relaxation based on second-level reactions for which the second-level objective value is known. The idea is to generate cuts that constrain the second-level objective function to be at least as good as it would be with any of the second-level solutions that have been observed. These constraints, presented in Eqn. (73), are included in the HP problem to model the reactions of the second level.

$$NPV^{\mathcal{C}}(x^{\mathcal{L}}, x^{\mathcal{C}}, y) \geq NPV^{\mathcal{C}}(x^{\mathcal{L}}, \hat{x}^{\mathcal{C},k}, y^k) \tag{73}$$

where $\hat{x}^{\mathcal{C},k}$ are parameters modeling a fixed second-level response, and y^k are duplicate variables that model optimal demand assignments for any first-level decision ($x^{\mathcal{L}}$) when the second-level response ($\hat{x}^{\mathcal{C},k}$) is fixed. In order to enforce the third-level optimality of demand assignments, a full set of duplicate variables ($y^k_{t,i,j}, \mu^k_{t,i}, u^k_{t,t',i}, \lambda^k_{t,k}$) and constraints must be appended to the HP problem for each solution that has been observed. The constraints correspond to the duality-based reformulation of the third-level problem; they are presented in Eqns. (74)-(80).

$$\sum_{t\in T}\sum_{i\in I}\sum_{j\in J} P_{t,i,j} y^k_{t,i,j} = \sum_{t\in T}\left(\sum_{j\in J} D_{t,j}\lambda^k_{t,j} - \sum_{i\in I} C_{0,i}\mu^k_{t,i} + \sum_{i\in I}\sum_{t'\in T^-_t} H_i u^k_{t,t',i}\right) \tag{74}$$

$$\sum_{j\in J} y^k_{t,i,j} \leq c_{t,i} \qquad \forall\, t\in T, i\in I^{\mathcal{L}} \tag{75}$$

$$\sum_{j\in J} y^k_{t,i,j} \leq \hat{c}^k_{t,i} \qquad \forall\, t\in T, i\in I^{\mathcal{C}} \tag{76}$$

$$\sum_{i\in I} y^k_{t,i,j} = D_{t,j} \qquad \forall\, t\in T, j\in J \tag{77}$$

$$\lambda^k_{t,j} - \mu^k_{t,i} \leq P_{t,i,j} \qquad \forall\, t\in T, i\in I, j\in J \tag{78}$$

$$u^k_{t,t',i} \geq \mu^k_{t,i} - M(1 - x_{t',i}) \qquad \forall\, t\in T, t'\in T^-_t, i\in I^{\mathcal{L}} \tag{79}$$

$$y^k_{t,i,j};\quad \mu^k_{t,i},\quad u^k_{t,t',i} \in \mathbb{R}^+;\quad \lambda^k_{t,j}\in\mathbb{R} \qquad \forall\, t\in T, i\in I, j\in J \tag{80}$$

We observe that the cuts modeled by Eqns. (73)-(80) do not exclude any solution that is trilevel feasible. All first-level solutions remain feasible after the cuts are appended to the HP problem because we assume that there is always enough capacity in the third level to satisfy all demands. Therefore, the duality-based reformulation of the third level always has a feasible solution for any fixed $x^{\mathcal{L}}$ and $x^{\mathcal{C}}$. Additionally, we can guarantee that no point in the *Inducible Region* of the trilevel problem is excluded from the tightened HP problem because Eqn. (73) provides lower bounds on $NPV^{\mathcal{C}}$ based on solutions that are feasible in the second- and third-level problems; solutions in the *Inducible Region* must be optimal in the second and third levels, which implies that their corresponding $NPV^{\mathcal{C}}$ must be greater or equal than any bound imposed by inequality (73). Note that Eqns. (73)-(80) are appended to the HP problem where the objective function is $NPV^{\mathcal{L}}$. Therefore in case of degeneracy it is resolved by the leader and the *rhs* of Eqn. (73) is pushed to the worst possible outcome for the competitors. Thus again it never cuts any trilevel solution, no matter the degeneration-resolution chosen. As will be seen in section 7, this is used to find the strategically optimistic solution.

The next two sections describe our two solution approaches that converge to the different *optimistic* solutions.

6. Algorithm 1: Constraint-directed exploration

We use the stability regions of the capacity planning problem and the equations describing them to design a constraint-directed exploration of the leader's decision space. Algorithm 1 performs an accelerated search on the inducible region by iteratively solving a restricted high-point problem HP^k where cuts are added to prevent selecting any leader decision $x^{\mathcal{L}}$ belonging to any of the previously computed stability regions (for which the solution is hence known). The details of the algorithm are presented below.

6.1. Reaching the sequentially optimistic solution

After solving the HP^k, the algorithm finds a trilevel feasible solution inside the stability region (R^k) where the obtained $x^{\mathcal{L}}_{HP^k}$ lies. This is done by fixing the leader decision to $x^{\mathcal{L}}_{HP^k}$ and solving the single-level reformulation (Q^k) of the second- and third-level problems. From that solution we don't only get a lower bound but we can now describe the stability region of this observed reaction of the followers (R^k) by using Eqn. (72). By adding this cut to the next HP^{k+1}, the search is directed towards unexplored first-level decisions. Convergence of the algorithm is guaranteed because the problem has a finite number of first-level decisions, and every iteration eliminates one stability region that contains at least one new point. The operations performed by the algorithm are divided in six steps.

Step 1: Solve HP^k over the unexplored first-level decision space. Identify the first-level solution ($x^{\mathcal{L}}_{HP^k}$). If HP^k is infeasible, terminate and return the incumbent.

Step 2: Update the upper bound (UB). If UB is less than the best lower bound (LB^*), terminate and return the incumbent.

Step 3: Solve Q^k with first-level variables fixed to $x^{\mathcal{L}}_{HP^k}$. Identify the second- and third-level solution ($x^{\mathcal{C}}_{Q^k}, y_{Q^k}, \mu_{Q^k}$).

Step 4: Identify the sets Γ^k_{x0}, $\Gamma^k_{\mu+}$, and $\Gamma^k_{\mu 0}$ describing the region R^k that contains $x^{\mathcal{L}}_{HP^k}$ and all other first-level solutions satisfying the condition given by Eqn. (59).

Step 5: Update LB^* if solution ($x^{\mathcal{L}}_{HP^k}, x^{\mathcal{C}}_{Q^k}, y_{Q^k}$) is better than the incumbent. Terminate if UB is equal to LB^*.

Step 6: Generate *no-good* cuts to exclude R^k from HP^{k+1}. Go back to Step 1.

Algorithm 1 has two possible stopping criteria:

C1: If $UB < LB^*$ in Step 2 or Step 5, return incumbent. In this case, no solution contained in the unexplored region of the first-level decision space can be better than the incumbent.

C2: If HP^k is infeasible in Step 1, return incumbent. In this case, the first-level decision space has been exhaustively analyzed.

It is worth noticing that Step 1 produces an improving UB because the feasible region of problem HP^k is successively reduced. On the other hand, Step 3 finds a trilevel feasible solution that corresponds to the *sequentially optimistic* model of degeneracy because problem Q^k resolves degeneracy in favor of the second level ($y \in \Psi_y^{\mathcal{C}}$). A *sequentially optimistic* solution might be very detrimental for the first level since demands assigned to the leader are degenerate according to the pricing model presented in Eqn. (1). Furthermore, instances with a degenerate third level might not close the gap between UB and LB^* because problems HP^k and Q^k use different degeneracy resolution models. In this case, an exhaustive search could be necessary to meet stopping criterion C2 and yield the incumbent, corresponding to the *sequentially optimistic* solution.

6.2. Reaching the hierarchically optimistic solution

Several additional operations are needed to instruct the algorithm to obtain the *hierarchically optimistic* solution. The idea is to modify Step 4 to find among the degenerate solutions the one that favors the leader according to the *hierarchically optimistic* model. Two additional optimization problems must be defined: the high-point problem ($HP_R^K(x^{\mathcal{C}}_{Q^k})$) constraint to region R^k with second-level variables fixed to $x^{\mathcal{C}}_{Q^k}$, and the high-point problem ($HP_R^K(NPV^{\mathcal{C}})$) constraint to region R^k with second-level objective value fixed to $NPV^{\mathcal{C}}(x^{\mathcal{L}}_{HP^k}, x^{\mathcal{C}}_{Q^k}, y_{Q^k})$.

Solving ($HP_R^K(x^{\mathcal{C}}_{Q^k})$) has two purposes: first, to find the best first-level solution in $x^{\mathcal{L}} \in R^k$ knowing that the second-level response will be $x^{\mathcal{C}}$. This also re-organizes the third level assignments to the leader so to fit the best supply scheme for it -without affecting the benefit of the market. Second, to detect if the market is degenerated in the sense of having different optimal assignments that yield different $NPV^{\mathcal{C}}$. This is the case if the new solution $y_{HP_R^k}$ has a different aggregated demand to every competitor's plant. If it is the case, we cannot conclude anything about the solution $(x^{\mathcal{L}}_{HP_R^k}, x^{\mathcal{C}}_{Q^k}, y_{HP_R^k})$ being trilevel hierarchically feasible because, if the competitors knew that the market would directly favor the leader, they might choose another expansion plan $x^{\mathcal{C}}$. Hence a penalty must be applied to the market in favor of the leader and go back to Step 3. Notice that by how the prices are constructed, it is enough to check that the total aggregated demand to the competitors is the same (if there are changes from one plant to another it can be proven that there is another optimal solution for the market that yields the same assignments to the competitor as in Q^k).

If the market was not found to be degenerated (or it was solved by a penalty), we still have to check whether the second level is degenerated in the sense of having another possible expansion $x^{\mathcal{C}}$ that yields the same value for it but a better one for the leader. This is checked by solving ($HP_R^K(_{NPV^c})$), and if the objective value of the leader changes, we need to add a penalty to the competitors and go back to step 3. A detailed description of the steps required to reach the hierarchically optimistic solution are presented below.

Step 4a: Identify the sets Γ_{x0}^k, $\Gamma_{\mu+}^k$, and $\Gamma_{\mu 0}^k$ describing the region R^k that contains $x_{HP^k}^{\mathcal{L}}$ and all other first-level solutions satisfying the condition given by Eqn. (2).

Step 4b: Solve $HP_R^K(x_{Q^k}^{\mathcal{C}})$ and identify the third-level response ($y_{HP_R^k}$). If the third-level solution is different from the one obtained in Step 3 ($\sum_{i\in I^c} y_{Q^k} \neq \sum_{i\in I^c} y_{HP_R^k}\ \forall\, t\in T$), add a penalty to the third-level objective to resolve degeneracy in favor of the first level. Go back to Step 3.

Step 4c: Solve $HP_R^K(_{NPV^c})$. If the first-level objective is different from the one obtained in Step 3 ($NPV^{\mathcal{L}}(_{x_{HP^k}^{\mathcal{L}}, x_{Q^k}^{\mathcal{C}}, y_{Q^k}}) \neq NPV^{\mathcal{L}}(_{x_{HP_R^k}^{\mathcal{L}}, x_{Q^k}^{\mathcal{C}}, y_{HP_R^k}})$), add a penalty to the second-level objective to resolve degeneracy in favor of the first level. Go back to Step 3.

The steps of the algorithm are presented schematically in Fig. 3; diamonds control the flow of the algorithm, light gray boxes are simple operations and dark gray boxes involve optimization problems.

7. Algorithm 2: Column-and-constraint generation algorithm

In opposition to Algorithm 1, Algorithm 2 finds optimal trilevel solutions by exploring the decision space of the second-level problem. Algorithm 2 is inspired in the column-and-constraint generation algorithm developed by Zeng and An [25] for linear bilevel problems with mixed-integer variables in both levels. However, our algorithm operates over the bilevel reformulation of the trilevel capacity planning problem, which already enforces optimality of the variables control by the markets; therefore, no additional reformulation is needed for the continuous variables. The details of the algorithm are presented below.

7.1. Reaching the strategically optimistic solution

Algorithm 2 uses the cuts presented in Section 5.3 to sequentially tighten the *high-point* relaxation of the trilevel capacity planning problem. The algorithm iterates between a master problem (MP^k) that provides upper bounds (UP^k) and the single-level reformulation of the second- and third-level problems (Q_R^k). Problem MP^k is the high-point relaxation of the bilevel reformulation with the cuts modeled by Eqns. (73)-(80). The search is directed towards unexplored second-level decisions by adding no-good cuts to problem Q_R^k, such that second-level decisions that were already observed are not considered in future iterations. The no-good cuts used to diversify the search in the second-level decision space are presented in Eqn. (81).

$$\sum_{(t,i)\in\left\{(t,i):\ [x_{i,j}^{\mathcal{C}}]_{Q_R^{k'}}=1\right\}} (1 - x_{t,i}^{\mathcal{C}}) \quad + \sum_{(t,i)\in\left\{(t,i):\ [x_{i,j}^{\mathcal{C}}]_{Q_R^{k'}}=0\right\}} x_{t,i}^{\mathcal{C}} \quad \geq \quad 1 \qquad \forall\, k' \in K \qquad (81)$$

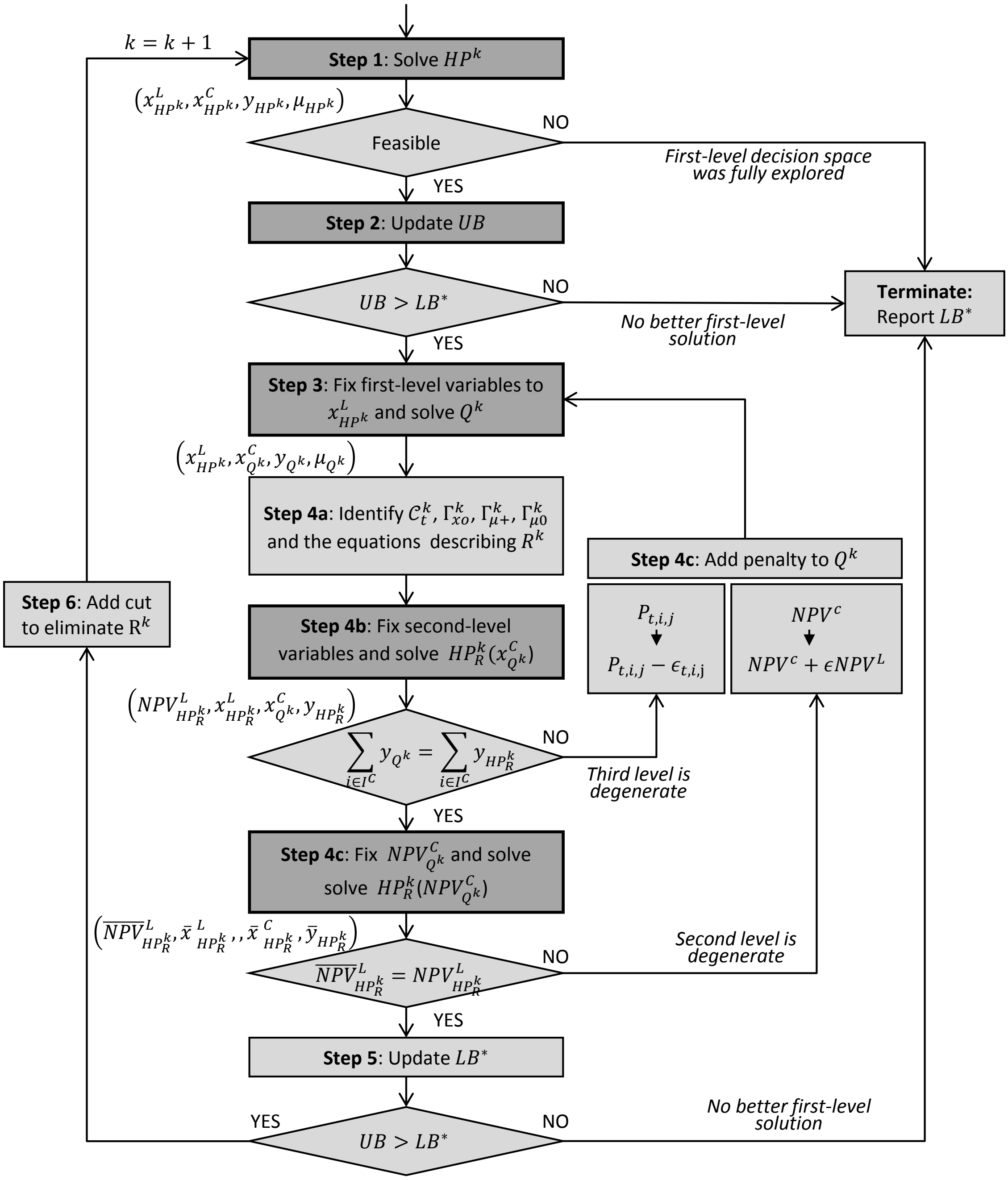

Figure 3: Algorithm 1 towards the hierarchically optimistic solution

where $[x^{\mathcal{C}}_{i,j}]_{Q^{k'}_R}$ denotes the second-level optimal solution for problem $Q^{k'}_R$ and $K = \{1, 2, \ldots, k\}$.

The algorithm is identified as a column-and-constraint generation approach because at every iteration, a new second-level candidate solution ($\hat{c}_{t,i}\ \forall\, t \in T, i \in I^{\mathcal{C}}$) is appended to MP^k, together with the constraints and variables modeling the third-level optimal response. Convergence of Algorithm 2 is guaranteed because the problem has a discrete number of second-level decisions, which implies that a finite number of different columns and constraints can be added to MP^k. The operations performed by Algorithm 2 are divided in five steps.

Step 1: Solve MP^k. Identify the first-level solution ($x^{\mathcal{L}}_{MP^k}$) and the second-level objective value $NPV^{\mathcal{C}}(x^{\mathcal{L}}_{MP^k}, x^{\mathcal{C}}_{MP^k}, y_{MP^k})$.

Step 2: Update the upper bound (UB). If UB is less or equal to the best lower bound (LB^*), terminate and return the solution yielding LB^*.

Step 3: Fix first-level variables to $x^{\mathcal{L}}_{MP^k}$ and solve $\bar{Q}^k_R$, which is Q^k_R including the no-good cuts from Eqn. (81). If infeasible, terminate and return the solution yielding UB . Otherwise, identify the second-level solution ($x^{\mathcal{C}}_{Q^k}$). If $NPV^{\mathcal{C}}(x^{\mathcal{L}}_{MP^k}, x^{\mathcal{C}}_{Q^k_R}, y_{Q^k_R}) < NPV^{\mathcal{C}}(x^{\mathcal{L}}_{MP^k}, x^{\mathcal{C}}_{MP^k}, y_{MP^k})$, terminate and return the solution yielding UB.

Step 4: Update the best LB^*. If UB is less or equal to the best lower bound (LB^*), terminate and return the solution yielding LB^*.

Step 5: Generate the columns and constraints tighten MP^{k+1} and the cuts to exclude $x^{\mathcal{C}}_{Q^k_R}$ from $\bar{Q}^k_R$. Go back to Step 1.

The steps of the algorithm are presented schematically in Fig. 4. Algorithm 2 has three possible stopping criteria:

C1: If $UB \leq LB^*$ in Step 2 or in Step 4, both problems MP^k and $\bar{Q}^*_R$ yield the same optimal value ($NPV^{\mathcal{L}}(x^{\mathcal{L}}_{MP^k}, x^{\mathcal{C}}_{MP^k}, y_{MP^k})$). This only happens if there is no third-level degeneracy favoring the second-level in $\bar{Q}^*_R$.

C2: If $\bar{Q}^k_R$ is infeasible in Step 3, return the solution $(x^{\mathcal{L}}_{MP^k}, x^{\mathcal{C}}_{MP^k}, y_{MP^k})$ obtained from MP^k. In this case, the second-level decision space has been exhaustively analyzed.

C3: If $NPV^{\mathcal{C}}(x^{\mathcal{L}}_{MP^k}, x^{\mathcal{C}}_{MP^k}, y_{MP^k}) \geq NPV^{\mathcal{C}}(x^{\mathcal{L}}_{MP^k}, x^{\mathcal{C}}_{Q^k_R}, y_{Q^k_R})$, return the solution $(x^{\mathcal{L}}_{MP^k}, x^{\mathcal{C}}_{MP^k}, y_{MP^k})$ obtained in MP^k. In this case, no other solution contained in the unexplored region can be better for the second level than $(x^{\mathcal{L}}_{MP^k}, x^{\mathcal{C}}_{MP^k}, y_{MP^k})$.

It is worth noticing that Step 1 produces an improving UB because the feasible region of problem HP^k is successively reduced. Also, the solutions obtained from HP^k correspond to *strategically optimistic* model of degeneracy since the control of all variables is granted to the first level and only a constraint on the second-level objective value is imposed. Neverhteless, Step 3 resolves third-level degeneracy in favor of the second level. Consequently, the gap between UB and LB^* might not close; in this case, either criterion C2 or C3 will be met.

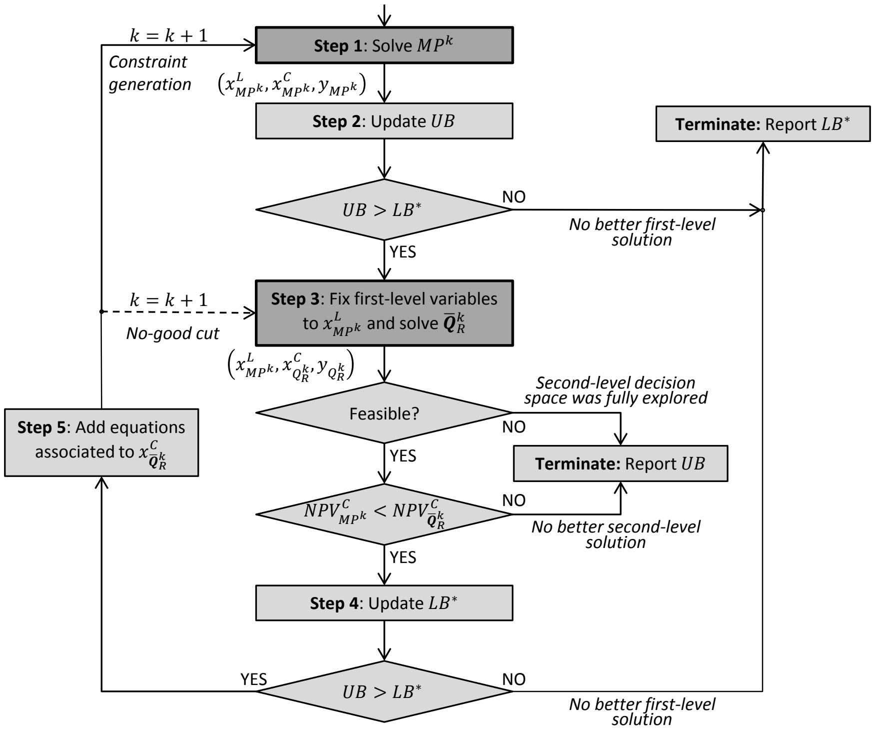

Figure 4: Algorithm 2 towards the strategically optimistic solution

Remark. Algorithm 1 and Algorithm 2 are guaranteed to find the same trilevel optimal solution in instances with no degeneracy at any level. If degeneracy is present, no result can be established about the relative performance of the algorithms because they look for different solutions, and these two problems can be arbitrarily difficult to solve with respect to the other. For non-degenerate instances we can establish that Algorithm 1 explores at least the same number iterations as Algorithm 2. This is the case because Algorithm 2 explores at most one point in each stability region, which is not true for Algorithm 1. However, it does not imply that Algorithm 2 outperforms Algorithm 1 in execution time because Algorithm 2 adds many variables and constraints to MP^k at every iteration, which increases the complexity of the iterations.

8. Capacity planning instances

We test Algorithm 1 and Algorithm 2 using two instances of the capacity planning problem with competitive decision-makers. The algorithms are implemented to find the *hierarchically* and *strategically optimistic* solutions, respectively. The first instance is an illustrative example that we use to provide insight about the performance of the algorithms; the second instance is an industrial example of practical interest for the air separation industry.

Instance 1. Illustrative instance of trilevel capacity planning

This example considers one existing plant ($\mathcal{L}_1$) and one potential plant ($\mathcal{L}_2$) controlled by the leader, as well as one existing plant ($\mathcal{C}_1$) and one potential plant ($\mathcal{C}_2$) controlled by the competition. The market comprises four customers (M_j) with demands for a single commodity. The planning problem has a horizon of 20 time periods, in which the plants are allowed to expand in periods 1, 5, 9, 13, and 17. A scheme representing the location of plants and markets is presented in Fig. 5; the parameters of the instance are given in Tables 1, 2, and 3.

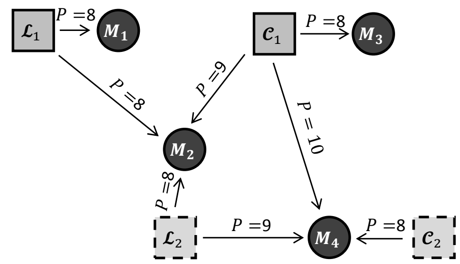

Figure 5: Network of plants and markets in Instance 1

Time (t)	Customer (j)			
	$D_{t,1}$	$D_{t,2}$	$D_{t,3}$	$D_{t,4}$
1-4	3.75	0	3	10
5-8	3.75	0	3	10
9-12	3.75	8	3	10
13-16	3.75	10	3	10
17-20	3.75	10	3	10

Table 1: Market demands in Instance 1

Costumer	Plant (i)			
(j)	$P_{t,\mathcal{L}_1,j}$	$P_{t,\mathcal{L}_2,j}$	$P_{t,\mathcal{C}_1,j}$	$P_{t,\mathcal{C}_2,j}$
M_1	8	8	17	17
M_2	8	8	9	17
M_3	17	17	8	17
M_4	9	9	10	8

Table 2: Selling prices in Instance 1

Parameter	Plant (i)			
	$\mathcal{L}_1$	$\mathcal{L}_2$	$\mathcal{C}_1$	$\mathcal{C}_2$
$A_{t,i}$	-	0	-	0
$B_{t,i}$	15	15	15	15
$E_{t,i}$	110	110	110	110
$F_{t,i}$	3	3	2	4
$G_{t,i,1}$	1	10	10	10
$G_{t,i,2}$	10	1	2	10
$G_{t,i,3}$	10	10	1	10
$G_{t,i,4}$	10	2	3	1
$C_{0,i}$	3.75	0	30	0
$H_{t,i}$	30	30	30	30

Table 3: Cost parameters and initial capacities in Instance 1

The optimal expansion strategy for the leader comprises expanding plant $\mathcal{L}_2$ at time 9 to capture the demand from M_2. The rational reaction of the competition is to expand plant $\mathcal{C}_2$ at time 9 to maintain M_4 by offering a lower price than the leader. The elements of the objective functions at the trilevel optimal solution are presented in Table 4.

Instance 1 has been designed such that Algorithm 1 and Algorithm 2 find exactly the same solution at every iteration. This is possible because the *hierarchically* and *strategically optimistic* solutions coincide (no degeneracy) and because the different restrictions on the HP in the 2 algorithms happen to have the same solution at every step of this instance. The convergence of the upper and lower bounds for both algorithms can be observed in Fig. 6.

Both algorithms were implemented in GAMS 24.4.1 and the optimization

Element of objective function	Leader	Competition
Income from sales [M$]:	1,496	2,240
Expansion cost [M$]:	110	110
Maintenance cost [M$]:	480	480
Production cost [M$]:	561	760
Transportation cost [M$]:	187	420
Total NPV [M$]:	158	470

Table 4: Optimal objective values in Instance 1

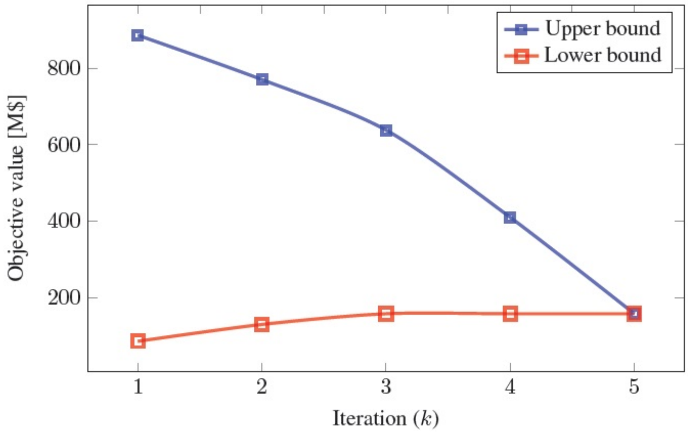

Figure 6: Convergence of Algorithms 1 and 2 in Instance 1

problems were solved using GUROBI 6.0.0 on an Intel Core i7 CPU 2.93 Ghz processor with 4 GB of RAM. Table 5 presents the computational statistics for problems HP^k of Algorithm 1 and MP^k of Algorithm 2 in the first and last iterations. We observe that both problems have the same the number of continuous variables and constraints in the first iteration, but they grow much faster in Algorithm 2 than in Algortihm 1; on the other hand, Algorithm 1 has a modest increase in the number of binary variables. Our analysis indicates that instances for which both algorithms explore the solution space in the same order, can be solve faster with Algorithm 1 because the complexity of iterations increases at a lower rate.

	First iteration	Final iteration	
Problem	HP^k & MP^k	HP^k	MP^k
Constraints:	1,015	1,035	3,596
Continuous variables:	835	835	3,331
Binary variables:	120	128	120
CPU time [s]:	2	5	9

Table 5: Computational statistics for Algorithms 1 and 2 in Instance 1

Instance 2. Industrial instance

This example is based on the instance Mid-size 1 presented by Garcia-Herreros et al. [10]; we extend the problem by considering expansions in the plants controlled by the competition. The problem comprises the production and distribution of one product to 15 customers. Initially, the leader has three plants with initial capacities equal to 27,000 ton/period, 13,500 ton/period, and 31,500 ton/period. Additionally, the leader considers the possibility of opening a new plant at a candidate location. As for the competition, they control three plants with an initial capacity of 22,500 ton/period, 45,000 ton/period and 49,500 ton/period; the competition also has a candidate location for a new plant. The investment decisions are evaluated over a time horizon of 5 years divided in 20 time periods; all producers are allowed to expand only every fourth time-period.

Selling prices and market demands follow an increasing trend during the

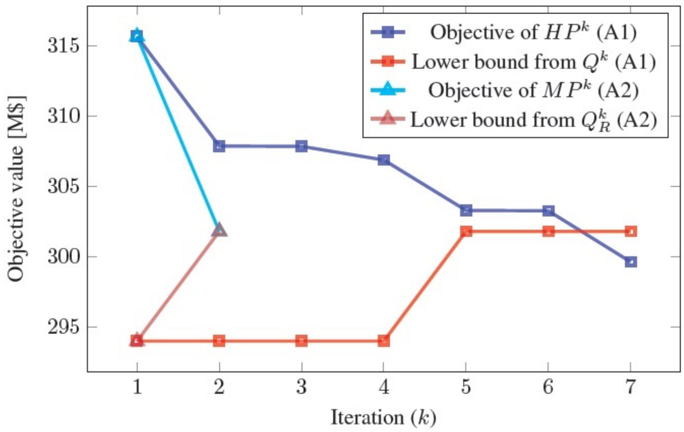

Figure 7: Convergence of Algorithm 1 (A1) and Algorithm 2 (A2) in Instance 2

time horizon. Investment and maintenance costs grow in time to adjust for inflation. The costs of production also have an increasing trend but exhibit a seasonal variation that relate to electricity prices. The exact data for this industrial instance can be found in the Supplementary material.

The same computational setup described above is used. In this industrial instance, Algorithm 2 is very efficient; it only needs two iterations to find the trilevel optimal solution, while Algorithm 1 requires 7 iterations. Both algorithms find the same solution because the *hierarchically* and *strategically optimistic* solutions coincide. The convergence of the upper and lower bounds to the optimal solution (M$302) can be observed in Fig. 7.

Table 6 presents the computational statistics for problems HP^k of Algorithm 1 and MP^k of Algorithm 2 in the first and last iterations. We observe that the number of continuous variables and constraints grows very quickly for problem MP^k, even though the number of binary variables stays constant. The total time required by Algorithm 1 to solve the instance is 46 s, in contrast with Algorithm 2 that only takes 8 s. This instance shows the advantage of Algorithm 2 for problems that are solved in few iterations.

Problem	First iteration HP^k & MP^k	Final iteration HP^k	MP^k
Constraints:	4,174	4,229	7,620
Continuous variables:	3,835	3,835	7,259
Binary variables:	240	264	240
Solution time [s]:	2	12	6

Table 6: Computational statistics for Algorithms 1 and 2 in Instance 2

The optimal investment plan for the leader in this industrial instance is to expand plant $\mathcal{L}_3$ at time 1 and 5. The rational reaction of the competition is not to expand at all. The elements of the objective functions at the trilevel optimal solution are presented in Table 7.

The optimal capacity expansion plan for the trilevel formulation differs from the results reported by Garcia-Herreros et al. [10] for the bilevel formulation in which the competitions cannot expand. Even though the optimal expansion strategy for the competition is not to expand, considering the competition as

Element of objective function	Leader	Competition
Income from sales [M$]:	816	504
Investment in new plants [M$]:	0	0
Expansion cost [M$]:	56	0
Maintenance cost [M$]:	94	97
Production cost [M$]:	288	171
Transportation cost [M$]:	76	41
Total NPV [MM$]:	302	195

Table 7: Optimal objective values in Instance 2

a rational decision-maker changes the optimal plan of the leader. This result exposes some of the counter-intuitive mechanisms present in multilevel optimization problems. In this particular instance, if the leader implements the bilevel optimal plan [10] prescribing three expansions instead of two, the rational reaction of the competition is to expand plant $\mathcal{C}_1$ at time 1. This expansion plans would produce a NPV for the leader equal to M$ 294, which is 2.5% lower than the trilevel optimal solution (M$302). This measure of regret illustrates the value of obtaining the trilevel optimal solution in comparison to a bilevel formulation that assumes static competitors.

9. Conclusions and future work

For the first time, a fully competitive model for the capacity planning problem has been formulated as a trilevel optimization. It allows simultaneously considering the conflicting interests of three rational decision-makers within a mathematical programming framework. We have also addressed for the first time the topic of degeneracy in multilevel decision problems. Our research found a void in definitions and models that induce ambiguity in the characterization of trilevel optimal solutions. We have introduced several extensions of the *optimistic* models from bilevel programming and we have provided algorithms that allow finding these different optimal solutions.

The proposed model belongs to a challenging class of mathematical problems: multilevel programming with integer variables in more than one level. The few general methods available to solve this type of problems are at an early stage. We have developed two problem specific solution methods that rely on different properties of the formulation. The examples show that none of the two algorithms strictly dominates the other in terms of performance, indicating that both are interesting approaches to solve this problem. The solutions obtained from the new formulation unveil complex interactions that are very difficult to predict. A significant improvement over previously proposed models is quantified in monetary terms for the industrial instances.

The type of problems that we have addressed are of interest in applications where discrete decisions are taken by different players. As the range of applications is expected to increase, we consider the generalization of the algorithms as an important direction for future research; additionally, efficiency and numerical stability of the algorithms can still improve. For the industrial application of the capacity expansion model, we believe that it is important to extend the model to include stochastic parameters.

Acknowledgments

The authors gratefully acknowledge the financial support received from the Fulbright program, and from Air Products & Chemicals through the Center for Advanced Process Decision-making (CAPD) at Carnegie Mellon University.

References

[1] G Anandalingam and Victor Apprey. Multi-level programming and conflict resolution. *European Journal of Operational Research*, 51(2):233–247, 1991.

[2] Jonathan F. Bard and James T. Moore. An algorithm for the discrete bilevel programming problem. *Naval Research Logistics*, 39:419–435, 1992.

[3] Jonathan F. Bard, John Plummer, and Jean Claude Sourie. A bilevel programming approach to determining tax credits for biofuel production. *European Journal of Operational Research*, 120:30–46, 2000.

[4] Wayne F Bialas and Mark H Karwan. Two-level linear programming. *Management Science*, 30(8):1004–1020, 1984.

[5] Suk-Gwon Chang and Bezalel Gavish. Telecommunications network topological design and capacity expansion: Formulations and algorithms. *Telecommunication Systems*, 1(1):99–131, 1993.

[6] Andrés Delgadillo, José Manuel Arroyo, and Natalia Alguacil. Analysis of electric grid interdiction with line switching. *Power Systems, IEEE Transactions on*, 25(2):633–641, 2010.

[7] Stephan Dempe, Boris S Mordukhovich, and Alain Bertrand Zemkoho. Necessary optimality conditions in pessimistic bilevel programming. *Optimization*, 63(4):505–533, 2014.

[8] S. T. DeNegre and T. K. Ralphs. A branch-and-cut algorithm for integer bilevel linear programs. *Operations Research/ Computer Science Interfaces Series*, 47:65–78, 2009.

[9] Nuno P. Faísca, Vivek Dua, Berç Rustem, Pedro M. Saraiva, and Efstratios N. Pistikopoulos. Parametric global optimisation for bilevel programming. *Journal of Global Optimization*, 38(4):609–623, 10 2006. doi: 10.1007/s10898-006-9100-6.

[10] Pablo Garcia-Herreros, Pratik Misra, Erdem Arslam, Sanjey Mehta, and Ignacio E. Grossmann. A duality-based approach for bilevel optimizatino of capacity expansion. In *12th International Symposium on Process Systems Engineering and 25th European Symposium on Computer Aided Process Engineering*, 2015.

[11] Fred Glover. Improved linear integer programming formulations of nonlinear integer problems. *Management Science*, 22:455–460, 1975.

[12] Ignacio E Grossmann. Advances in mathematical programming models for enterprise-wide optimization. *Computers & Chemical Engineering*, 47: 2–18, 2012.

[13] Zeynep H Gümüş and Christodoulos A Floudas. Global optimization of mixed-integer bilevel programming problems. *Computational Management Science*, 2(3):181–212, 2005.

[14] Robert G Jeroslow. The polynomial hierarchy and a simple model for competitive analysis. *Mathematical programming*, 32(2):146–164, 1985.

[15] Hanan Luss. Operations research and capacity expansion problems: A survey. *Operations Research*, 30(5):907–947, 1982.

[16] Christos T Maravelias and Ignacio E Grossmann. Simultaneous planning for new product development and batch manufacturing facilities. *Industrial & engineering chemistry research*, 40(26):6147–6164, 2001.

[17] Athanasios Migdalas. Bilevel programming in traffic planning: models, methods and challenge. *Journal of Global Optimization*, 7(4):381–405, 1995.

[18] Alexander Mitsos, George M Bollas, and Paul I Barton. Bilevel optimization formulation for parameter estimation in liquid–liquid phase equilibrium problems. *Chemical Engineering Science*, 64(3):548–559, 2009.

[19] James T Moore and Jonathan F Bard. The mixed integer linear bilevel programming problem. *Operations research*, 38(5):911–921, 1990.

[20] Frederic H Murphy and Yves Smeers. Generation capacity expansion in imperfectly competitive restructured electricity markets. *Operations research*, 53(4):646–661, 2005.

[21] J Cole Smith, Churlzu Lim, and A Alptekinoglu. Optimal mixed-integer programming and heuristic methods for a bilevel stackelberg product introduction game. *Naval Research Logistics*, 56(8):714–729, 2009.

[22] Susara A Van Den Heever and Ignacio E Grossmann. Disjunctive multiperiod optimization methods for design and planning of chemical process systems. *Computers & Chemical Engineering*, 23(8):1075–1095, 1999.

[23] Pan Xu and Lizhi Wang. An exact algorithm for the bilevel mixed integer linear programming problem under three simplifying assumptions. *Computers & operations research*, 41:309–318, 2014.

[24] Yiming Yao, Thomas Edmunds, Dimitri Papageorgiou, and Rogelio Alvarez. Trilevel optimization in power network defense. *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*, 37 (4):712–718, 2007.

[25] Bo Zeng and Yu An. Solving bilevel mixed integer program by reformulations and decomposition. *Optimization On-line*, 2014.