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 with the highest profit. Traditional formulations neglect the rational behavior of some external decision-makers by assuming either static competition or captive markets. We propose a mathematical programming formulation with three levels of decision-makers to capture the dynamics of duopolistic 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 superior in terms of performance.

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

1. Introduction

Industrial and manufacturing companies rely on capacity expansion models to plan the investments that allow them to satisfy future demands. In this sector, the proximity of producers to customers increases supply reliability and reduces transportation costs, which provides a key competitive advantage [22, 7]. This feature makes location and capacity planning a major strategic decision that impacts the market share obtained in an environment with rational customers. In this article, we study the case of two companies competing for the same market in a hierarchical framework, where the leading company first performs its capacity expansions, then the competition reacts with its own, and finally an inelastic market observes all available capacities and minimizes its total cost of supply. This is a common situation in the market of industrial gases, which has been the motivation for our research; we apply our model and algorithms to case studies from the air separation industry.

The purpose of this paper is threefold: to introduce a new model based on a trilevel mathematical programming formulation, to elaborate on the characteristics of general trilevel optimization problems, and to propose two novel

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

algorithms exploiting the problem structure. The remaining part of this section, presents a literature review of the available hierarchical capacity planning models, of general multilevel programs, and of the advances in algorithms to solve them; each part specifies our contribution to the corresponding body of work.

1.1. Industrial capacity planning

Capacity planning is a widely studied problem in areas requiring large capital investments whose feasibility, effectiveness, and profitability can only be assessed in a long time horizon. Some examples are electrical power supply [31], communication networks [8], and Enterprise-Wide Optimization [17]. When dealing with industrial expansions, a considerable difficulty that arises is the discrete nature of the decisions, as it is only possible to build new plants or production lines of a specific size. If we are additionally interested in developing a robust expansion plan that anticipates the rational reaction of the competitors and potential costumers, it is necessary to adopt a game theoretic approach.

A foundational model in the location literature is the discrete $(r|p)$-centroid problem. It investigates what are the best locations in a network for the leading company to place p new facilities, knowing that the competitors will react by choosing r other locations [19]. There has been many applications of sequential competitive location on networks [25]. Recent advances to solve this problem include branch-and-cut procedures [33] or exact iterative local search methods [1]. Nevertheless, these models do not include the cost of expansion, maintenance or production, and the strategies do not contemplate any time-horizon. These assumptions, that greatly simplify the problem, are not considered in our model. We also allow the number of expansions to be a variable in the optimization problems solved by the leader and by the follower, instead of being predefined. Furthermore, most of the models available in the literature only consider uncapacitated facilities; therefore, costumers can always be served by their prefered facility without considering the influence of other costumers [22].

The work presented by Karakitsiou and Migdalas [23] recognizes the failure to capture a rational market in previous models. They propose a more realistic model using Nash equilibrium and a demand curve that is assigned to the costumers. In their model, only the leader expands and all competitors have fixed capacity; they only compete by deciding their production quantities. Models with similar characteristics have already been proposed [35] and extensions also consider shipping dynamics or multi-period time horizon [14, 28]. These Stackelberg-Nash-Cournot formulations usually yield Equilibrium Problems with Equilibrium Constraints and are tackled by complete enumeration or heuristics [7]. In particular, given that most solution approaches are based on KKT reformulation of the lower level, this framework does not allow the competitors to have discrete decisions.

In markets like the one of industrial gases, the demands are highly inelastic: costumers need to fulfill their exact demand and prices are fixed in advance from each supplier. Hence, instead of having demand curves and simple market clearing equations, a more realistic model must have a market that minimizes its total cost of supply as presented by Garcia-Herreros et al. [15]. The motivation of our research is to extend that work to find strategies that foresee not only the behavior of rational markets, but also the reaction of a rational competition and its resulting expansion plan. We will show in our numerical examples that considering a static competitor can lead to expansion plans that yield severe losses.

Similarly to the model presented by Garcia-Herreros et al. [15], we include several time steps with deterministic forecasted demands. Hence, we also solve the expansion planning problem and properly evaluate the investment returns. Given that the construction of new plants is publicly announced to the market, we consider perfect information among all players. Finally, industrial producers often compete regionally with only one major company, and one of them can be considered a follower that observes the expansion plan of the leader and then decides its own strategy. Therefore, our model represents a hierarchical duopoly. In contrast to previous research [15], we consider an additional sequential decision maker, resulting in a three-level Stackelberg game. The natural formulation is a trilevel optimization with integer variables controlled by the two upper levels.

Even Bilevel Linear Programs are NP-hard [21, 6] and Bilevel Mixed Integer Linear Programs (BMILP) are still considered an unsolved problem in Operations Research [9]. In the following, we review the corresponding literature upon which our studies of general trilevel programs are built in Section 3, as well as our algorithms in Section 5-6.

1.2. Multilevel Programming

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* (she), that first decides her strategy with perfect information of the criterion ruling the behavior of the lower-level player, called *follower* (he). Once the follower observes the leader's decision, he reacts according to his own interests. The potential to coordinate decision-making in decentralized systems using bilevel optimization has been recognized for decades [3]. Interesting bilevel programming models have been developed for traffic planning [27], optimal taxation of biofuels [5], parameter estimation [29], and product introduction [36].

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 [39]. However, the solution procedures for this formulation are problem specific [2] and there is scarce theoretical study of the general properties of trilevel optimization problems [20].

In particular, there is no work dealing with degenerate (with multiple optima) solutions in multilevel programs. This topic has attracted considerable attention in the bilevel case, where the *Optimistic* and *Pessimistic* solutions have been defined to specify whether the ties are broken in favor of the leader or against her. It is a very active area of research [37, 11, 40] and some algorithms are designed to find one type of solution or the other [24, 41]. Our research discusses new ideas about how degeneracy affects multilevel problems and our algorithms addresses the issue of degeneracy, so we do not assume the optimal solution to be unique unlike in most of the literature.

1.3. Solution approaches for BMILPs

If the last player is represented by an LP (or a convex program) a trilevel problem can be reformulated as a bilevel problem by replacing the third-level by its optimality conditions [15]. In the bilevel reformulation, the second level models the capacity expansion of the competitor 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.

The numerical solution of BMILPs has been receiving increasing attention, but the existing literature only considers academic examples with a few discrete variables. The first Branch-&-Bound algorithm was developed by Moore and Bard [30]; 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 iteratively solving a MILP after fixing the decision of the leader [4]; in the worst case, both algorithms conduct an exhaustive exploration of the leader's decision space. DeNegre and Ralphs [12] derived a locally valid cut that can be added to the latter Branch-&-Bound procedure; however, these cuts tend to be weak in problems with parameters of different magnitudes or with non-integer coefficients.

The framework proposed by Gümüs and Floudas [18] 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. [13] 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 very 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 [38] proposed a general spatial Branch-&-Bound search that splits the variables of the leader in polyhedral sets characterizing the decisions of the leader that share the same optimal reaction of the follower. Also, Zeng and An [42] 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. Other state-of-the-art methods require special assumptions that do not hold in our case, like restricting the influence of the follower in the leader problem to be only through his objective value [32].

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 competitor 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 and propose a mathematical formulation. Section 3 explores the implications of degeneracy in trilevel optimization problems. In Section 4, 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 5 and 6. In Section 7, we illustrate the implementation of the algorithms on two instances of the capac-

MODEL VARIABLES		RELATED PARAMETERS	
Facility status:		**Fixed costs & initial values:**	
$x_{t,i} \in \{0,1\}$	Plant i expands at t	$E_{t,i}$	Cost of expanding an open plant
$w_{t,i} \in \{0,1\}$	Plant i is open at t	$B_{t,i}$	Maintenance cost of open plants
$v_{t,i} \in \{0,1\}$	Plant i is built at t	$A_{t,i}$	Cost of opening a new plant
		$V_{0,i}$	Plant i is initially open (binary)
$c_{t,i} \in \mathbb{R}$	Capacity of i at t	$C_{0,i}$	Initial capacity at i
		H_i	Possible capacity increment at i
$\mathscr{C}_{\mathcal{L},t} \in \mathbb{R}$	Total leader cap. at t		
Costumer related:		**Unit costs:**	
$y_{t,i,j} \in \mathbb{R}$	Demand assigned to i from costumer j at t	$P_{t,i,j}$	Unit total price $(S_{t,i} + G_{t,i,j})$
		$S_{t,i}$	Unit selling price (with margin)
$\mathscr{D}_{\mathcal{L},t} \in \mathbb{R}$	Total demand assigned to leader at time t	$G_{t,i,j}$	Unit transportation cost
		$F_{t,i,j}$	Unit production cost

Table 1: Variables and parameters of the model

ity planning problem. Finally, Section 8 reviews the novelty of this work and indicates directions for future research.

2. Problem and model description

We address the problem of two industrial producers sequentially deciding their capacity expansions to maximize their individual profit, subject to the demand assignments controlled by a rational market minimizing its total cost of supply. One company is called the *leader*, as she plans and announces its expansion strategy first. The other company is called the *follower*, as he observes and then reacts with his own expansion plan.

The set I of possible plant locations is partitioned among the two firms, yielding the sets $I_{\mathcal{L}}$ and $I_{\mathcal{C}}$ respectively. In every time step ($t \in T$) they can act on their locations by building new facilities or expanding the capacity of existing facilities, always in discrete increments (H_i). Using the variables and parameters presented in Tab. 1 and the description therein, the capacity management constraints can be described by Eqn. ($1_{\mathcal{L}}$)-($5_{\mathcal{L}}$). Apart from the hierarchical distinction, both firms are subject to the same model.

$$w_{t,i} = V_{0,i} + \sum_{t' \in T_t^-} v_{t',i} \qquad \forall \ t \in T, i \in I_{\mathcal{L}} \tag{$1_{\mathcal{L}}$}$$

$$x_{t,i} \leq w_{t,i} \qquad \forall \ t \in T, i \in I_{\mathcal{L}} \tag{$2_{\mathcal{L}}$}$$

$$c_{t,i} = C_{0,i} + \sum_{t' \in T_t^-} H_i x_{t',i} \qquad \forall \ t \in T, i \in I_{\mathcal{L}} \tag{$3_{\mathcal{L}}$}$$

$$c_{t,i} \in \mathbb{R}^+ \qquad \forall \ t \in T, i \in I_{\mathcal{L}} \tag{$4_{\mathcal{L}}$}$$

$$v_{t,i}, \ w_{t,i}, \ x_{t,i} \in \{0,1\} \qquad \forall \ t \in T, i \in I_{\mathcal{L}} \tag{$5_{\mathcal{L}}$}$$

The letter subscript in the equation number indicates whether they refer to the leader $\mathcal{L}$ or to the competitor $\mathcal{C}$. Also, any variable with superscript will denote the value it took when solving a particular optimization problem specified by the superscript or the context. Other used shorthands include $x_{\mathcal{L}}$ to denote the vector of all leader variables $[x_{t,i}, v_{t,i}, w_{t,i}, c_{t,i}]_{t\in T, i\in I_{\mathcal{L}}}$, $X_{\mathcal{L}}$ represents its domain, and T_t^- and T_t^+ denote all times before or after t respectively. Following again the notation from Tab. 1, the objective function of the firms is defined as the Net Present Value ($NPV_{\mathcal{L}}$ or $NPV_{\mathcal{C}}$) given by Eqn. ($6_{\mathcal{L}}$). The first sum is the total income of the firm. The second are the fixed costs related to capacity investments and maintenance. Finally the third includes the variable costs. Any desired discount factor is included in the time-varying parameters.

$$\begin{aligned} NPV_{\mathcal{L}}(x_{\mathcal{L}}, y) = & \sum_{t\in T}\sum_{i\in I_{\mathcal{L}}}\sum_{j\in J} P_{t,i,j} y_{t,i,j} \\ & - \sum_{t\in T}\sum_{i\in I_{\mathcal{L}}} (A_{t,i} v_{t,i} + B_{t,i} w_{t,i} + E_{t,i} x_{t,i}) \\ & - \sum_{t\in T}\sum_{i\in I_{\mathcal{L}}}\sum_{j\in J} (F_{t,i} + G_{t,i,j})\, y_{t,i,j} \end{aligned} \tag{$6_{\mathcal{L}}$}$$

Note that the only coupling between both companies comes from the demands $y_{t,i,j}$ each costumer $j \in J$ assigns to every facility (i.e. firms do not directly act on the opponent's plants, as in Defender-Attacker problems [34]). Given our rational market model, these assignments must be a solution of the optimization problem solved by the market to minimize its total cost $P(y)$ -Eqn. (7)- of satisfying its demand -Eqn. (8)- subject to the available capacities -Eqn. (9). The resulting optimization problem is called the market problem $M(c)$. Its solution vector of demand assignments y must lie in the region $\Psi_y(x_{\mathcal{L}}, x_{\mathcal{C}})$, called the market *Basic Rational Reaction* set as described in Sec. 3.

$$\Psi_y(x_{\mathcal{L}}, x_{\mathcal{C}}) = \arg\min_{y_{t,i,j}\geq 0} \quad P(y) = \sum_{t\in T}\sum_{i\in I}\sum_{j\in J} P_{t,i,j} y_{t,i,j} \tag{7}$$

$$\text{s.t.} \quad \sum_{i\in I} y_{t,i,j} = D_{t,j} \qquad \forall\, t \in T, j \in J \tag{8}$$

$$\sum_{j\in J} y_{t,i,j} \leq c_{t,i} \qquad \forall\, t \in T, i \in I \tag{9}$$

The only influence of the suppliers' decisions $(x_{\mathcal{L}}, x_{\mathcal{C}})$ is the right hand side of Eqn. (9): if a highly demanded and saturated plant increases its capacity, it will gain market share - and others will lose it. Therefore, it is the central competition point among suppliers and the motivation to expand well located plants.

Note that $M(c)$ is a linear transportation model; hence, it is always feasible as long as $\sum_{j\in J} D_{t,j} \leq \sum_{i\in I} c_{t,i}$ $\forall t \in T$. Given that we are mainly interested by the competitive dynamics arising in this hierarchical framework -and less by the companies being artificially forced to expand against their profit- we will consider that the initially installed capacity is always greater than the total demand. This can be done without loss of generality. If the two companies do not have that much installed capacity initially (possibly in old, far, unattractive plants), we can always consider a third party plant, with unlimited capacity but very high prices, so the market will keep maximal pressure on the plants of our two competing firms.

We also point out that our Stackelberg model is completely deterministic and hence we can also consider without loss of generality all decisions for all time steps to be taken at the beginning in "open loop" fashion [26].

Now we have all the pieces to finally state the full model in Eqns. (10)-(14). This is a trilevel program with integer constraints in the two upper level, which is at the very frontier of the area.

$$\text{“}\max_{x_\mathcal{L}}\text{”} \quad NPV_\mathcal{L}(x_\mathcal{L}, y) \tag{10}$$

$$\text{s.t.} \quad (1_\mathcal{L}) - (5_\mathcal{L}) \tag{11}$$

$$x_\mathcal{C} \in \text{“}\arg\max_{x_\mathcal{C}}\text{”} \quad NPV_\mathcal{C}(x_\mathcal{C}, y) \tag{12}$$

$$\text{s.t.} \quad (1_\mathcal{C}) - (5_\mathcal{C}) \tag{13}$$

$$y \in \Psi_y(x_\mathcal{L}, x_\mathcal{C}) \tag{14}$$

We have used quotation marks on the "max" and "arg max" operators to indicate that the model as stated is not yet well defined [40]. Each player only controls its own variables, and hence we need to specify how the lower level players will break their ties in case of having multiple globally optima solutions. In other words, if $\Psi_y(x_\mathcal{L}, x_\mathcal{C})$ or the "arg max" do not yield singletons, the perfect information assumption requires that all players know which particular element of these sets will be selected. This topic has never been thoroughly studied in the trilevel case, and it is the subject of our next section.

3. Multilevel programming and degeneracy

For any non-strictly convex minimization problem, there might be several alternative global optima. In the multilevel case, if for at least one decision of the leader, a lower level program exhibits this phenomena, the problem is called *degenerate*. This ambiguity can have a profound impact on the final solution of the multilevel problem as two different optimal responses of the follower will give him the same objective value, but might produce opposed effects on the higher levels. It is therefore critical to specify how this ambiguities will be resolved, what are the model interpretation of them, and to develop an algorithm that converges to the desired solution. The question has only been addressed in the bilevel case, so in this section we start by reviewing some related definitions and then we extend them to the trilevel case.

3.1. Review on bilevel optimization

When there are no degeneracies in the lower level, we can write a general bilevel problem as in Eqns. (15)-(17), where $x \in X$ are the variables controlled by the leader and $y \in Y$ are the variables controlled by the follower.

$$\max_{x \in X} \quad f_1(x, y) \tag{15}$$

$$\text{s.t.} \quad g_1(x, y) \leq 0 \tag{16}$$

$$y \in \Psi(x) = \arg\max_{y \in Y} \{f_2(x, y) : g_2(x, y) \leq 0\} \tag{17}$$

Definition 1. Given the bilevel program presented in Eqns. (15)-(17), let

- Ω be the *Bilevel Constraint Region*:

$$\Omega = \{(x, y) \in X \times Y : g_1(x, y) \le 0,\ g_2(x, y) \le 0\} \tag{18}$$

- $\Omega_y(x)$ be the *Follower Constraint Region* for a fixed $x \in X$:

$$\Omega_y(x) = \{y \in Y : g_2(x, y) \le 0\} \tag{19}$$

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

$$\Psi(x) = \arg\max_{y \in Y} \{f_2(x, y) : g_2(x, y) \le 0\} \tag{20}$$

If $\Psi(x)$ is not a singleton, keeping the formulation as in Eqns. (15)-(17) suggests that the leader also can choose its most favorable y, as long as it is optimal for the follower. This is called the *Optimistic* formulation and it is the most common approach in the literature. Formally, we could either write $y \in Y$ as a variable under the first max or replace $\Psi(x)$ in Eqn. (17) by the *Bilevel Optimistic Reaction set* defined by Eqn. (21). Note that it is irrelevant whether this set is now a singleton, given that any point $y \in \Psi_{\mathcal{L}}(x) \subset \Psi(x)$ will grant the same objective value to both players.

$$\Psi_{\mathcal{L}}(x) = \arg\max_{y \in \Psi(x)} \{f_1(x, y)\} \tag{21}$$

Apart from the mathematical necessity to address the degeneracy issue, it is also a modeling concern. The *Optimistic* approach is adequate when some degree of collaboration, or "ϵ-influence", is allowed between levels -as side-payments from the leader for example. This interpretation and some reformulation aspects described in Sec. 4 make it the most commonly widespread approach. However, there is an increasing interest on extending the treatment of degeneracies, studying other alternative formulations. 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 [11]. These alternative models are considered harder to solve than the *Optimistic* approach, although new algorithms are weakening this condition [41].

The idea outlined in Eqn. (21) of reducing the *Basic Rational Reaction* set so that it captures the chosen tie breaking strategy, is very powerful and will be extended in the next subsection also to trilevel problems, allowing to formally define the possible "*Optimistic*" formulations.

3.2. Degeneracy in trilevel programming

In order to comply with the perfect information assumption, the decision criteria under degeneracy must be completely specified at all levels. In this way, decision-makers that are hierarchically higher can calculate the response of the lower levels. Unfortunately, the definition of bilevel *Optimistic* solution does not extend trivially to trilevel programs. Next, we propose three meaningful extensions according to the order in which the objective functions of the upper-levels are favored: *Sequentially Optimistic*, *Hierarchically Optimistic* and *Strategically Optimistic*. The analysis herein is completely general, but for better use in Sec. 4, we have used the capacity expansion variables and objectives introduced in Sec. 2.

Definition 2. 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.

Definition 3. 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.

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 4. 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.

For a simple case where these definitions yield different solutions, refer to Example 1 at the end of this section. To characterize mathematically each of these solutions, as was outlined in the bilevel case, we will replace the *Basic Rational Reaction set* by the appropriate subset. First let us introduce some notation. The definitions of *Trilevel Constraint Region* Ω and *Third Level Constraint Region* $\Omega_y(x_\mathcal{L}, x_\mathcal{C})$ can be extended directly from the bilevel case in Def. 1. The *Rational Reaction* sets require some more work. Their use will become clear in Prop. 1-3, where they enforce the different degeneracy resolution criteria.

Definition 5. In a trilevel program we define the following sets:

- The *Basic Rational Reaction* set of the third level:

$$\Psi_y(x_\mathcal{L}, x_\mathcal{C}) = \underset{y \in \Omega_y(x_\mathcal{L}, x_\mathcal{C})}{\arg\min} \{P(y)\} \tag{22}$$

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

$$\Psi_{\mathcal{C},y}(x_\mathcal{L}, x_\mathcal{C}) = \underset{y \in \Psi_y(x_\mathcal{L}, x_\mathcal{C})}{\arg\max} \{NPV_\mathcal{C}(x_\mathcal{C}, y)\} \tag{23}$$

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

$$\Psi_{\mathcal{L},y}(x_\mathcal{L}, x_\mathcal{C}) = \underset{y \in \Psi_y(x_\mathcal{L}, x_\mathcal{C})}{\arg\max} \{NPV_\mathcal{L}(x_\mathcal{L}, y)\} \tag{24}$$

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

$$\Psi_{x_\mathcal{C}}^{Seq}(x_\mathcal{L}) = \underset{x_\mathcal{C} \in X_\mathcal{C}(x_\mathcal{L})}{\arg\max} \{NPV_\mathcal{C}(x_\mathcal{L}, x_\mathcal{C}, y) : y \in \Psi_{\mathcal{C},y}(x_\mathcal{L}, x_\mathcal{C})\} \tag{25}$$

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

$$\Psi_{x_\mathcal{C}}^{Hie}(x_\mathcal{L}) = \underset{x_\mathcal{C} \in X_\mathcal{C}(x_\mathcal{L})}{\arg\max} \{NPV_\mathcal{C}(x_\mathcal{C}, y) : y \in \Psi_{\mathcal{L},y}(x_\mathcal{L}, x_\mathcal{C})\} \tag{26}$$

- The *Strategically Optimistic* reaction set of the two followers is:

$$\begin{aligned}\Psi_{(x_\mathcal{C},y)}^{Str}(x_\mathcal{L}) = \Big\{ &(x_\mathcal{C}, y) \in X_\mathcal{C}(x_\mathcal{L}) \times \Psi_y(x_\mathcal{L}, x_\mathcal{C}) : \\ &\forall\, \tilde{x}_\mathcal{C} \in X_\mathcal{C}(x_\mathcal{L}), \exists\, \tilde{y} \in \Psi_y(x_\mathcal{L}, \tilde{x}_\mathcal{C}) : \\ &NPV_\mathcal{C}(x_\mathcal{C}, y) \geq NPV_\mathcal{C}(\tilde{x}_\mathcal{C}, \tilde{y}) \Big\}\end{aligned} \tag{27}$$

Next, we are interested on precisely describing all points $(x_{\mathcal{L}}, x_{\mathcal{C}}, y)$ that satisfy each of the Defs. 2-4. Apart from the theoretical interest, we point out that most algorithms for multilevel problems essentially solve single level problems of the type of Eqns. (15)-(17), somehow enforcing the lower level variables to be in their *Rational Reaction* set. Hence, it is critical to constrain enough the variables of the followers so that their control can be given to the leader without waving their rational behavior. In the following, we state what are these constraints in the case of the three trilevel *Optimistic* solution types from Defs. 2-4.

Proposition 1. *The set of Sequentially Optimistic solutions is:*

$$\arg\max_{(x_{\mathcal{L}}, x_{\mathcal{C}}, y)} \left\{ NPV_{\mathcal{L}}(x_{\mathcal{L}}, y): \ x_{\mathcal{L}} \in X_{\mathcal{L}}, \ x_{\mathcal{C}} \in \Psi^{Seq}_{x_{\mathcal{C}}}(x_{\mathcal{L}}), \ y \in \Psi_{\mathcal{C},y}(x_{\mathcal{L}}, x_{\mathcal{C}}) \right\} \quad (28)$$

Proof. The market first breaks its ties in favor of the competitor, as imposed by $y \in \Psi^{Seq}_{\mathcal{C},y}(x_{\mathcal{L}}, x_{\mathcal{C}})$. The competitor is aware of it and plans according to the definition of $\Psi^{Seq}_{x_{\mathcal{C}}}$ in Eqn. (25). Note that in fact this definition could simply have $y \in \Psi_y(x_{\mathcal{L}}, x_{\mathcal{C}})$ instead of $y \in \Psi_{\mathcal{C},y}(x_{\mathcal{L}}, x_{\mathcal{C}})$, given that the objective function being optimized is $NPV_{\mathcal{C}}$. □

Proposition 2. *The set of Hierarchically Optimistic solutions is:*

$$\arg\max_{(x_{\mathcal{L}}, x_{\mathcal{C}}, y)} \left\{ NPV_{\mathcal{L}}(x_{\mathcal{L}}, y): \ x_{\mathcal{L}} \in X_{\mathcal{L}}, \ x_{\mathcal{C}} \in \Psi^{Hie}_{x_{\mathcal{C}}}(x_{\mathcal{L}}), \ y \in \Psi_{\mathcal{L},y}(x_{\mathcal{L}}, x_{\mathcal{C}}) \right\} \quad (29)$$

Proof. Again, the market is imposed to always break its ties in favor of the leader first, and the competitor plans accordingly. □

Proposition 3. *The set of Strategically Optimistic solutions is:*

$$\arg\max_{(x_{\mathcal{L}}, x_{\mathcal{C}}, y)} \left\{ NPV_{\mathcal{L}}(x_{\mathcal{L}}, x_{\mathcal{C}}, y): \ x_{\mathcal{L}} \in X_{\mathcal{L}}, \ (x_{\mathcal{C}}, y) \in \Psi^{Str}_{(x_{\mathcal{C}}, y)}(x_{\mathcal{L}}) \right\} \quad (30)$$

Proof. The idea behind the *Strategically Optimistic* model is that the resolution of the degeneracy of the third level is fully controlled by the leader. And the competitor knows it. More precisely, for each possible reaction of the competitor $\tilde{x}_{\mathcal{C}} \in X_{\mathcal{C}}(x_{\mathcal{L}})$, the leader could force the market to take a $\tilde{y} \in \Psi_y(x_{\mathcal{L}}, \tilde{x}_{\mathcal{C}})$ that gives him the lowest $NPV_{\mathcal{C}}(\tilde{x}_{\mathcal{C}}, \tilde{y})$. Therefore, the competitor is willing to take any strategy $x_{\mathcal{C}}$ as long as the leader commits to steer the market decision y to a point granting him anything higher than the best $NPV_{\mathcal{C}}(\tilde{x}_{\mathcal{C}}, \tilde{y})$ he could achieve having the market completley against him. □

Example 1. We illustrate how all the different *Optimistic* solution types from Defs. 2-4 might not coincide. Consider the case depicted in Fig. 1, where for a fixed decision of the leader $x^1_{\mathcal{L}}$, the second level has two feasible startegies $x^1_{\mathcal{C}}$, $x^2_{\mathcal{C}}$, and that in both cases the third level is degenerate and has two possible optimal decisions, $\{y_A, \ y_B\}$ and $\{y_C, \ y_D\}$ respectively. The objective values of the leader $NPV_{\mathcal{L}}$ and the follower $NPV_{\mathcal{C}}$ for each case are explicited therein. According to them, outcome D is the optimal solution under the *Sequentially Optimistic* model because degeneracy in the third level favors the objective of the competition ($NPV_{\mathcal{L}} = 200$, $NPV_{\mathcal{C}} = 400$). Under the *Hierarchically Optimistic* model, the optimal solution of the problem is given by outcome C

($NPV_{\mathcal{L}} = 100$, $NPV_{\mathcal{C}} = 200$). This result is counter-intuitive because resolution of third-level degeneracy locally favors the first level, but it forces the second-level to avoid $x^1_{\mathcal{C}}$ and select instead $x^2_{\mathcal{C}}$, which is detrimental for the leader. The *Strategically Optimistic* solution is outcome A ($NPV_{\mathcal{L}} = 300$, $NPV_{\mathcal{C}} = 300$) and in general it is the best the leader can achieve in a trilevel setup. Outcome B will never happen under any degeneracy resolution model because the competitor will never accept it (he will rather have any of the outcomes associated with $x^2_{\mathcal{C}}$).

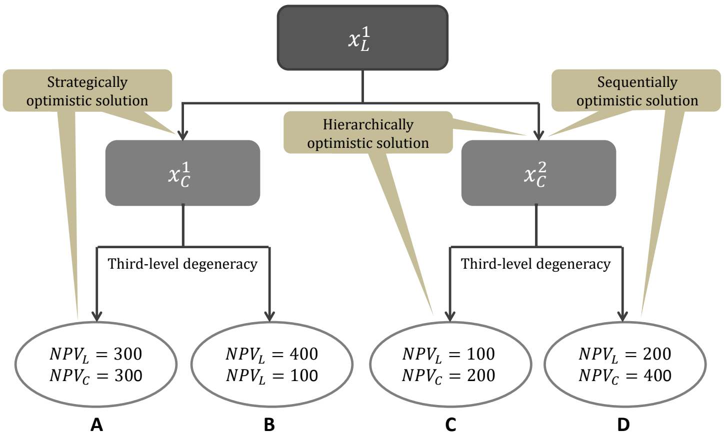

Figure 1: The *Strategically Optimistic* model is the most beneficial for the first level. The *Sequentially Optimistic* is the most beneficial for the second level. The *Hierarchical* may vary.

Depending on the application, each type of *Optimistic* solution has a different interpretation that the modeler needs to be aware of (also to pick the right algorithm). In the particular case of capacity expansion, the *Sequential* approach implies that although we consider undifferentiated products, in case of a tie for the market the competitor offers a slightly better service. The *Hierarchical* is the exact reverse, where inevitably the market would have a preference for the leader's products in case of a tie. Finally, the *Strategic* implies an additional control of the leader over the market, where she is able to "ϵ" modify the perception of the market over the products. 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.

4. Three solution paradigms applied to trilevel capacity planning

In this section we recall three concepts from the bilevel literature and outline how they will be useful to solve the trilevel capacity planning problem. First we will define the *Inducible Region* and how for easy problems, like our third level LP, it can be found with reformulations. Then, we introduce the *High Point* relaxation and how it can be tightened to yield improving upper bounds. Finally, we study the leader *Stability Regions* of our problem and define cuts to eliminate them, allowing to accelerate a search over the leader decision space. All the results here will be used to construct the algorithms in Sec. 5-6 and prove their convergence to a specific *Optimistic* trilevel solution.

4.1. Inducible Region and reformulations

In bilevel programming, Eqn. (31) defines the *Inducible Region* (IR) characterized by the feasible upper-level decisions and their corresponding rational response in the lower level problem. Therefore, the IR is the feasible region of its bilevel program.

$$IR = \big\{(x, y) : x \in X, y \in \Psi(x)\big\} \tag{31}$$

Any degeneration-resolution can be conveyed in $\Psi(x)$. And using the sets in Def. 5, this definition can be extended to the trilevel case. An explicit description of the IR would allow solving the problem as a single-level program over it. Unfortunately this region is usually non-convex, non-connected, and in general very hard to describe because of the constraint $y \in \Psi(x)$. In the simpler case of having a bilevel problem with an LP in the lower level, reformulation techniques can be used to replace the lower level optimization by its optimality constraints, hence exactly imposing $y \in \Psi(x)$. In the following, we detail how to apply this to reformulate our market problem $M(c)$.

The most common approach to reformulate a bilevel problem with a convex lower-level is to replace the inner program by its Karush-Kuhn-Tucker (KKT) optimality conditions. However, in linear lower-level problems with inequality constraints, the KKT approach might be ineffective because it requires the addition of many complementarity constraints. To reformulate our rational market $M(c)$ described in Eqns. (7)-(9), the duality-based approach described by Garcia-Herreros et al. [15] is better suited because it does not require adding 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 $M(c)$ is described by Eqns. (32)-(36).

$$\sum_{t\in T}\sum_{i\in I}\sum_{j\in J} P_{t,i,j} y_{t,i,j} = \sum_{t\in T}\left[\sum_{j\in J} D_{t,j}\lambda_{t,j} - \sum_{i\in I} c_{t,i}\mu_{t,i}\right] \tag{32}$$

$$\sum_{j\in J} y_{t,i,j} \leq c_{t,i} \qquad \forall\ t \in T, i \in I \tag{33}$$

$$\sum_{i\in I} y_{t,i,j} = D_{t,j} \qquad \forall\ t \in T, j \in J \tag{34}$$

$$\lambda_{t,j} - \mu_{t,i} \leq P_{t,i,j} \qquad \forall\ t \in T, i \in I, j \in J \tag{35}$$

$$y_{t,i,j}\ ,\ \mu_{t,i} \in \mathbb{R}^{+}; \quad \lambda_{t,j} \in \mathbb{R} \qquad \forall\ t \in T, i \in I, j \in J \tag{36}$$

where Eqn. (32) enforces strong duality and Eqn. (35) are the dual constraints corresponding to primal variables $y_{t,i,j}$. Dual variables associated to Eqn. (9) and Eqn. (8) are denoted by $\mu_{t,i} \in \mathbb{R}^{+}$ and $\lambda_{t,j,k} \in \mathbb{R}$, respectively.

It is important to note that Eqn. (32) contains bilinear terms in the product of upper-levels variables $c_{t,i}$ and dual variables $\mu_{t,i}$. Bilinear terms are non-convex; however, we can apply an exact linearization procedure [16] because variables $c_{t,i}$ only take discrete values according to Eqn. ($3_{\mathcal{L}}$) and ($3_{\mathcal{C}}$). Then, a big-M reformulation is applied by introducing new variables $u_{t,t',i}$. These are imposed to be the product of dual variables $\mu_{t,i}$ and expansion variables $x_{t',i}$. It can be shown that Eqn. (38) is sufficient when these equations are inside the maximization of an NPV (refer to [15] for more details).

$$\sum_{t\in T}\sum_{i\in I}\sum_{j\in J} P_{t,i,j} y_{t,i,j} = \sum_{t\in T}\left(\sum_{j\in J} D_{t,j}\lambda_{t,j} - \sum_{i\in I} C_{0,i}\mu_{t,i} - \sum_{i\in I}\sum_{t'\in T_t^-} H_i u_{t,t',i}\right) \tag{37}$$

$$u_{t,t',i} \geq \mu_{t,i} - M(1 - x_{t',i}) \qquad \forall\, t \in T, t' \in T_t^-, i \in I \tag{38}$$

$$u_{t,t',i} \in \mathbb{R}^+ \qquad \forall\, t \in T, t' \in T_t^-, i \in I \tag{39}$$

Hence, Eqn. (32) can be replaced by Eqns. (37)-(39) and the set of the market *Basic Rational Reaction set* $\Psi_y(x_{\mathcal{L}}, x_{\mathcal{C}})$ is completely described by Eqns. (33)-(39). Applied to our present trilevel problem, we can replace Eqn. (14) by these constraints, obtaining a Bilevel Mixed Integer Linear Program (BMILP) with integer variables at both levels. Hence, this problem cannot be further reformulated with the same technique and the next subsections introduce new tools.

4.2. High Point relaxation and tightenings

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 (see [30] for a graphical example). Therefore, a special type of relaxation is required to develop iterative solution methods for these problems. Here we introduce the *High-Point* (*HP*) relaxation for bilevel programs, which is obtained by removing the lower-level objective function but keeping its constraints. The resulting single-level problem is a relaxation of the original bilevel problem [30]. In terms of the general bilevel program described by Eqns. (15)-(17), the *High-Point* problem is defined by Eqn. (40).

$$HP \;:\; \max_{(x,y)\in\Omega} f_1(x, y) \tag{40}$$

The *HP* extends to any multilevel problem: removing all lower levels objective functions yields an upper bound. Nevertheless, solving the *HP* problem usually yields a weak upper bound on the original problem because in this relaxation all variables are controlled by the upper level, waiving any rationality of the lower levels. And the more rational levels are waived, the worse the relaxation can be expected to be. Hence, we will never use the *High-Point* relaxation HP_T of the original trilevel problem. Instead, we will do it on the BMILP reformulation introduced in the subsection above, calling it HP_{BMILP} or *HP* for short. Given that the reformulation already imposes the rationality of the third level, this is a tighter relaxation. This is formalized in definitions Eqns. (41)-(42) and in Prop. 4.

$$HP_T : \max_{(x_{\mathcal{L}}, x_{\mathcal{C}}, y)} \left\{ NPV_{\mathcal{L}}(x_{\mathcal{L}}, y) : (1_{\mathcal{L}})-(5_{\mathcal{L}}),\ (1_{\mathcal{C}})-(5_{\mathcal{C}}),\ y \in \Omega_y(x_{\mathcal{L}}, x_{\mathcal{C}}) \right\} \tag{41}$$

$$HP_{\text{BMILP}} : \max_{(x_{\mathcal{L}}, x_{\mathcal{C}}, y)} \left\{ NPV_{\mathcal{L}}(x_{\mathcal{L}}, y) : (1_{\mathcal{L}})-(5_{\mathcal{L}}),\ (1_{\mathcal{C}})-(5_{\mathcal{C}}),\ (33)-(39) \right\} \tag{42}$$

Proposition 4. *The problems HP_T and HP_{BMILP} are both relaxations of the original trilevel problem, and HP_{BMILP} yields a tighter upper bound than HP_T.*

Proof. Any feasible solution for the original trilevel problem is obviously feasible for these problems, no matter what is the degeneracy resolution chosen. Furthermore, $y \in \Omega_y(x_{\mathcal{L}}, x_{\mathcal{C}})$ are the primal constraints of the market, which are exactly Eqns. (33)-(34). Hence, any point feasible for HP_{BMILP} is also feasible for HP_T yielding the desired result. □

Now we present a method to tighten our HP based on the column-and-constraint generation procedure developed by Zeng and An [42]. Their method tackles general BMILPs by the three steps described below. Notation-wise, remember that any variable with a superscript denotes the value it took when solving a particular problem, and hence it is fixed. Just for this paragraph, we also consider again the general bilevel formulation from Eqns. (15)-(17).

1. Solve the HP problem and collect the solution (x^k, y^k).
2. Fix the leader variables to x^k and solve to optimality the follower's problem $Q\text{-}k = Q(x^k)$, obtaining $y^{Q\text{-}k}$.
3. Add to the HP the equations imposing that the follower has to obtain at least what it would obtain if its integer variables were fixed to the value in $y^{Q\text{-}k}$ and its continuous variables were optimizing its objective value (imposed with KKT for example). Denote this new problem $HP\text{-}k$ or k^{th} master problem $MP\text{-}k$ and iterate.

This proceedure could be directly applied to solve our BMILP, where the integer variables of the lower level are exactly the competitor variables $x_{\mathcal{C}}$ and the continuous ones are the market primal and dual variables. Nevertheless, this would be quite inefficient as we would impose the optimality of a set of variables that happen to already be involved in equilibrium constraints. Below we describe how simply duplicating the constraints with duplicated market (and dual) variables is enough to get as good as a tighening - up to degeneracy.

As above, the idea is to generate constraints that impose the second-level objective function $NPV_{\mathcal{C}}$ to be at least as large as it would be with any of the second-level solutions $\{x_{\mathcal{C}}^k\}_{k=1}^K$ that have been observed. For each of them, this can be expressed with Eqn. (43^k),

$$NPV_{\mathcal{C}}(x_{\mathcal{L}}, x_{\mathcal{C}}, y) \geq NPV_{\mathcal{C}}(x_{\mathcal{L}}, x_{\mathcal{C}}^k, y_k) \tag{43^k}$$

where y_k are duplicate market assignment variables that we want to represent how the market behaves if the competitor chooses $x_{\mathcal{C}}^k$ and the leader whatever $x_{\mathcal{L}}$. In order to enforce this third-level optimality of demand assignments, a full set of duplicate variables $(y_k, \mu_k, u_k \lambda_k)$ and constraints must be appended to the HP problem for each solution $x_{\mathcal{C}}^k$ that has been observed. The constraints correspond to the duality-based reformulation of the third-level problem; they are presented in Eqns. (33^k)-(39^k), where the added superscript k denotes that all competitor variables have been fixed to $x_{\mathcal{C}}^k$.

Proposition 5. *Adding the constraints* (43^k) *and* (33^k)-(39^k) *to the HP yields a tighter relaxation, for any $x_{\mathcal{C}}^k$.*

Proof. The tightening is obvious as we are simply adding equations. So we need to show that Eqns. (43^k) and (33^k)-(39^k) do not exclude any solution that is trilevel feasible. All first level solutions remain feasible after the constraints 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 $x_{\mathcal{C}}^k$. Additionally, we can guarantee that no point in the *Inducible Region* of the trilevel problem is excluded from the tightened HP problem, because Eqn. (43^k) 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 (43^k). Note that Eqns. (43^k) and (33^k)-(39^k) 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. (43^k) is set to the worst possible outcome for the competitor. Thus, again it never cuts any trilevel solution, no matter the chosen degeneration resolution. As will be seen in Section 6, this is used to find the Strategically Optimistic solution. □

4.3. Stability Regions and cuts

To solve a general bilevel program like in Eqns. (15)-(17), the leader has to know the reaction to each of her actions, or at least to the ones that seem attractive. This burden can be aliviated if the leader can group her actions x into regions that are certified to produce the same (fixed) reaction $y^k \in Y$ from the follower. This concept has already been used in bilevel optimization [10, 38]. Here, we formalize them by Eqn. (44) and denote them *Stability Regions* $R(y^k) \subset X$.

$$R(y^k) := \left\{x \in X : y^k \in \Psi(x)\right\} \tag{44}$$

Once the leader has computed the reaction y^k to a certain x^k, she does not need to re-compute it for any other $x \in R(y^k)$, and hence these can be excluded from any future search. Similarly to the *Inducible Region* (IR) and the *High Point* (HP) concepts, using the *Rational Reaction sets* in Defs. 1 and 5, we can extend (44) to the trilevel case. Here we first study the regions $R_{x_{\mathcal{L}}}(x_{\mathcal{C}}^k, y^k) \subset X_{\mathcal{L}}$ given by Def. 6.

Definition 6. A *Stability Region* $R_{x_{\mathcal{L}}}(x_{\mathcal{C}}^k, y^k)$ is the set of first-level decisions that produce the same rational reactions $(x_{\mathcal{C}}^k, y^k)$ in the second and third levels.

$$R_{x_{\mathcal{L}}}(x_{\mathcal{C}}^k, y^k) = \left\{x_{\mathcal{L}} \in X_{\mathcal{L}} :\ (x_{\mathcal{C}}^k, y^k) \in \Psi_{x_{\mathcal{C}}}^0(x_{\mathcal{L}}) \times \Psi_y^0(x_{\mathcal{L}}, x_{\mathcal{C}}^k)\right\} \tag{45}$$

where $\Psi_{x_{\mathcal{C}}}^0(x_{\mathcal{L}})$ and $\Psi_y^0(x_{\mathcal{L}}, x_{\mathcal{C}}^k)$ refer to one of the degeneracy resolution models described in Section 3.2.

Intuition suggests that in the trilevel capacity planning formulation, expanding plants that already have slack capacity do not change the rational response of the second and third levels. This allows us to exactly describe the stability regions with Prop. 6. Refer to Appendix A for the full proof.

Proposition 6. *Let Q-k be the bilevel problem obtained after fixing the first-level decisions to $x^k_{\mathcal{L}}$ in the second and third level problems presented in Eqns.* (12)-(14)*. We denote by $(x^k_{\mathcal{C}}, y^k)$ a corresponding optimal reaction and by μ^k the optimal multipliers associated with capacity constraints in Eqn.* (9)*. Then,*

$$x_{\mathcal{L}} \in R_{x_{\mathcal{L}}}(x^k_{\mathcal{C}}, y^k) \iff \left[c_{t,i} = c^k_{t,i}\right] \vee \begin{bmatrix} c_{t,i} \geq c^k_{t,i} \\ \mu^k_{t',i} = 0 \quad \forall\ t' \in T^+_t \end{bmatrix}, \quad \begin{matrix} \forall\ t \in T \\ \forall i \in I_{\mathcal{L}} \end{matrix} \tag{46}$$

As shown at the end of this section, this disjunctive characterization allows to generate cuts that exclude every point $x_{\mathcal{L}}$ in them. Hence, tighter cuts based on $R_{x_{\mathcal{L}}}(x^k_{\mathcal{C}}, yk)$ are not possible: any other expansion decision that is not excluded by these cuts will generate a different reaction, at least from the market.

To deepen this study, we introduce the *Extended Stability Regions* as the sets $R_{x_{\mathcal{L}}}(x^k_{\mathcal{C}})$, which contain all leader decisions that will not modify the competitor's reaction.

Definition 7. The *Extended Stability Region* $R_{x_{\mathcal{L}}}(x^k_{\mathcal{C}})$ is the set of first level decisions that produce the same rational reaction from the second level $x_{\mathcal{C}}$.

$$R_{x_{\mathcal{L}}}(x^k_{\mathcal{C}}) = \left\{ x_{\mathcal{L}} \in X_{\mathcal{L}} \ :\ \exists y : (x^k_{\mathcal{C}}, y) \in \Psi^0_{x_{\mathcal{C}}}(x_{\mathcal{L}}) \times \Psi^0_y(x_{\mathcal{L}}, x^k_{\mathcal{C}}) \right\} \tag{47}$$

It is clear that $R_{x_{\mathcal{L}}}(x^k_{\mathcal{C}}) \subset R_{x_{\mathcal{L}}}(x^k_{\mathcal{C}}, y^k)\ \forall y^k \in \Psi_y(x_{\mathcal{L}})$, and hence more powerful cuts could be generated. Nevertheless, two problems should be addressed:

1. A change in the market decision directly affects the leader objective function. Hence, it is not clear how the enitre region $R_{x_{\mathcal{L}}}(x^k_{\mathcal{C}})$ could be cut as we do with $R_{x_{\mathcal{L}}}(x^k_{\mathcal{C}}, y^k)$.

2. Finding these regions would be equivalent to finding sensitivity conditions for the middle level, which happens to have integer variables. It is well known that there are no general sensitivity conditions for MILPs; we cannot apply duality theory as we could to derive $R_{x_{\mathcal{L}}}(x^k_{\mathcal{C}}, y^k)$.

The first issue will be addressed in next Section when the constraint-directed exploration algorithm is presented. The second issue is not solvable in general. Fortunately, introducing Assumption 1 on the leader prices, we can partially address it and derive stronger cuts in Prop. 7. This is a common modeling assumption in the industrial environment, basically stating that each costumer sees the leader as a single plant and does not have to worry about what exact plant is serving it.

Assumption 1. *All plants controlled by the leader ($i \in I_{\mathcal{L}}$) offer unique prices to each market ($j \in J$).*

$$P_{t,i_1,j} = P_{t,i_2,j} \quad \forall t \in T,\ (i_1, i_2) \in I^2_{\mathcal{L}},\ j \in J \tag{48}$$

The interest on this assumption is that the market only sees $\mathscr{C}^k$, and hence so does the competitor. Therefore, any rearrangement inside the leader, both in terms of what plants to open or from where to serve the assigned demand, are indifferent -in terms of the objective value- to all the other players. In particular, applying the same proof as for Prop. 6, we can show that any internal rearrangement of the leader's capacity will not change the amount of demand assigned to the leader $\mathscr{D}_t$ (even though the exact y will vary depending on where is the capacity), and, hence, the reaction of the competitor will not change neither.

Proposition 7. *Under Assumption 1 and the notation from Prop. 6,*

$$[\mathscr{C}_t = \mathscr{C}_t^k] \vee \begin{bmatrix} \mathscr{C}_t \geq \mathscr{C}_t^k \\ \sum_{i \in I_{\mathcal{L}}} \mu_{t',i}^k = 0 \quad \forall\, t' \in T_t^+ \end{bmatrix}, \forall\, t \in T \implies x_{\mathcal{L}} \in R_{x_{\mathcal{L}}}(x_{\mathcal{C}}^k) \quad (49)$$

As will be exploited in the algorithm of Section 5, we do not need to solve again the lower level bilevel problem for any $x_{\mathcal{L}}$ in this subset of the *Extended Stability Region.* Hence, here we will deduce a cut that eliminates every point described by the disjunction in Eqn. (49). Note that if Assumption 1 is not satisfied, we are restricted to only generate cuts that remove $R_{x_{\mathcal{L}}}(x_{\mathcal{C}}^k, y^k)$. These can be derived exactly as the ones below, but disagregating also with respect to the plants $i \in I_{\mathcal{L}}$, given that we would express the disjunction in Eqn. (46) instead of (49). In the rest of the paper we will work under Assumption 1, and we will abuse language refering to the *"Stability Region"* R^k as the set of $x_{\mathcal{L}}$ satisfying the Eqn. (49). Next, we introduce some sets to concisely express our cuts in the following Prop. 8.

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

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

$$\Gamma_{x0}^k = \left\{ t \in T : \sum_{i \in I_{\mathcal{L}}} x_{t,i}^k = 0, \quad \sum_{t' \in T_t^+} \mu_{t',i}^k > 0 \right\} \quad (50)$$

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

$$\Gamma_{\mu+}^k = \left\{ t \in T : \sum_{i \in I_{\mathcal{L}}} x_{t,i}^k > 0, \quad \sum_{t' \in T_t^+} \mu_{t',i}^k > 0 \right\} \quad (51)$$

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

$$\Gamma_{\mu0}^k = \left\{ t \in T : \sum_{i \in I_{\mathcal{L}}} x_{t,i}^k > 0, \quad \sum_{t' \in T_t^+} \mu_{t',i}^k = 0 \right\} \quad (52)$$

Proposition 8. *The "Stability Region" R^k described by the disjunction in Eqn.* (49) *is characterized by the use of binary variables $z_{0,t}^k,\ z_{1,t}^k \in \{0,1\} \forall t \in T$, and:*

$$\sum_{t \in \Gamma_{x0}^k} \sum_{i \in I_{\mathcal{L}}} x_{t,i}^k + \sum_{t \in \Gamma_{\mu0}^k} z_{1,t}^k + \sum_{t \in \Gamma_{\mu+}^k} (1 - z_{0,t}^k) = 0 \quad (53)$$

$$\begin{bmatrix} z_{0,t}^k = 1 \\ \mathscr{C}_t = \mathscr{C}_t^k \end{bmatrix} \vee \begin{bmatrix} z_{1,t}^k = 1 \\ \mathscr{C}_t < \mathscr{C}_t^k \end{bmatrix} \vee \begin{bmatrix} z_{0,t}^k + z_{1,t}^k = 0 \\ \mathscr{C}_t > \mathscr{C}_t^k \end{bmatrix} \quad \forall\, t \in T \quad (54)$$

Proof. At any time $t \in T$, the binary variables $z^k_{0,t}$ and $z^k_{1,t}$ indicate if the leader offers more, less, or the same capacity than $\mathscr{C}^k_t$, as expressed by Eqn. (54). The result follows from combining this with the sets from Def. 8 and the disjunction in Eqn. (49) on when the total leader capacity can increase. □

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. (53) to be greater or equal than one (≥ 1). The next two sections describe how to use this and the other results introduced in this seccion to obtain two solution approaches that converge to the different *Optimistic* solutions.

5. 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 the cuts from Prop. 8 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 known). The details of the algorithm are presented below.

5.1. 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 $x^{HP\text{-}k}_{\mathcal{L}}$ lies. This is done by fixing the leader decision to $x^{HP\text{-}k}_{\mathcal{L}}$ and solving the single-level reformulation *Q-k* of the second- and third-level problems. From that solution we not only get a lower bound, but we can now partially describe the *Extended Stability Region* of this observed reaction of the followers R^k by using Prop. 7. Hence, by adding the cuts from Eqn. (53) 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 a first-level solution $x^{HP\text{-}k}_{\mathcal{L}}$. 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 $x^{HP\text{-}k}_{\mathcal{L}}$. Identify a second- and third-level solution ($x^{Q\text{-}k}_{\mathcal{C}}, y^{Q\text{-}k}, \mu^{Q\text{-}k}$).

Step 4: Identify the sets Γ^k_{x0}, $\Gamma^k_{\mu+}$, and $\Gamma^k_{\mu0}$ describing the region R^k that contains $x^{HP\text{-}k}_{\mathcal{L}}$ and all other first-level solutions satisfying the condition given by Eqn. (49).

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

Step 6: Generate *no-good* cuts to exclude R^k from $HP\text{-}k{+}1$. 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 remaining unexplored region of the first-level decision space can be better than the incumbent already found.

C2: If $HP\text{-}k$ 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 $HP\text{-}k$ 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 $Q\text{-}k$ resolves degeneracy in favor of the second level ($y \in \Psi_{\mathcal{C},y}$). 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 Assumption 1. Furthermore, instances with a degenerate third level might not close the gap between UB and LB^* because problems $HP\text{-}k$ and $Q\text{-}k$ 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.

5.2. 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:

- Let $HP\text{-}k_R(x_{\mathcal{C}}^{Q\text{-}k})$ be the *High-Point* problem $HP\text{-}k$ constrained to the region $x_{\mathcal{L}} \in R^k$, and with second-level variables fixed to $x_{\mathcal{C}}^{Q\text{-}k}$
- Let $HP\text{-}k_R(NPV_{\mathcal{C}})$ be the *High-Point* problem $HP\text{-}k$ constrained to $x_{\mathcal{L}} \in R^k$, and with second-level objective value fixed to $NPV_{\mathcal{C}}(x_{\mathcal{L}}^{HP\text{-}k}, x_{\mathcal{C}}^{Q\text{-}k}, y^{Q\text{-}k})$.

Solving $HP\text{-}k_R(x_{\mathcal{C}}^{Q\text{-}k})$ 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 as to fit the best supply scheme for her -without affecting the benefit of the market. Second, to detect if the market is degenerate in the sense of having different optimal assignments that yield different $NPV_{\mathcal{C}}$. This is the case if the new solution $y_{HP\text{-}k_R}$ has a different aggregated demand to every competitor's plant. If it is the case, we cannot conclude anything about the solution $(x_{\mathcal{L}}^{HP\text{-}k_R}, x_{\mathcal{C}}^{Q\text{-}k}, y_{HP\text{-}k_R})$ being trilevel hierarchically feasible because, if the competitor knew that the market would directly favor the leader, he 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 leader $\mathscr{D}_{\mathcal{L},t}$ is the same for all t (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 $Q\text{-}k$).

If the market is not found to be degenerate (or it was solved by a penalty), we still have to check whether the second level is degenerate in the sense of having another possible expansion $x_{\mathcal{C}}$ that yields the same value for him but a better one for the leader. This is checked by solving $HP\text{-}k_R(_{NPV_{\mathcal{C}}})$, and if the objective value of the leader changes, we need to add a penalty to the competitor 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^k_{x0}, $\Gamma^k_{\mu+}$, and $\Gamma^k_{\mu 0}$ describing the region R^k that contains $x^{HP\text{-}k}_{\mathcal{L}}$ and all other first-level solutions satisfying the condition given by Eqn. (49).

Step 4b: Solve $HP\text{-}k_R(x^{Q\text{-}k}_{\mathcal{C}})$ and identify the third-level response ($y^{HP\text{-}k_R}$). If the third-level solution is "aggregately different" from the one obtained in Step 3 ($\exists t \in T : \mathscr{D}^{HP\text{-}k_R(x^{Q\text{-}k}_{\mathcal{C}})}_{\mathcal{L},t} \neq \mathscr{D}^{Q\text{-}k}_{\mathcal{L},t}$), add a penalty to the third-level objective to resolve degeneracy in favor of the first level. Go to Step 3.

Step 4c: Solve $HP\text{-}k_R(_{NPV_{\mathcal{C}}})$. If the first-level objective is different from the one in Step 3 ($NPV_{\mathcal{L}}(x^{HP\text{-}k}_{\mathcal{L}}, x^{Q\text{-}k}_{\mathcal{C}}, y^{Q\text{-}k}) \neq NPV_{\mathcal{L}}(x^{HP\text{-}k_R}_{\mathcal{L}}, x^{Q\text{-}k}_{\mathcal{C}}, y^{HP\text{-}k_R})$), add a penalty to the second-level objective to resolve degeneracy in favor of the first level. Go to Step 3.

The steps of the algorithm are presented schematically in Fig. 2; diamonds control the flow of the algorithm, light gray boxes are simple operations and dark gray boxes involve optimization problems.

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

As opposed 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 [42] 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 controlled by the markets; therefore, no additional reformulation is needed for the continuous variables. The details of the algorithm are presented below.

6.1. Reaching the Strategically Optimistic solution

Algorithm 2 uses the tightening presented in Section 4.2 to sequentially strengthen the *High-Point* relaxation of the trilevel capacity planning problem. The algorithm iterates between a master problem $MP\text{-}k$ that provides upper bounds UP^k and the single-level reformulation of the second- and third-level problems with extra cuts $\overline{Q}\text{-}k$. Problem $MP\text{-}k$ is the *High-Point* relaxation of the bilevel reformulation with the cuts modeled by Eqns. (43^k) and (33^k)-(39^k). The search in $\overline{Q}\text{-}k$ is directed towards unexplored second-level decisions by adding no-good cuts to the problem $Q\text{-}k$, such that second-level decisions that were already observed are not considered in future iterations. These cuts are presented in Eqn. (55).

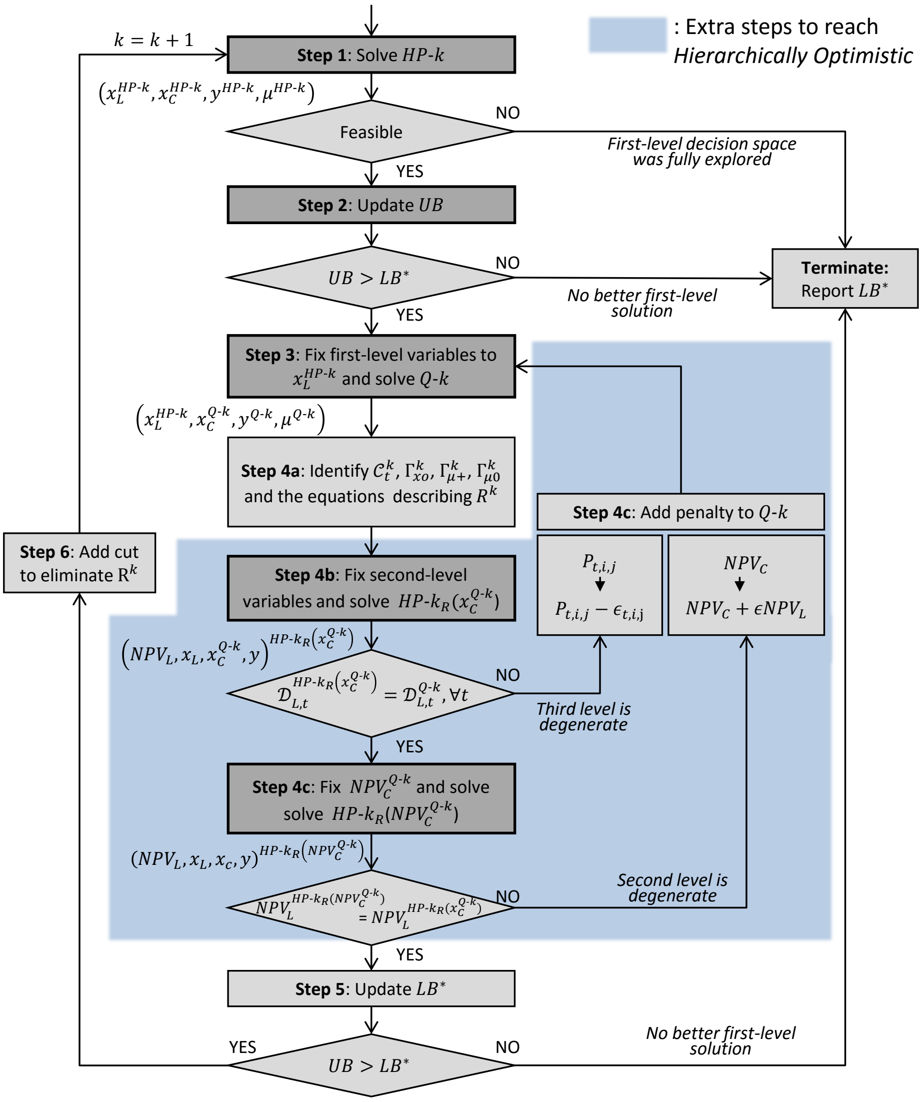

Figure 2: Algorithm 1 for *Sequentially* or *Hierarchically Optimistic* solutions

$$\sum_{i\in I_{\mathcal{C}}}\left[\sum_{t:\ x_{t,i}^{\overline{Q}\text{-}k'}=1}(1-x_{t,i})\ +\sum_{t:\ x_{t,i}^{\overline{Q}\text{-}k'}=0}x_{t,i}\right]\ \geq\ 1 \qquad \forall\ k'\in K \tag{55}$$

where $K=\{1,2,\ldots,k\}$, and $x_{t,i}^{\overline{Q}\text{-}k'}$ denotes the values of a second-level optimal solution for problem $\overline{Q}$-k'.

The algorithm is identified as a column-and-constraint generation approach because at every iteration, a new second-level candidate solution ($x_{\mathcal{C}}^{\overline{Q}\text{-}k}$) 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 into five steps.

Step 1: Solve MP-k. Identify the first-level solution ($x_{\mathcal{L}}^{MP\text{-}k}$) and the second-level objective value $NPV_{\mathcal{C}}(x_{\mathcal{L}}^{MP\text{-}k},x_{\mathcal{C}}^{MP\text{-}k},y^{MP\text{-}k})$.

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

Step 3: Fix first-level variables to $x_{\mathcal{L}}^{MP\text{-}k}$ and solve $\overline{Q}$-k, which is Q-k_R including the no-good cuts from Eqn. (55). If infeasible, terminate and return the solution yielding UB. Otherwise, identify the second-level solution ($x_{\mathcal{C}}^{\overline{Q}\text{-}k}$). If $NPV_{\mathcal{C}}(x_{\mathcal{L}}^{MP\text{-}k},x_{\mathcal{C}}^{\overline{Q}\text{-}k},y^{\overline{Q}\text{-}k}) < NPV_{\mathcal{C}}(x_{\mathcal{L}}^{MP\text{-}k},x_{\mathcal{C}}^{MP\text{-}k},y^{MP\text{-}k})$, 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 tightening MP^{k+1} and the cuts to exclude $x_{\mathcal{C}}^{\overline{Q}\text{-}k}$ from $\overline{Q}$-k+1. Go back to Step 1.

The steps of the algorithm are presented schematically in Fig. 3. Algorithm 2 has three possible stopping criteria:

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

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

C3: If $NPV_{\mathcal{C}}(x_{\mathcal{L}}^{MP\text{-}k},x_{\mathcal{C}}^{MP\text{-}k},y^{MP\text{-}k}) \geq NPV_{\mathcal{C}}(x_{\mathcal{L}}^{MP\text{-}k},x_{\mathcal{C}}^{\overline{Q}\text{-}k},y^{\overline{Q}\text{-}k})$, return the solution ($x_{\mathcal{L}}^{MP\text{-}k}, x_{\mathcal{C}}^{MP\text{-}k}, y^{MP\text{-}k}$) obtained in MP-k. In this case, no other solution contained in the unexplored region can be better for the second level than ($x_{\mathcal{L}}^{MP\text{-}k}, x_{\mathcal{C}}^{MP\text{-}k}, y^{MP\text{-}k}$).

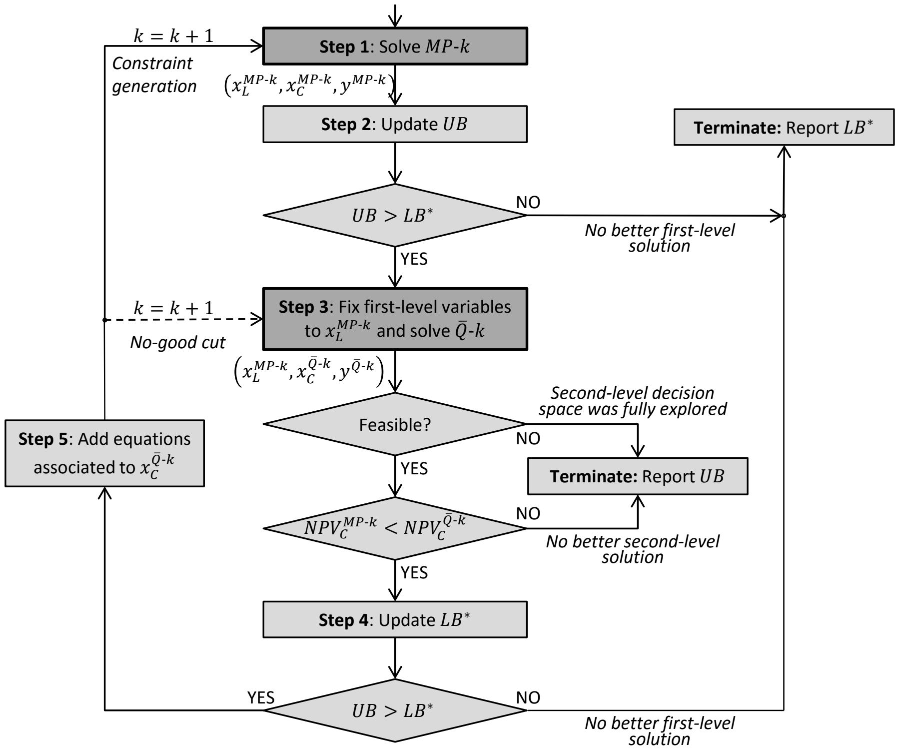

Figure 3: Algorithm 2 for *Strategically Optimistic* solution

It is worth noticing that Step 1 produces an improving UB because the feasible region of problem HP-k is successively reduced according to Prop. 5. 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.

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 seek 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 requires at least the same number iterations as Algorithm 2. This is the case because Algorithm 2 explores at most one point in each *Extended Stability Region*, which is not true for Algorithm 1 (remember that Prop. 7 only provided one direction of the implication). 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.

7. Capacity planning instances

We test Algorithms 1 and 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. 4; the parameters of the instance are given in Tables 2, 3, and 4.

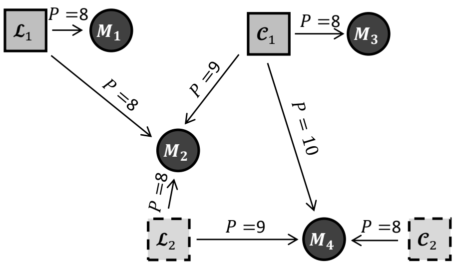

Figure 4: Network of plants and markets in Instance 1

Costumer (j)	Plant (i) $P_{t,\mathcal{L}_1,j}$	$P_{t,\mathcal{L}_2,j}$	$P_{t,\mathcal{C}_1,j}$	$P_{t,\mathcal{C}_2,j}$
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 3: Selling prices [\$/ton] 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 2: Market demands [M ton/period] in Instance 1

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

Table 4: Cost parameters and initial capacities in Instance 1

The optimal expansion strategy for the leader induces 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 5.

Instance 1 has been designed such that Algorithms 1 and 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 two 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. 5.

Items 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 5: Optimal objective values in Instance 1

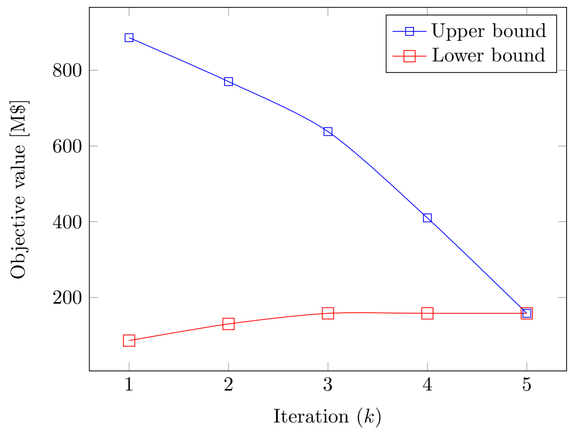

Figure 5: Convergence of Algorithms 1 and 2 in Instance 1

Both algorithms were implemented in GAMS 24.4.1 and the optimization problems were solved using GUROBI 6.0.0 on an Intel Core i7 CPU 2.93 Ghz processor with 4 GB of RAM. 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 both problems have the same 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 6: 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. [15]; 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, he controls 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 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. 6.

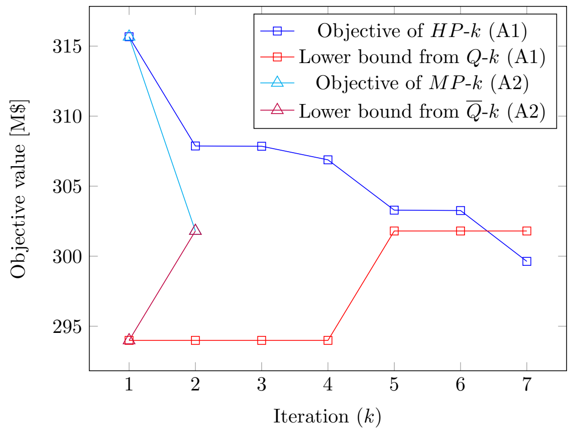

Figure 6: Convergence of Algorithm 1 (A1) and Algorithm 2 (A2) in Instance 2

Table 7 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 takes only 8 s. This instance shows the advantage of Algorithm 2 for problems that are solved in few iterations.

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 8.

The optimal capacity expansion plan for the trilevel formulation differs from

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 7: Computational statistics for Algorithms 1 and 2 in Instance 2

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 8: Optimal objective values in Instance 2

the results reported by Garcia-Herreros et al. [15] 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 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 [15] prescribing three expansions instead of two, the rational reaction of the competition is to expand plant $\mathcal{C}_1$ at time 1. This expansion plan 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 a static competitor.

8. Conclusions and future work

In this work we propose a fully competitive model for capacity planning in a duopoly and formulate it as a trilevel optimization problem. It allows simultaneously considering the conflicting interests of three rational decision-makers within a mathematical programming framework. We also address for the first time the topic of degeneracy in multilevel decision problems, clearing the ambiguity in the characterization of trilevel optimal solutions. For this, we introduce several extensions of the *Optimistic* models from bilevel programming, and we provide 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 economic terms for the industrial instance.

The type of problems that we address are of interest in applications where discrete decisions are taken by different players in a hierarchy. 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 like demand forecasts or costs. In the case of using the model for different market environments, it might be interesting to modify the competition assumption to be Nash-Cournot, while still allowing the competitors to expand.

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.

Appendix A. Proof of Prop. 6

We want to prove that a first-level decision $x_{\mathcal{L}}$ satisfying conditions (46) produces the same rational reaction $(\hat{x}_{\mathcal{C}}, \hat{y})$ as $\hat{x}_{\mathcal{L}}$. The superscript k is replaced by the hat to simplify notation. We divide the proof of Proposition 6 in three steps.

Step 1. *In the third-level market problem* $M(c)$ *from Eqn.* (7)-(9)*, increasing the capacity of one plant cannot increase the demand assigned to any other plant.*

Let us denote by $\hat{M}$ the third-level problem $M(c^k_{t,i})$, with capacities fixed to $c^k_{t,i}$. Let us also denote by $\tilde{M}$ the problem in which plant i' increases its capacity by $\Delta C_{i'} > 0$. Their optimal assignments are $\hat{y}_{t,i,j}$ and $\tilde{y}_{t,i,j}$ respectively. We want to show that they satisfy the conditions presented in Eqn. (A.1).

$$\sum_{j \in J} \tilde{y}_{t,i,j} \leq \sum_{j \in J} \hat{y}_{t,i,j} \qquad \forall\ t \in T,\ i \in I \backslash \{i'\} \tag{A.1}$$

First, we notice that fully utilized plants ($\sum_{j \in J} \hat{y}_{t,i,j} = \hat{c}_{t,i}$) in problem $\hat{M}$ cannot increase the demand assigned to them. For all other plants with slack capacity ($\sum_{j \in J} \hat{y}_{t,i,j} + \hat{s}_{t,i} = \hat{c}_{t,i},\ \hat{s}_{t,i} > 0$), the Lagrange multiplier $\hat{\mu}_{t,i}$ associated with the capacity constraint (9) must be zero according to complementary slackness of the third-level LP.

It was proved by Garcia-Herreros et al. [15] that increasing the capacity of one plant cannot produce an increase in the optimal Lagrange multipliers associated with any capacity constraint (9). Therefore, the Lagrange multipliers $\hat{\mu}_{t,i}$ of plants that had slack capacity in problem $\hat{M}$ remain at zero in the optimum of problem $\tilde{M}$, as expressed in Eqn. (A.2).

$$0 \leq \tilde{\mu}_{t,i} \leq \hat{\mu}_{t,i} = 0 \qquad \forall\ (t,i) \in \{(t,i):\ t \in T,\ i \in I \backslash \{i'\},\ \hat{s}_{t,i} > 0\} \tag{A.2}$$

Since the slack $\hat{s}_{t,i}$ in plants that are not fully utilized in problem $\hat{M}$ could be arbitrarily small, we conclude that the condition in Eqn. (A.1) must be satisfied. More formally, note that the solution to $M' = M(\hat{c}_{t,i} - \hat{s}_{t,i} + \epsilon),\ \epsilon > 0$ has the same optimal solution $\hat{y}$ than $M(\hat{c})$ because it has a smaller feasible region and still contains the optimal solution of problem $M(\hat{c})$. But any arbitrarily small increase in demand would saturate any plant, yielding $\mu'_{t,i} > 0$ and hence contradiction.

Step 2. *The optimal objective value of the second level cannot improve if the capacity of the leader increases* ($\exists i \in I_{\mathcal{L}},\ t \in T:\ \tilde{x}_{t,i} > \hat{x}_{t,i}$) *and capacities of the competitor remain constant.*

Recall from Tab. 1 that the unit total price is $P_{t,i,j} = S_{t,i} + G_{t,i,j}$. Rewriting the objective function of the competition in Eqn. ($6_{\mathcal{C}}$) as in Eqn. (A.3), we observe that the margin obtained from every unit sold only dependents on the production cost $F_{t,i}$ and the selling price $S_{t,i}$ of each plant.

$$\begin{aligned} NPV_{\mathcal{C}} = & \sum_{t\in T}\sum_{i\in I_{\mathcal{C}}}\sum_{j\in J} (S_{t,i} - F_{t,i})\, y_{t,i,j} \\ & - \sum_{t\in T}\sum_{i\in I_{\mathcal{C}}} (A_{t,i}v_{t,i} + B_{t,i}w_{t,i} + E_{t,i}x_{t,i}) \end{aligned} \tag{A.3}$$

Furthermore, any expansion of the leader fits exactly the case of Step 1, that can be applied repeatedly to all leader expansions. Therefore, the condition presented in Eqn. (A.1) implies that the objective function of the second level cannot improve from problem $\hat{M}$ to $\tilde{M}$. This is formalized in Eqn. (A.4).

$$NPV_{\mathcal{C}}(\tilde{x}_{\mathcal{L}}, x_{\mathcal{C}}, \tilde{y}) \leq NPV_{\mathcal{C}}(\hat{x}_{\mathcal{L}}, x_{\mathcal{C}}, \hat{y}) \quad \forall\, x_{\mathcal{C}} \in X_{\mathcal{C}},\ \begin{array}{l} \tilde{y} \in \Psi^0_y(\tilde{x}_{\mathcal{L}}, x_{\mathcal{C}}) \\ \hat{y} \in \Psi^0_y(\hat{x}_{\mathcal{L}}, x_{\mathcal{C}}) \end{array} \tag{A.4}$$

Step 3. *If the expansion strategy of the leader $\tilde{x}_{t,i}$ satisfies the conditions presented in Eqn.* (46)*, the bilevel problems $\hat{Q}$ and $\tilde{Q}$, resulting from fixing the leader to $\hat{x}_{\mathcal{L}} = x^k_{\mathcal{L}}$ and $\tilde{x}_{\mathcal{L}}$ respectively, have the same rational reactions.*

First, we use the duality-based reformulation presented in Eqns. (32)-(36) to verify that optimal solutions $\hat{y}$ to problem $\hat{M}$ are feasible in $\tilde{M}$. This is the case because the capacity constraint Eqn. (33) are relaxed with the additional expansions of the leader, and the dual objective function (right-hand side of Eqn. (32)) 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. (12)-(14), is feasible in the bilevel problem $\tilde{Q}$ since second-level constraints ($1_{\mathcal{C}}$)-($5_{\mathcal{C}}$) are not affected by first-level decisions. Therefore, because the objective function $NPV_{\mathcal{C}}$ 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. (A.5).

$$NPV_{\mathcal{C}}(\hat{x}_{\mathcal{L}}, \hat{x}_{\mathcal{C}}, \hat{y}) = NPV_{\mathcal{C}}(\tilde{x}_{\mathcal{L}}, \hat{x}_{\mathcal{C}}, \hat{y}) \leq NPV_{\mathcal{C}}(\tilde{x}_{\mathcal{L}}, \tilde{x}_{\mathcal{C}}, \tilde{y}) \tag{A.5}$$

where $(\tilde{x}_{\mathcal{C}}, \tilde{y})$ is the optimal solution of $\tilde{Q}$. Furthermore, we can establish the inequalities given in Eqn. (A.6),

$$NPV_{\mathcal{C}}(\tilde{x}_{\mathcal{L}}, \tilde{x}_{\mathcal{C}}, \tilde{y}) \leq NPV_{\mathcal{C}}(\hat{x}_{\mathcal{L}}, \tilde{x}_{\mathcal{C}}, \hat{\tilde{y}}) \leq NPV_{\mathcal{C}}(\hat{x}_{\mathcal{L}}, \hat{x}_{\mathcal{C}}, \hat{y}) \tag{A.6}$$

where $\hat{\hat{y}}$ 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. (A.4), and the inequality on the right follows from optimality of $(\hat{x}_{\mathcal{C}}, \hat{y})$ in problem $\hat{\hat{Q}}$. Eqns. (A.5) and (A.6) together demonstrate that problems $\hat{Q}$ and $\tilde{Q}$ have the same optimal objective value that can be given by the optimal solution $(\hat{x}_{\mathcal{C}}, \hat{y})$. 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 6.

References

[1] Ekaterina Alekseeva, Yury Kochetov, and Alexandr Plyasunov. An Exact Method for the Discrete (r-p)-centroid Problem. *Journal of Global Optimization*, 63:445–460, 2015.

[2] Natalia Alguacil, Andrés Delgadillo, and José M. Arroyo. A Trilevel Programming Approach for Electric Grid Defense Planning. *Computers and Operations Research*, 41(1):282–290, 2014. ISSN 03050548.

[3] G. Anandalingam and Victor Apprey. Multi-Level Programming and Conflict Resolution. *European Journal of Operational Research*, 51(2):233–247, 1991.

[4] Jonathan F. Bard and James T. Moore. An Algorithm for the Discrete Bilevel Programming Problem. *Naval Research Logistics*, 39(3):419–435, apr 1992. ISSN 0894069X.

[5] 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.

[6] Omar Ben-Ayed and Charles E. Blair. Computational Difficulties of Bilevel Linear Programming. *Operations Research*, 38(3):556–560, 1990.

[7] Paul G. Berglund and Changhyun Kwon. Solving a Location Problem of a Stackelberg Firm Competing with Cournot-Nash Firms. *Networks and Spatial Economics*, 14(1):117–132, 2014. ISSN 15729427.

[8] Suk-Gwon Chang and Bezalel Gavish. Telecommunications Network Topological Design and Capacity Expansion: Formulations and Algorithms. *Telecommuncation Systems*, 1(1):99–131, 1993.

[9] Andrés Delgadillo, José Manuel Arroyo, and Natalia Alguacil. Analysis of Electric Grid Interdiction with Line Switching. *IEEE Transactions on Power Systems*, 25(2):633–641, 2010. ISSN 08858950.

[10] Stephan Dempe. *Foundations of Bilevel Programming*. Springer Science & Business, 2002.

[11] Stephan Dempe, Boris S. Mardukhovich, and Alain B. Zemkoho. Necessary Optimality Conditions in Pessimistic Bilevel Programming. *Optimization*, 63(4):505–533, 2014.

[12] Scott T. DeNegre and Ted K. Ralphs. A Branch-and-Cut Algorithm for Integer Bilevel Linear Programs. *Operations Research and Computer Science Interfaces Series*, 47:65–78, 2009.

[13] 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, oct 2006. ISSN 0925-5001.

[14] Terry L. Friesz, Matthew A. Rigdon, and Reetabrata Mookherjee. Differential Variational Inequalities and Shipper Dynamic Oligopolistic Network Competition. *Transportation Research Part B*, 40:480–503, 2006.

[15] Pablo Garcia-Herreros, Lei Zhang, Pratik Misra, Erdem Arslan, Sanjay Mehta, and Ignacio E. Grossmann. Mixed-Integer Bilevel Optimization for Capacity Planning with Rational Markets. *Computers and Chemical Engineering*, 86:33–47, 2016.

[16] Fred Glover. Improved Linear Integer Programming Formulations of Nonlinear Integer Problems. *Management Science*, 22(4):455–460, 1975.

[17] Ignacio E. Grossmann. Advances in Mathematical Programming Models for Enterprise-Wide Optimization. *Computers and Chemical Engineering*, 47:2–18, 2012.

[18] Zeynep H. Gümüs and Christodoulos A. Floudas. Global Optimization of Mixed-Integer Bilevel Programming Problems. *Computational Management Science*, 2(3):181–212, 2005.

[19] Louis S. Hakimi. On Locating New Facilities in a Competitive Environment. *European Journal of Operational Research*, 12(1):29–35, 1983.

[20] Jialin Han, Jie Lu, Yaoguang Hu, and Guangquan Zhang. Tri-level Decision-Making with Multiple Followers: Model, Algorithm and Case Study. *Information Sciences*, 311:182–204, 2015.

[21] Robert G. Jeroslow. The Polynomial Hierarchy and a Simple Model for Competitive Analysis. *Mathematical Programming*, 32(2):146–164, 1985.

[22] Athanasia Karakitsiou. *Modeling Discrete Competitive Facility Location*. Springer, 2015.

[23] Athanasia Karakitsiou and Athanasios Migdalas. Locating Facilities in a Competitive Environment. *Optimization Letters*, 2015. ISSN 1862-4472.

[24] Polyxeni-M. Kleniati and Claire S. Adjiman. Branch-and-Sandwich: a Deterministic Global Optimization Algorithm for Optimistic Bilevel Programming Problems. Part II: Convergence Analysis and Numerical Results. *Journal of Global Optimization*, 60(3):459–481, 2014. ISSN 0925-5001.

[25] Dominik Kress and Erwin Pesch. Sequential Competitive Location on Networks. *European Journal of Operational Research*, 217:483–499, 2012.

[26] Peter B. Luh, Shi-Chung Chang, and Tsu-Shuan Chang. Solutions and Properties of Multi-stage Stackeiberg Games. *Automatica*, 20(2):251–256, 1984.

[27] Athanasios Migdalas. Bilevel Programming in Traffic Planning: Models, Methods and Challenge. *Journal of Global Optimization*, 7:381–405, 1995.

[28] Tan C. Miller, Terry L. Friesz, Roger L. Tobin, and Changhyun Kwon. Reaction Function Based Dynamic Location Modeling in Stackelberg-Nash-Cournot Competition. *Networks and Spatial Economics*, 7:77–97, 2007.

[29] Alexander Mitsos, George M. Bollas, and Paul I. Barton. Bilevel Optimization Formulation for Parameter Estimation in LiquidLiquid Phase Equilibrium Problems. *Chemical Engineering Science*, 64:548–559, 2009.

[30] James T. Moore and Jonathan F. Bard. The Mixed Integer Linear Bilevel Programming Problem. *Operations Research*, 38(5):911–921, 1990.

[31] Frederic H. Murphy and Yves Smeers. Generation Capacity Expansion in Imperfectly Competitive Restructured Electricity Markets. *Operations Research*, 53(4):646–661, 2005. ISSN 0030-364X.

[32] Pierre-Louis Poirion, Sonia Toubaline, Claudia D'ambrosio, and Leo Liberti. Bilevel Mixed-Integer Linear Programs and the Zero Forcing Set. *optimization-online.org: 5210*, 2015.

[33] Marcos Costa Roboredo and Artur Alves Pessoa. A Branch-and-Cut Algorithm for the Discrete (r-p)-centroid Problem. *European Journal of Operational Research*, 224(1):101–109, 2013.

[34] Maria P. Scaparra and Richard L. Church. A Bilevel Mixed-Integer Program for Critical Infrastructure Protection Planning. *Computers & Operations Research*, 35(6):1905–1923, 2008.

[35] Hanif D. Sherali, Allen L. Soyster, and Frederic H. Murphy. Stackelberg-Nash-Cournot Equilibria: Characterizations and Computations. *Operations Research*, 31(2):253–276, 1983.

[36] J. Cole Smith, Churluzu Lim, and Aydin Alptekinoglu. Optimal Mixed-Integer Programming and Heuristic Methods for a Bilevel Stackelberg Product Introduction Game. *Naval Research Logistics*, 56(8):714–729, 2009.

[37] Wolfram Wiesemann, Angelos Tsoukalas, and Polyxeni-Margarita Kleniati. Pessimistic Bilevel Optimization. *SIAM Journal on Optimization*, 23(1): 353–380, 2013.

[38] Pan Xu and Lizhi Wang. An Exact Algorithm for the Bilevel Mixed Integer Linear Programming Problem under Three Simplifying Assumptions. *Computers and Operations Research*, 41:309–318, 2014. ISSN 03050548.

[39] Yiming Yao, Thomas Edmunds, Dimitri Papageorgiou, and Rogelio Alvarez. Trilevel Optimization in Power Network Defense. *IEEE Transactions on Systems, Man and Cybernetics Part C: Applications and Reviews*, 37(4):712–718, 2007. ISSN 10946977.

[40] Alain B. Zemkoho. Solving Ill-posed Bilevel Programs. *Set-Valued and Variational Analysis*, 24(3):423–448, 2016.

[41] Bo Zeng. Easier than We Thought : A Practical Scheme to Compute Pessimistic Bilevel Optimization Problem. *Available at SSRN 2658342*, 2015.

[42] Bo Zeng and Yu An. Solving Bilevel Mixed Integer Program by Reformulations and Decomposition. *Optimization On-line*, 2014.