POLYTECHNIQUE MONTRÉAL
affiliée à l'Université de Montréal

Bilevel models for demand response in smart grids

MATHIEU BESANÇON
Département de mathématiques et de génie industriel
Polytechnique Montréal
et
CRIStAL, INOCS & INRIA
Centrale Lille

Thèse en cotutelle présentée en vue de l'obtention du diplôme de *Philosophiæ Doctor*
Mathématiques

Décembre 2020

POLYTECHNIQUE MONTRÉAL
affiliée à l'Université de Montréal

Cette thèse intitulée :

Bilevel models for demand response in smart grids

présentée par **Mathieu BESANÇON**
en vue de l'obtention du diplôme de *Philosophiæ Doctor*
a été dûment acceptée par le jury d'examen constitué de :

Gilles SAVARD, président
Michel GENDREAU, membre et directeur de recherche
Frédéric SEMET, membre et codirecteur de recherche
Wolfram WIESEMANN, membre
Ivana LJUBIC, membre externe

ACKNOWLEDGEMENTS

This thesis made me cross the path of many people since the beginning of the adventure in August 2017, so this will take a while. I will first deeply thank Luce and Miguel for the tireless discussions, frank answers and advice on research and academic life, during and after the PhD. I thank Frédéric Semet and Michel Gendreau for accepting to supervise this doctoral project in an unconventional setting and dedicating their time to it. Directly next to advisors, I thank Juan Gomez for taking some time to discuss his research with this first-year PhD student storming into his office on a regular basis. His support and advice have been invaluable to build the first parts of this thesis.

I thank all the people I have shared an apartment with across these years, Mickaël, Romain, Capucine, Pierre-Yves, Chelsea, Sally, and of course Luther. It is no small feat to cope with paper drafts and optimization problems scattered in a kitchen. I will keep a special mention for Aude, we missed several months of apartment sharing due to schedule problems, family issues, and not the least, a pandemic, we will catch it up in the future.

The work environment is probably one of the essential factors for the success of a doctoral program and of research in general. I thank the people from GERAD and Polytechnique, past and present, for this productive yet appreciable work environment, starting a PhD could not have been in better condition than with you. On the other side of the Atlantic, I thank all the people from INOCS Lille & Brussels for being a great team and helping each other out.

Thanks to my family for being there from the beginning of this journey, my parents wondering why their son is still a student, my sisters probably amused by the same fact, my grandmothers Anne-Marie and Colette, for the scientific debates on the best method for traditional quiche or a daring courgette cake. I'll have a special mention for my grand-fathers who both helped me grow into who I am, I know they would be proud.

The work presented in this thesis has also been supported by the hands of many open-source contributors without whom every project would start by reinventing a wheel. You have provided me with both exceptional tools and a great community to discuss, discover, improve, and teach. I'll thank for this the Julia community at large, the JuMP & JuliaPolyhedra developers and maintainers, and the LightGraphs team. I will also thank all the local developer communities I interacted with in Montréal, Lille, Paris, Nantes, Berlin and online. I will also have a mention for all the tools developed and shared by dedicated individuals, these are the modern invisible shoulders of giants we stand on.

Other hands and eyes have played a key role during the redaction of this manuscript, many thanks to all those who took on their limited time to read, discuss, and review various parts of this manuscript.

Academia can be a tough environment, certainly more at the beginning, I want to give a special mention to all researchers who took some time to reach out, provide me with invaluable feedback and a different angle on my work or doctoral project: Stefano Gualandi, Lars Schewe, Akshay Gupte, and Michael Poss. In the same line, my deep thanks go to Mikel Leturia and Aissa Ould Dris from the Process Engineering department at the UTC whose inputs have been significant in choosing the path of academic research.

Research is the pillar of an academic career, but sharing the knowledge produced is what creates its impact. I had the pleasure of helping train new generations of engineers on various topics, both in Lille and Montreal, I'd like to thank all people who invited me to teach with them, Frédéric, Diego, Maxime in Lille on the operations research and programming courses, Roland in Montréal on dynamical systems for power grids. If some students read these words, I hope you took something away from the hours we spent in the physical or virtual classroom and enjoyed it.

This PhD would not have happened without the financial support from multiple institutions. I'd like to thank the Région Haut-de-France for funding my project on the French side, and the NSERC Energy Storage Technology (NEST) Network on the Canadian side. I'll also thank the support of the Mermoz scholarship and of the GdR RO young researcher funding, they allowed me to organize the visit to Edinburgh.

Doing this PhD on two continents meant calling many places "home" at different points. Each of these places has left its mark and attachments, I would like to thank all the friends, close or far away, from school, university and former jobs, and look forward to asking you if you have read this thesis, you're allowed to say no. I'll also mention the PhD Discord community, its tireless hours of discussion and debate amidst the chaos of 2020.

ABSTRACT

This thesis investigates mathematical optimization models with a bilevel structure and their application to price-based Demand Response in smart grids.

The increasing penetration of renewable power generation has put power systems under higher tension. The stochastic and distributed nature of wind and solar generation increases the need for adjustment of the conventional production to the net demand, which corresponds to the demand minus the renewable generation.

Demand Response as a means to this adjustment of demand and supply is receiving growing attention. Instead of achieving the adjustment thanks to generation units, it consists in leveraging the flexibility of a part of the demand, thus changing the aggregated demand curve in time.

The first part of this thesis focuses on a Time-and-Level-of-Use Demand Response system based on a price of energy that depends on the time of consumption, but also on a capacity that is self-determined by each user of the program. This capacity is booked by the user for a specific time frame, and determines a limit for energy consumption. Several key properties of the pricing system are studied, focusing on the perspective of the supplier setting the pricing components. The supplier anticipates the decision of the customers to the prices they set, the sequential decision created by this situation is modelled as a Stackelberg or Leader-Follower game formulated as a bilevel optimization problem.

Bilevel optimization problems embed the optimality condition of other optimization problems in their constraints. Their range of applications includes optimization for engineering, economics, power systems, or security games. The inherent computational difficulty of bilevel problems has motivated the development of customized algorithms for their resolution.

In the second part of the contributions, a variant of the bilevel optimization problem is developed, where the upper level protects its feasibility against deviations of the lower-level solution from optimality. More specifically, this near-optimal robust model maintains the upper-level feasibility for any lower-level solution that is feasible and almost optimal for the lower-level. This model introduces a robustness notion that is specific to multilevel optimization. We derive a single-level closed-form reformulation when the lower level is a convex optimization problem and an extended formulation when it is linear. The near-optimal robust bilevel problem is a generalization of the optimistic bilevel problem and is in general harder to solve. Nonetheless, we obtain complexity results for the near-optimal robust bilevel problem, showing it belongs to the same complexity class as the optimistic problem under mild assumptions. Finally, we design exact and heuristic solution methods that significantly improve the solution time of the extended formulation.

RÉSUMÉ

La thèse porte sur les modèles d'optimisation mathématique biniveau et leurs applications à la réponse de la demande dans les réseaux électriques.

L'augmentation de la production d'énergie renouvelable et l'apparition de nouveaux acteurs ont complexifié les opérations et décisions dans les réseaux électriques. La nature aléatoire et distribuée de la génération solaire et éolienne entraîne un besoin d'ajustement de la production conventionnelle à la demande nette, correspondant à la demande après prise en compte de la production renouvelable.

La réponse de la demande est une des solutions utilisées pour faire face à ces nouveaux besoins des réseaux électriques. Au lieu d'étudier l'adaptation de la production à la charge, son principe est d'exploiter la flexibilité d'une partie de la consommation, ajustant ainsi la courbe de demande au cours du temps.

Dans la première partie de cette thèse, nous étudions un système de réponse de la demande par prix dynamique, TLOU. Dans ce système, un usager réserve une capacité pour une période donnée, et paie un prix dépendant du dépassement de sa capacité par la consommation sur la période donnée. Nous étudions les propriétés de ce système de tarification, en particulier du point de vue du fournisseur déterminant les paramètres de prix. L'interaction entre le fournisseur et les usagers est modélisée comme un jeu de Stackelberg ou meneur-suiveur qui est résolu par une approche d'optimisation mathématique biniveau.

Les problèmes d'optimisation biniveau sont caractérisés par un problème d'optimisation imbriqué dans les contraintes d'un autre problème d'optimisation. Leurs champs d'applications incluent les problèmes de conception en ingénierie, les modèles économiques, les réseaux électriques ou encore la sûreté des systèmes.

Dans la deuxième partie de la thèse, une formulation du problème biniveau est proposée dans laquelle le deuxième niveau n'est plus nécessairement résolu exactement, mais peut dévier de son optimum d'une quantité limitée. Nous développons une formulation biniveau robuste à la quasi-optimalité (NORBiP) dans laquelle le premier niveau s'assure de trouver une solution dont la faisabilité est garantie pour l'ensemble des solutions quasi optimales du deuxième niveau. Ce modèle introduit une notion de robustesse spécifique à l'optimisation multiniveau. Une reformulation à un seul niveau est développée dans le cas où le deuxième niveau est un problème d'optimisation convexe, basée sur la dualisation des contraintes de robustesse. Dans le cas où le deuxième niveau est un problème linéaire, une formulation étendue linéarisée est proposée. Bien que cette formulation robuste soit plus difficile à résoudre que le problème biniveau classique, nous établissons des résultats sur sa complexité, démontrant que le problème robuste à la quasi-optimalité appartient à la même classe de complexité que le problème optimiste équivalent sous certaines hypothèses. Enfin, des algorithmes exacts et heuristiques sont proposés pour accélérer la résolution de problèmes biniveaux robustes à la quasi-optimalité dans le cas linéaire.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS AND ACRONYMS

DR	Demand Response
LCP	Linear Complementarity Problem
LP	Linear (Optimization) Problem
MIBLP	Mixed-Integer Bilevel Linear (Optimization) Problem
MILP	Mixed-Integer Linear (Optimization) Problem
MINLP	Mixed-Integer Non-Linear (Optimization) Problem
NORBiP	Near-Optimal Robust Bilevel Problem
TLOU	Time-and-Level-of-Use
TOU	Time-of-Use

LIST OF APPENDICES

CHAPTER 1 INTRODUCTION

This thesis tackles some forms of bilevel optimization problems and applications to pricing mechanisms for demand response in power grids. This chapter introduces the context of power grids motivating the work and foundational concepts in mathematical optimization and hierarchical games.

1.1 Context

In the past decades, power grids across the world have evolved to become complex systems in which multiple actors act simultaneously. Their purpose, that is, instantaneously and reliably bringing power to multiple users of various needs, is almost taken for granted. Nonetheless, current trends such as the growing electrification of transport have been broadening the areas of societies relying on their operations. At the same time, the growth of renewable energy sources has changed the nature of supply. Wind generation and solar panels are now playing a major role in the energy mix of several major power systems. These technologies are characterized by low operating CO_2 emissions, but also by the uncertainty on their production, even on a range of a few hours. Unlike thermal power plants, they also come in smaller scales and can therefore represent an affordable investment at the level of households or micro-grids. These private investments have been accelerated by government incentives across the globe. Such equipment, often combined with an energy storage capacity transform the household from a simple consumer of electric energy to a prosumer. Prosumers are economic entities that are partly power consumer, partly power producer, sometimes simultaneously [1, 2]. These actors with dual roles create the needs for market mechanisms of greater complexity than the former paradigm separating entities in terms of supply and demand.

These trends can become opportunities to reduce the carbon footprint of power systems by making electricity production less carbon-intensive. However, they also increase the complexity of power grid operations, both through the uncertainty on the production and the distributed and fragmented nature of power generators.

Economic and organizational approaches have received growing attention to face these challenges to complement technical solutions such as storage. Aggregators, in particular, are a new type of actor playing the role of buffers between the grid operator and consumers or prosumers. By regrouping multiple consumers, they can also mutualize investment and install a higher storage capacity and generation.

1.2 Hierarchical games and bilevel optimization

Models from mathematical optimization and game theory can be used in order to understand and predict the decision-making process of economic agents. Game theory focuses on economic problems where multiple agents make inter-dependent decisions. Hierarchical games in particular are decision-making processes in which each agent observes the decisions of agents deciding before them and anticipate the decisions of agents deciding after them. They offer a rich toolset for the new forms of economic interactions of power grids. Hierarchical games where players are divided into two sets are referred to as Stackelberg or leader-follower games from the seminal work [3]. Fig. 1.1 illustrates the structure and differences between Stackelberg and Nash games, in which no player may observe the decisions of the others before making their own.

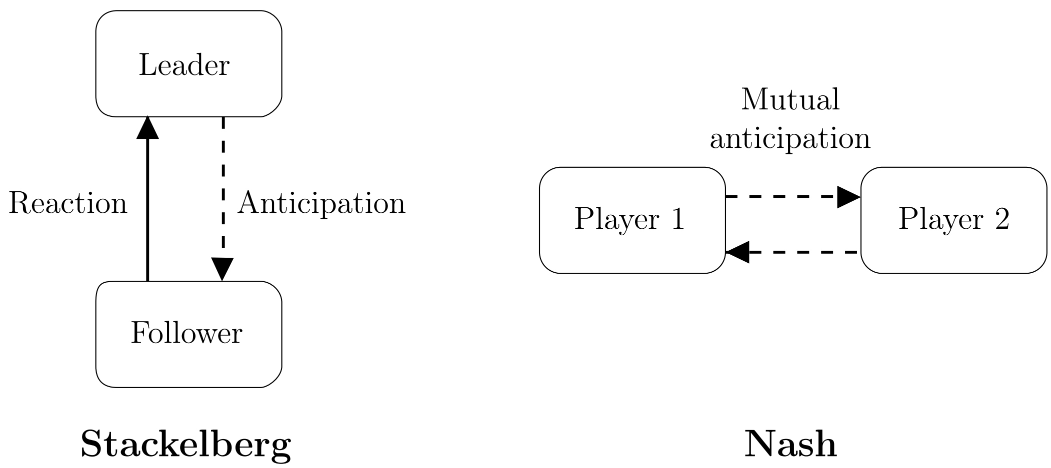

Figure 1.1 Structure of Stackelberg and Nash games

Mathematical optimization, also referred to as mathematical programming, is a domain of applied mathematics, focused on finding the solution of highest quality to a problem depending on decision variables. We define a solution as an assignment of values to the decision variables of the problem and the quality of a solution is assessed based on an objective function.

In a more formal notation, an optimization problem can be stated as:

$$\min_{x \in \mathcal{X}} F(x) \tag{1.1a}$$

$$\text{s.t.} \quad G_i(x) \in \mathcal{S}_i \qquad \forall i \in [\![m]\!] \tag{1.1b}$$

where:

- $[\![m]\!] \equiv \{1, 2, \dots, m\}$
- x is the collection of decisions of the problem, which take values in the domain $\mathcal{X}$;
- $F : \mathcal{X} \mapsto \mathbb{R}$ is the objective function, which is used to compare different decisions. The value of $F(x)$ is referred to as the *objective value*. Since the problem is formulated as a minimization problem, a lower objective value is preferred;
- Each of the m constraints is composed of a mapping $G_i : \mathcal{X} \mapsto \mathcal{A}_i$ and a set $\mathcal{S}_i \subseteq \mathcal{A}_i$.

$\mathcal{X}$ is typically considered to be a compact set, and in many cases a compact subset of $\mathbb{R}^n$, although optimization over other objects can be of interest, such as functions [4,5], matrices, graphs, sets [6], geometrical shapes [7] or text strings [8].

Completely generic optimization problems represented by functions and sets without any structure are seldom studied. Instead, investigations on theoretical properties and solution methods have focused on specific forms of optimization problems i.e. problems with particular forms of F, G, $\mathcal{S}_i$ and $\mathcal{X}$. Some examples of specific optimization problems are Linear Problems (LP), Mixed-Integer Linear Problems (MILP), Constraint Integer Programming (CIP) [9] or unconstrained non-linear problems. When the objective function does not depend on the decision variables, any solution is acceptable and the problem is typically referred to as a feasibility problem.

Although we often consider the optimization problem as solved by a person or an economic entity, we can also view some systems in physics and engineering as optimization problems solved by reaching an equilibrium:

> *"Nature optimizes. Physical systems tend to a state of minimum energy."* [10]

A distinction is often made between *exact*, *approximation* and *heuristic* methods. Exact methods seek both a solution and a guarantee that this solution is optimal. Approximation methods provide by design a worst-case guarantee on the quality of the solution which typically comes as a worst-case ratio between the objective value of a solution found by the method and the optimal objective value. Heuristic methods focus on searching for a good solution without any guarantee on its quality compared to the optimum and are computationally less demanding than exact methods. Most of the thesis focuses on exact methods for the problems considered. Some heuristics are also developed in Chapter 8.

Mathematical optimization formulations typically represent a single decision-making process and are thus not suited to represent hierarchical games. Optimization problems with two decision-makers taking successive actions are referred to as *bilevel optimization* problems and will be the main focus of this thesis. Keeping a notation similar to Problem (1.1), the bilevel problem is expressed as:

$$\text{“}\min_{x\in\mathcal{X}}\text{”}\; F(x,v) \tag{1.2a}$$
$$\text{s.t. } G_k(x,v)\in\mathcal{S}_k \qquad \forall k\in[\![m_u]\!] \tag{1.2b}$$
$$\text{where } v \text{ solves: } \min_{y\in\mathcal{Y}}\{f(x,y) \text{ s.t. } g_i(x,y)\in s_i\ \forall i\in[\![m_l]\!]\}. \tag{1.2c}$$

Quotes were added to the minimization operator “$\min_{x\in\mathcal{X}}$” because this formulation of the bilevel problem presents an ambiguity and is not well-posed. Indeed, there may exist multiple lower-level solutions v which are feasible and yield an optimal value. In such a setting with multiple lower-level optimal solutions, it is not clear which one would be chosen. Solutions approaches to alleviate this ambiguity are developed in Section 2.2.

The problem solved by Eq. (1.2c) is called the lower-level problem, it depends on the upper-level variable x, and reacts to it by choosing a solution v that minimizes its own objective value subject to its own constraint set. The upper-level decision-maker must make a decision x which is feasible for them while anticipating the reaction of the lower level. By formulating economic settings as bilevel optimization problems, the modeller can capture rich interactions between the first and second agent and represent the interest and constraints of both.

1.3 Research objectives and outline

The objectives of the thesis are two-fold. First, we want to design demand response programs adapted to the current smart grid context, capturing interactions between the entity setting prices and the entity receiving them in a leader-follower setting formulated as a bilevel optimization problem. The second objective is to extend and adapt the framework of bilevel optimization to these pricing problems, and in particular to deal with the uncertainty of the upper-level decision-maker when anticipating the lower-level reaction. For this second objective, we study a new robust framework based on bilevel optimization where the lower level is assumed to make decisions that may fail to reach optimality.

CHAPTER 2 LITERATURE REVIEW

This chapter offers an overview of the recent research on topics connected to the thesis. Section 2.1 highlights recent models for demand response applied to power grids. Section 2.2 lays out the mathematical formulations, solution methods, and results leveraged in the literature to handle bilevel optimization problems. Finally, Section 2.3 provides the reader with research on Nash and Stackelberg games integrating a form of uncertainty.

2.1 Demand response in power grids

Demand Response (DR) regroups different programs allowing the power grid operator to leverage the consumption flexibility of customers. The final aim for the supplier is to create an additional control lever when optimizing power grid operations [11], increasing or decreasing the flexible consumption at a given time to adapt to the current net consumption level. Typical examples of actions performed through DR include:

- Valley filling: increasing the consumption at a time point where it is expected to be low (off-peak);
- Peak shaving: decreasing the consumption at a time point where it is expected to be high (on-peak);
- Load shifting: displacing some energy-consuming activities in time, from an on-peak to an off-peak period.

DR programs have been classified based on the way agents responsible for the consumption are influenced as either *price-based* or *incentive-based* programs [12,13]. Price-based DR programs use the price as the sole mechanism to request and encourage customers to adapt their consumption. Incentive-based DR programs include other technical and economic levers and distinguish between classical programs, where the supplier has partial control of a flexible consumption unit, and market-based programs, where the supplier taps into different pools of flexibility. The DR program studied in this thesis is based on dynamic prices as price-based programs but includes a capacity picked by the user to encourage self-limiting their consumption.

In [14], the barriers to and enablers for the successful implementation of Demand Response are reviewed, along the technological, societal, and economic axes. Barriers notably include heavy communication and computation requirements for users, privacy, and data security concerns when sensitive information has to be collected, stored, and exchanged, and behavioural barriers with the potential non-rationality of users. These barriers have been major considerations when designing the DR program presented here. In [15], comparisons of DR mechanisms on different aspects are provided. Price-based programs and especially real-time pricing are considered the most promising as they require less involvement of the supplier in the customer consumption and do not limit the level of consumption.

The development of DR broadens the scope of the task given to the operator handling the network: not only must they ensure the balance of supply and demand, but also decide which levers to use to change the demand. Finding the best way to utilize flexible resources has given birth to *aggregators*, which are middle-entities between the grid and final consumers [16]. Aggregators are of particular interest to draw upon the DR potential of smaller electricity customers, such as residential customers who represent a large part of the total energy consumption but do not have the technical capacity to operate complex energy management systems. Moreover, DR programs with great numbers of households would lead to scalability issues, with the supplier handling and anticipating reactions from numerous entities. Aggregating the consumption of individual customers allows for simplified management of the grid, with fewer entities (the aggregators), each setting DR programs for their residential or commercial end-users. In incentive-based DR programs based on mean-field games [17, 18], the supplier takes control of a large number of consumption units with similar underlying dynamics, such as water heaters or air conditioning systems. The mean-field control ensures an ability to leverage consumption time flexibility, while respecting comfort and technical constraints. Incentive-based programs based on economic programs rather than direct load control are investigated in [19, 20]. In these programs, the supplier offers cost reductions for users consuming below their estimated baseline when they are required to do so. Two critical components to the design of these programs are on one hand the estimation of the baseline that can be subject to gaming from users [19], and on the other hand the estimation of responsiveness to changes in the level of incentive [21].

The work in [22,23] focuses on the response of *microgrids* that are small networks integrating multiple consuming units (households or commercial users), with local renewable production, storage, and an energy management system. The considered microgrids integrate the consumption patterns of residential, commercial, and industrial users with distributed generation sources, both from wind and solar systems. In both models, the uncertain power generation from wind and solar systems is modelled using continuous probability distributions, which are sampled using e.g. a Monte-Carlo approach. In [24], a real-time pricing mechanism is considered for Demand Response between a supplier and microgrids integrating storage and dealing with consumption uncertainty. The resulting bilevel optimization problem is solved using a heuristic particle swarm algorithm with an exact branch-and-bound method for the lower level. In [25, 26], residential DR programs are developed based on varying quality of service or reliability. The reliability menus are designed by the supplier and then self-determined by consumers depending on their reliability requirements at different time frames.

Time-and-Level-of-Use (TLOU) pricing policies have been studied as extensions of the Time-of-Use (TOU) pricing scheme. They allow the supplier to set prices not only as a function of time but also of the power or energy consumption itself. The price of energy is often defined as a step-wise constant function of time and consumption. In [27–29], the optimization of consumption of smart homes, and smart buildings aggregating multiple users is studied under a TLOU pricing. The prices as a function of time and level are known in advance and considered fixed. In [30], TLOU prices are applied to a manufacturing use case. The industrial unit owns a storage system and on-site stochastic renewable energy sources and co-optimizes its consumption, internal energy management, and demand to the grid.

The framework presented in [31] defines a two-stage DR program combining incentive-based elements with time-dependent electricity pricing. The bilevel formulation of the problem maximizes the supplier's profit while maintaining the profit and comfort of each user. A mixed-integer non-linear bilevel model is presented for TOU pricing in [32]. A custom solution method is developed, leveraging the fact that the feasible domains of the two levels are independent. In [33–35], TOU prices are optimized by a power supplier facing independent users, each minimizing both their total cost and an inconvenience function, defined as a required shifting of loads in their schedule.

2.2 Stackelberg games and bilevel optimization

Stackelberg or Leader-Follower games are a type of non-cooperative game, first introduced in [3] for a market with multiple firms, including a player making the first decision often referred to as the leader, and at least one player which observes the decision of the leader and can take it into account before solving their own optimization problem. The leader's objective and potentially constraints depends on the decision of the follower, which they need to anticipate.

Bilevel optimization studies mathematical optimization problems where the optimality of other optimization problems are embedded in the constraints. Its first definition is given in [36]. Bilevel optimization is the way of formulating Stackelberg games in a mathematical optimization formalism, even though it is not the only area it has been applied to. Instead of representing a second player, the lower level can model a system optimizing itself or reaching an equilibrium, with practical applications in chemical engineering [37–39], design of industrial processes [40] or mechanical systems [41–43].

We provide below results on the complexity classes of different bilevel problems. Complexity classes are defined for decision problems that only admit a boolean answer. They are typically extended to optimization problems by formulating the decision problem of finding a solution that is feasible and attains a given bound on the objective value. Two complexity classes commonly used to classify decision problems arising from optimization are $\mathcal{P}$ and $\mathcal{NP}$. Decision problems belonging to $\mathcal{P}$ can be answered in polynomial time, while for a problem belonging to $\mathcal{NP}$, a positive answer can be verified in polynomial time given a certificate. Typical examples of optimization problems for which the corresponding decision versions are in $\mathcal{P}$ and $\mathcal{NP}$ are the linear and mixed-integer linear optimization problems respectively. Given a complexity class X, a problem L is X-hard if it at least as hard as all problems from X i.e. all problems from X can be reduced to L in polynomial time. Furthermore, L is X-complete if it is X-hard and belongs to X. The linear bilevel problem, for which the objective and constraint functions are linear, is strongly $\mathcal{NP}$-hard; a reduction to a pure binary linear problem in [44] shows that the corresponding decision problem is $\mathcal{NP}$-complete. If the lower-level problem contains variables that must take integer values or more generally if it cannot be solved in polynomial time, other complexity classes of the polynomial hierarchy [45] are needed to characterize the complexity of the bilevel problem. It is shown in [46] that the bilevel problem where the lower level itself is $\mathcal{NP}$-complete is in Σ_2^P, meaning it can be solved in non-deterministic polynomial time if equipped with an oracle solving problems in $\mathcal{NP}$ in constant time. The paper also generalizes this result to multilevel problems, where a number of players (potentially greater than 2) make decisions sequentially, with linear objective functions and linear constraints. In such a setting with $n+1$ players, the resulting multilevel problem belongs to Σ_n^P, or to Σ_{n+1}^P if the lower-level itself is not in $\mathcal{P}$, with integrality constraints for instance. One main takeaway from these complexity results is that a polynomial-time solution to bilevel problems should not be expected unless $\mathcal{P} = \mathcal{NP}$, which is considered unlikely in the complexity theory community.

We highlighted the ill-posedness of the initial formulation of the bilevel Problem Eq. (1.2), due the possible non-uniqueness of the lower-level optimal solution. Several approaches have been developed to alleviate this amiguity that is often described as "an unpleasant situation" [47]. The most common, referred to as the *optimistic approach*, consists in assuming the lower level chooses the optimal solution which is most favourable to the upper level. In this setting, the upper level can assign values to the lower-level variables as long as these values are an optimal solution for the lower level problem, resulting in the following modification of Problem (1.2):

$$\min_{x\in\mathcal{X},v\in\mathcal{Y}} \; F(x,v) \tag{2.1a}$$

$$\text{s.t.}\; G_k(x,v) \in \mathcal{S}_k \qquad \forall k \in [\![m_u]\!] \tag{2.1b}$$

$$v \in \arg\min_{y\in\mathcal{Y}}\{f(x,y) \text{ s.t. } g_i(x,y) \in s_i \;\forall i \in [\![m_l]\!]\}. \tag{2.1c}$$

Even though the optimistic approach may appear as a strong assumption, it is often reasonable in the context of Stackelberg games. Indeed, the leader acting as the upper level can break the tie in favour of their preferred option by offering an arbitrarily small incentive to the follower acting as the lower level.

For some problems, anticipating a cooperative reaction of the lower level remains unrealistic. If the lower level may choose among multiple solutions, the upper level can assume any of these solutions could be selected, including the worst one, which is the motivation for the pessimistic approach. The pessimistic bilevel model is commonly formulated as:

$$\begin{aligned} &\min_{x\in X}\max_{v\in Y(x)} \; F(x,v) \\ &\text{where: } Y(x) = \arg\min_{y\in\mathcal{Y}}\{f(x,y) \text{ s.t. } g_i(x,y) \in s_i \,\forall i \in [\![m_l]\!]\} \\ &X = \{x \in \mathcal{X}, G_k(x) \in \mathcal{S}_k \forall k \in [\![m_u]\!]\}. \end{aligned} \tag{2.2}$$

By selecting the lower-level response v through a "max" operator, the pessimistic model assumes the response that is worst with respect to the upper-level objective. This model requires all upper-level constraints to be independent of the lower-level variables, and therefore captured in the upper-level feasible set $x \in X$. Such requirement is not suitable for many models of interest, where the upper-level feasibility is also determined by lower-level decisions. This motivated a more general constraint-based pessimistic bilevel model as defined in [48], presented here with the notation of Problem (1.2):

$$\begin{aligned}
\min_{x}\ & F(x) && (2.3)\\
\text{s.t. }\ & G_k(x,v) \in \mathcal{S}_k\ \forall v \in Y(x) && \forall k \in [\![m_u]\!]\\
& x \in \mathcal{X}\\
& \text{where: } Y(x) = \arg\min_{y\in\mathcal{Y}}\{f(x,y) \text{ s.t. } g_i(x,y) \in s_i\ \forall i \in [\![m_l]\!]\}.
\end{aligned}$$

The universal quantifier $\forall v \in Y(x)$ requires feasibility of the upper-level constraint for all possible optimal responses of the lower-level. Changing the universal quantifier $\forall v \in Y(x)$ for an existential one $\exists$ turns the problem into the optimistic formulation [48]. This formulation is a generalization of the objective-based min-max pessimistic Problem (2.2), which can be expressed in the form of Problem (2.3) using an epigraph formulation:

$$\begin{aligned}
\min_{x,\tau}\ & \tau\\
\text{s.t. }\ & \tau \geq F(x,v)\ \ \forall v \in Y(x)\\
& x \in X = \{x \in \mathcal{X}, G_k(x) \in S_k \forall k \in [\![m_u]\!]\}\\
& \text{where: } Y(x) = \arg\min_{y\in\mathcal{Y}}\{f(x,y) \text{ s.t. } g_i(x,y) \in s_i\, \forall i \in [\![m_l]\!]\}.
\end{aligned}$$

Two alternative models are developed in [49]; the *rewarding* solution is obtained when the upper-level decision is made in a pessimistic fashion and the selected lower-level decision is best with respect to the upper-level problem. The *deceiving* solution is obtained when the upper-level decision is based on an optimistic assumption and the selected lower-level decision is worst with respect to the upper-level objective. In [50], a single-leader multi-follower Stackelberg game is developed, where the followers play a Nash game among themselves. A *semi-optimistic* case is developed, where the chosen Nash equilibrium is not optimistic nor pessimistic but determined by specific properties of the lower-level Nash game.

2.2.1 Single-level reformulations

Bilevel problems are often reformulated in a single level that is then solved using existing methods. Three single-level reformulations are presented in [47, Chapter 3]. When the lower level is a convex optimization problem for any fixed upper-level decision x and when Slater's constraint qualifications hold, its Karush-Kuhn-Tucker (KKT) conditions are necessary and sufficient for optimality. Let the lower level be written as:

$$\min_{y \in \mathcal{Y}} \; f(x, y) \tag{2.4a}$$

$$\text{s.t.} \;\; g_i(x, y) \leq 0 \qquad \forall i \in [\![m_l]\!] \tag{2.4b}$$

where $\mathcal{Y}$ is a convex set and $f(x, \cdot)$ and $g_i(x, \cdot)$ are convex functions. The feasible set of the lower level function-based constraints is given by:

$$Y(x) = \{y \mid g(x, y) \leq 0\}.$$

A solution $\hat{y}$ is optimal for the lower-level problem iff:

$$0 \in \partial_y f(x, \hat{y}) + N_{Y(x) \cap \mathcal{Y}}(\hat{y}) \tag{2.5}$$

where $\partial_y f(x, \hat{y})$ denotes the subdifferential of f with respect to y evaluated at $(x, \hat{y})$ and $+$ denotes the Minkowski sum of the two sets. If f is continuously differentiable, the subdifferential is a singleton $\{\nabla_y f(x, \hat{y})\}$. $N_{Y(x) \cap \mathcal{Y}}(\hat{y})$ is the normal cone of $Y(x) \cap \mathcal{Y}$ at $\hat{y}$. The Primal KKT transformation consists in replacing Problem (2.4) with Eq. (2.5) in the bilevel problem, resulting in a single-level optimization problem with equilibrium constraints (MPEC) [51,52], which still requires specialized solution approaches. When $\mathcal{Y} = \mathbb{R}^{n_l}$, with n_l the dimension of the lower-level decision space, i.e. all constraints on the lower-level decision are contained in the functions g_i, and if Slater's regularity conditions hold, the expression of the KKT conditions can be simplified to:

$$0 \in \partial_y f(x, y) + \sum_{i=1}^{m_l} \lambda_i \partial_y g_i(x, y) \tag{2.6a}$$

$$g_i(x, y) \leq 0 \qquad \forall i \in [\![m_l]\!] \tag{2.6b}$$

$$\lambda_i \geq 0 \qquad \forall i \in [\![m_l]\!] \tag{2.6c}$$

$$\lambda_i g_i(x, y) = 0 \qquad \forall i \in [\![m_l]\!] \tag{2.6d}$$

with λ_i the dual variable associated with constraint i. Constraints (2.6b-2.6d) are often formulated using the complementarity notation:

$$0 \leq \lambda_i \perp - g_i(x, y) \geq 0.$$

This approach is referred to as the classical KKT transformation [47]. The single-level reformulation using Eq. (2.6) as necessary and sufficient conditions for lower-level optimality is an optimization problem with complementarity constraints. Slater's constraint qualifications require the lower-level problem to be strictly feasible for any feasible x:

$$\exists y,\, g(x, y) < 0.$$

In [53], examples are developed, where Slater's constraint qualifications are not respected. Computing local solutions to bilevel optimization problems is a particularly challenging task; in [54], the authors highlight that Problem (2.6) explicitly including the dual variables has local minimizers which do not correspond to local solutions to the original bilevel problem.

The last approach commonly used does not require convexity of the lower-level problem. We can thus use the more general notation of Problem (1.2), based on the value function of the lower-level defined as:

$$\phi(x) = \min_y \{f(x, y) \text{ s.t. } g_i(x, y) \in s_i \,\forall i \in [\![m_l]\!]\}.$$

The lower-level problem can be replaced by a condition on the value function:

$$\begin{aligned}
&v \in \mathcal{Y} \\
&f(x, v) \leq \phi(x) \\
&g_i(x, v) \in s_i \,\forall i \in [\![m_l]\!].
\end{aligned} \tag{2.7}$$

At optimality, Constraint (2.7) is tight. The key difficulty of the value function reformulation is that $\phi(x)$ is non-smooth, non-convex, and potentially hard to compute since evaluating it at any point requires solving the lower-level problem. A value-function approach is used in [55] to tackle mixed-integer bilevel problems. The properties of the KKT and value function reformulations and in particular partial calmness at local solutions are studied in [56, 57]. In [58], the value function of a bilevel problem with a mixed-integer lower level is reformulated as a subproblem for which the feasible set is independent of the upper-level variables. The single-level reformulation results in a MILP with a large number of constraints added on the fly during a branch-and-cut procedure.

2.2.2 Solution approaches for particular bilevel problems

A special case of particular interest is the linear bilevel problem, which we write as:

$$\begin{aligned}
\min_{x,v}\ & \langle c_x, x\rangle + \langle c_y, v\rangle \\
\text{s.t. }\ & Gx + Hv \leq q \\
& v \in \arg\min_y \{\langle d, y\rangle \text{ s.t. } Ax + By \leq b\} \\
& x \geq 0
\end{aligned}$$

resulting in the following KKT reformulation:

$$\begin{aligned}
\min_{x,v,\lambda}\ & \langle c_x, x\rangle + \langle c_y, v\rangle && \text{(2.8a)}\\
\text{s.t. }\ & Gx + Hv \leq q && \text{(2.8b)}\\
& Ax + By \leq b && \text{(2.8c)}\\
& B^T\lambda + d = 0 && \text{(2.8d)}\\
& 0 \leq \lambda_i \perp (b_i - (Ax)_i - (Bv)_i) \geq 0 \qquad \forall i \in [\![m_l]\!] && \text{(2.8e)}\\
& x \geq 0, v \geq 0, \lambda \geq 0 && \text{(2.8f)}
\end{aligned}$$

Problem (2.8) is linear, except for the complementarity Constraints (2.8e). These complementarity constraints alone make the feasible region of the single-level formulation non-convex and even disjoint, as illustrated in [59, Example 2]. A common reformulation of Constraint (2.8e) uses so-called "big-M" constraints:

$$\begin{aligned}
& \lambda \geq 0 \\
& Ax - Bv \leq b \\
& \lambda_i \leq M_D z_i \\
& b_i - (Ax)_i - (Bv)_i \leq M_P(1 - z_i) \\
& z \in \{0,1\}^{m_l}.
\end{aligned}$$

The binary variables z_i introduced are equal to 1 if the primal constraint is active and 0 otherwise. An advantage of this approach is that the non-linearities are removed and replaced with binary variables. The complete single-level reformulation becomes a Mixed-Integer Linear Problem (MILP) well-suited to general-purpose solvers computing the global optimum. The choice of primal and dual bounds M_P M_D is often made through trial and error and can lead to the optimal solution being cut off, even when the bound constraint is not tight [60]. Finding suitable big-M bounds is itself an NP-hard problem, as shown in [61]. An alternative solution is to relax the equality of the complementarity into an inequality:

$$\lambda \geq 0 \tag{2.9a}$$

$$b - Ax - Bv \geq 0 \tag{2.9b}$$

$$\lambda_i(b_i - (Ax)_i - (Bv)_i) \leq \epsilon \qquad \forall i \in [\![m_l]\!] \tag{2.9c}$$

as proposed in [62] and studied numerically in [63]. This results in a non-linear problem with non-convex bilinear constraints. This problem is often solved either with local methods or using relaxations and spatial branching. MINLPs are tackled in a similar fashion in [64], first transforming integrality constraints into complementarity constraints and then relaxing these similarly to (2.9). In [65], a similar approach penalizes a duality gap to converge to feasible solutions for bilevel problems.

Instead of using non-linear solvers to relax the problem or artificial bounds on the primal and dual variables, a solution method consists of branching on the complementarity constraint directly. MILP solvers allow this branching scheme on the complementarity constraint directly using Special Ordered Sets of type 1 (SOS1). These constraints initially defined in [66] are implemented in most Constraint Programming and MILP solvers and express the constraint that, for a set of variables, at most one can take a non-zero value. Using SOS1 constraints allows practitioners to avoid managing a branching tree that can be algorithmically challenging while ensuring the correctness of a model without big-M bounds. Recent computational investigations in [67] suggest that the SOS1 approach can even be preferred over big-M constraints when combined with inequalities strengthening the formulation.

When the lower-level problem is not convex, the KKT equation system cannot be used to fully describe necessary and sufficient optimality conditions. The simplest bilevel optimization problem with such structure is the mixed-integer bilevel linear problem (MIBLP), which is Σ_2^P-hard [46]. General-purpose exact solvers for MIBLP have been developed and described in [68, 69], [70], [71, 72], and [73], based on custom branching and cutting rules in a branch-and-cut algorithm starting from the relaxation of the lower-level optimality condition, also called *High Point Relaxation*.

In [74], MIBLPs are considered, where the lower-level problem is purely binary. The authors define a notion of k-optimality for a lower-level solution that is at least as good as any feasible lower-level solution within a Hamming distance of k. This defines a relaxation of the bilevel problem, where the lower level can be solved to k-optimality. This relaxation defines a hierarchy of bounds converging to the bilevel-optimal solution

2.3 Hierarchical games under uncertainty

In this section, we review some models developed to capture uncertainty in optimization problems and games. Lots of decision-making problems deal with uncertain parameters, a common metaphor considers "nature" as a decision-maker choosing the uncertain parameter against the optimizer [75].

In stochastic optimization, the source of uncertainty is represented by random variables with a probability distribution over the possible outcomes. Typical examples include the minimization of the expected value of a function $F(x, \xi)$ over all possible outcomes of the random variable ξ, or a constraint on the total variance of a function $g(x, \xi)$, which is often used as a measure for the risk of a decision. The work in Chapter 5 is based on a stochastic approach, with a bilevel model where the upper and lower level both know the probability distribution of uncertain energy consumption, but with neither of the agents knowing the outcome at the time of their decision.

In some cases, the probability distribution of the uncertain parameter is unknown or makes the stochastic formulation of the optimization problem computationally intractable. In such a case, only the set of possible values of the uncertain parameter would be known. The robust optimization approach consists of finding a solution to the problem assuming any potential outcome of the uncertain parameter can be realized, including the "worst" one. If the worst-case is determined with respect to the objective, the resulting problem is formulated as:

$$\min_{x \in \mathcal{X}} \max_{\xi \in \Xi} F(x, \xi)$$

with Ξ the set of values that may be taken by the uncertain parameter. When the uncertainty affects the feasibility of the optimization problem, a constraint-based formulation is preferred:

$$\begin{aligned} &\min_{x \in \mathcal{X}} \; F(x) \\ &\text{s.t.} \;\; G_i(x, \xi) \in S_i \;\; \forall \xi \in \Xi \qquad\qquad \forall i \in [\![m]\!] \end{aligned}$$

The model with uncertainty on the objective can be transformed into an equivalent model with uncertainty on the constraints with an epigraph form [76]:

$$\begin{aligned} \min_{x\in\mathcal{X},\tau}\ & \tau \\ \text{s.t.}\ & \tau \geq F(x,\xi)\ \ \forall \xi \in \Xi. \end{aligned}$$

The near-optimal robust model presented in Chapter 6 applies concepts of robust optimization to a generalized pessimistic bilevel model similar to the model introduced in [48]. A variant of robust optimization we will consider focuses on robust optimization under decision-dependent uncertainty [77], under which the uncertain set $\Xi(x)$ depends on the decision x. Problems with such constraints have also been coined as generalized semi-infinite optimization problems [78].

In [79], the authors consider a bilevel problem and relax the assumption of pure rationality of the lower level, considering instead that the lower level may optimize their problem to near-optimality using an algorithm from a predetermined set. Another form of incomplete rationality of the lower level is considered in [80]. The authors develop a bilevel model where the lower level optimizes an objective function that is opposite to the upper-level objective function while preserving a certain quality on their own objective. A parameter α controls this tradeoff between the lower-level objective and the minimization of the upper-level objective. The pessimistic linear bilevel problem and the min-max linear model are special cases of the proposed formulation. They also generalize the pessimistic-optimistic bilevel model defined in [81] using this α-pessimistic approach. This approach is comparable to the objective-based model of the near-optimal robustness developed in Chapter 6. In [82], Linear Complementarity Problems (LCPs) are studied under uncertainty in a robust framework. They highlight that strict complementarity cannot be guaranteed for all realizations of the uncertain parameter and instead minimize the relative complementarity gap.

Bilevel models with uncertainty on some parameters of the lower level have been investigated in a stochastic optimization framework in [83, 84] and [85, 86]. The right-hand side of the lower level is stochastic for the upper level, but revealed before the upper level decides, such that the lower-level problem is a parameterized deterministic problem. This form of uncertainty, affecting the upper level only, is similar to the robustness approach developed in Chapter 6. In [87], a formulation is defined to unify multilevel and multi-stage stochastic optimization through the use of risk functions capturing a part of the objective function that is unknown at a given stage, either because of a random parameter or of the decision of another player. In [88], the optimistic and pessimistic versions of a stochastic bilevel optimization problem are investigated with linear objective and constraints. The authors derive a MILP providing a lower bound on the optimistic version and another providing an upper bound on the pessimistic version, both based on decision rules. Finally, in [89], a stochastic approach is designed to alleviate the ambiguity of the lower-level response in Multi-Leader-Follower Games. Instead of relying on the optimistic assumption considered naive in the case of multiple leaders or the pessimistic assumption perceived as excessively risk-averse, the authors define a belief measure for each leader, associating an upper-level solution to a probability measure on the potential lower-level response.

CHAPTER 3 THESIS ORGANIZATION

The work presented in this thesis is structured around two aspects. The first one is the development of a bilevel framework for the optimal price-setting of a Time-and-Level-of-Use program. The second component is the development of the concept of near-optimal robustness for bilevel problems.

In the context of setting prices in a Demand Response program, using a bilevel approach allows the energy supplier to anticipate a reaction of the users to the pricing decision. Modelling the supplier decision as a Stackelberg game formulated using bilevel optimization relies on the assumption that the user, acting as a lower level, is a strategic player. This implies that they make decisions based on their own objective, which in that case corresponds to minimizing their expected cost as presented in Chapter 4, which was published in the IEEE Transactions on Smart Grids [90]. In Chapter 5, we formulate a multi-user TLOU pricing problem where each user's consumption follows a mixed continuous-discrete probability distribution. We construct several variants of the multi-user pricing problem, each associated with different ways to group the users into clusters.

Bilevel optimization raises the problem of lower-level ambiguity; when the set of optimal solutions for the lower-level problem is not reduced to a singleton, it is unclear which solution would be chosen. In a pricing context such as the TLOU application, if a user is given a set of options with the same cost, they may pick any of these optimal choices, or make a decision based on other secondary objectives. The optimistic assumption is the most commonly-made choice and works well when all optimal solutions are equivalent for the lower level. However, the TLOU program developed in this thesis has a particular solution which is not to book any capacity, corresponding to the flat TOU pricing. All other things equal, a user could be expected to prefer TOU which offers less variability, while the optimistic model implies that the user is accepting any solution resulting in the optimal objective value. The net gain necessary to make the user change their decision is an assumption of the problem and can be interpreted as a "conservativeness" of the user. They would be assumed not to choose a "better" solution over the baseline if it does not result in a net improvement of their objective, thus modifying the central assumption of bilevel optimization.

In the second part of the thesis, we explore bilevel optimization where the lower level may not be solved to perfect optimality but instead deviates by a limited tolerance. Instead of assuming the lower-level to be a strategic and perfectly rational player, one can assume the lower level makes decisions with so-called bounded rationality [91, 92], meaning the decision may be sub-optimal, but with the objective not excessively deviating from the optimal objective value. It is a restriction of the optimistic bilevel problem and includes the pessimistic bilevel problem as a special case. In Chapter 6, near-optimal robust bilevel optimization is defined on generic bilevel problems and a duality-based solution method is defined when convexity of the lower level holds. When the lower level is linear, an extended formulation of the single-level problem can be derived, allowing for the use of MILP solvers. In Chapter 7, we consider near-optimal robust versions of bilevel and multilevel optimization problems and establish several results on the complexity of solving these variants compared to their non-robust counterpart. Finally, Chapter 8 defines two groups of algorithms to solve near-optimal robust linear bilevel problems, both based on the extended formulation proposed in Chapter 6. One is an exact method exploiting the independence of near-optimality robustness of each of the upper-level constraints, and the other is a group of heuristic methods computing a bilevel-feasible, near-optimal robust solution by solving a smaller MILP problem.

CHAPTER 4 ARTICLE 1: A BILEVEL APPROACH TO OPTIMAL PRICE-SETTING OF TIME-AND-LEVEL-OF-USE TARIFFS

Authors: Mathieu Besançon, Miguel F. Anjos, Luce Brotcorne, Juan A. Gomez-Herrera, published in [90][1].

Abstract

The Time-and-Level-of-Use (TLOU) system is a recently developed approach for electric energy pricing, extending Time-of-Use with an energy capacity that customers can book in advance for a given consumption time. We define a bilevel optimization model for determining the pricing parameters of TLOU, maximizing the supplier revenue while anticipating an optimal reaction of the customer. A solution approach is built, based on the discrete finite set of optimality candidates of the lower-level customer problem.

4.1 Introduction

With the increasing proportion of solar and wind generation in the energy mix, power systems have to accommodate greater variability on the supply side, along with more complex decisions on the demand side with generation and storage units down to the residential level. Demand Response (DR) has been seen as one of the promising approaches to these challenges, leveraging customers' flexibility to provide services for better operation of power systems. DR programs are often classified as incentive-based or price-based programs [12]. Different price-based programs are compared in [93] from the perspective of a supplier designing the pricing program, anticipating in a bilevel framework the reaction of a prosumer with storage and shifting capacity.

A detailed review of the literature on bilevel optimization for price-based DR is available in [94]. Incentive-based programs require more commitment from the demand side, thus are often less suitable for targeting residential customers. Price-based DR requires less commitment and constraints on the customer side, thus offering a greater flexibility, but does not provide the supplier with strong guarantees on the actual or even expected demand. An approach taken in price-based DR is to offer varying reliability of electricity to customers [25], leaving the supplier free to adapt the power effectively served to the customer. The insufficiency of short-term estimation methods was identified in a technical report [95] as as one of the critical barriers to the effective implementation of DR. Even though Time-and-Level-of-Use (TLOU) is more related to price-based DR, the self-determined capacity creates an incentive for respecting the upper bound on the consumption on the part of the customer. It was defined in [27] as an extension of the Time-of-Use (TOU) pricing scheme, targeting specifically the issue with current large-scale DR programs identified in the FERC report [95]. Specifically, in the TLOU context, TOU is just a special setting with a null booked capacity. The authors of [27] develop the optimal planning and operation of a smart building under TLOU pricing. In this work, we propose a bilevel optimization model to assist in determining the optimal TLOU pricing structure for a supplier.

The reaction of the customer to the proposed pricing is integrated in the supplier decision problem, thus turning the customer-supplier interaction into a Stackelberg game solved as a bilevel optimization problem. Using specific properties of the customer problem, the optimal capacity decision can be reduced from a continuous set to a discrete finite number of choices which can be computed independently of other decisions. Through this transformation, the necessary and sufficient conditions for lower-level optimality are expressed as a set of linear constraints. For each time frame and corresponding consumption distribution, the set of optimal pricing options can be computed as solutions to the bilevel problem with fixed lower level. These options can be computed in advance, and one can be selected by the supplier ahead of the consumption time to create an incentive for the customer to book and consume a given capacity.

The contributions of this letter are the following: developing further the conceptual basis of the TLOU pricing and some of its key properties, building a bilevel model for the supplier's problem and a specialized solution method, and highlighting through numerical experiments the ability of TLOU to create different incentives depending on the supplier's needs.

This letter is structured as follows. Section 4.2 introduces the TLOU pricing, along with the variant used in this work. In Section 4.3, the model of the supplier decision problem is developed, and necessary optimality conditions are defined to design an efficient solution method. Computational experiments are presented in Section 4.4 for a supplier offering prices to incentivize the customer to follow a certain capacity profile across the day. Section 4.5 concludes the work.

4.2 TLOU pricing

The TLOU policy extends TOU by allowing a customer to book an energy capacity that they self-determine at each time frame depending on their planned requirements. Through this mechanism, they provide the supplier with information on the energy they would possibly consume. The capacity is the energy booked by the customer for a given time frame, following the same terminology as the initial description of TLOU in [27]. As in TOU, the price of energy depends on the time frame within the day, but also on the capacity booked by the customer. TLOU is applied in a three-phase process:

1. The supplier sends the pricing information $(K, \pi^L(c), \pi^H(c))$ to the customer.
2. The customer books a capacity from the supplier for the time frame before a given deadline.
3. After the time frame, the energy cost is computed depending on the energy consumed x_t and booked capacity c_t:
   - If $x_t \leq c_t$, then the applied price of energy is $\pi^L(c_t)$ and the energy cost is $\pi^L(c_t) \cdot x_t$.
   - If $x_t > c_t$, then the applied price of energy is $\pi^H(c_t)$ and the energy cost is $\pi^H(c_t) \cdot x_t$.

The first step corresponds to the pricing decision of the supplier for a given time frame. They then send the pricing to the customer, who takes in a second time their decision by booking a capacity $c_t \geq 0$, minimizing their expected cost. The pricing system is described by three elements: a booking fee K, a step-wise decreasing function $\pi^L(c_t)$ representing the lower energy price and a step-wise increasing function $\pi^H(c_t)$ representing the higher energy price. $\pi^L(c_t)$ will refer to the function of the capacity and π^L_j to the value of the lower price at step j.

An important feature of this DR program is the minimal information that must be exchanged by the parties involved. Unlike DR programs which require users to provide their projected needs or delegate the scheduling decisions to the supplier, TLOU only requires a capacity from the user, with the option of falling back on TOU pricing by not booking a capacity for any time frame.

We use the TLOU definition presented in [94]. Unlike the original convention of [27], the totality of the energy consumed is paid at the lower tariff if it remains below the booked capacity, and at the higher tariff otherwise, as described in Eq. (4.1). In other words, if the consumption over the time frame remains below the booked capacity, the effective energy price is given by the lower tariff curve; if the consumption exceeds the booked capacity, the energy price is given by the higher tariff. This asymmetry in the pricing system creates a strong incentive to make the capacity act as an upper bound on the consumption, while under-consumption is penalized by a soft term proportional with the deviation as highlighted in Fig. 4.2. Customers are still able to consume above the capacity if necessary, and the supplier is compensated for this deviation by the higher price that applies. The total cost for the customer associated with a booked capacity c and a consumption X_t for a time frame t is:

$$\mathcal{C}(c_t; X_t) = \begin{cases} K \cdot c_t + \pi^L(c_t) \cdot X_t, & \text{if } X_t \ \leq \ c_t, \\ K \cdot c_t + \pi^H(c_t) \cdot X_t & \text{otherwise.} \end{cases} \tag{4.1}$$

In the rest of this letter, the index of the considered time frame is dropped when not necessary in an expression to keep the notation succinct.

Proposition 4.2.1. *If a customer books a capacity $c > 0$ for a given time frame and assuming $K + \pi^L(c) < \pi^H(c)$, the lowest cost per $kW \cdot h$ is reached when the consumption is exactly equal to the booked capacity.*

Proof. If $X \leq c$, the total cost is given by $Kc + X\pi^L(c)$, hence the relative cost per consumed unit of energy is $\frac{Kc}{X} + \pi^L(c)$, which is strictly decreasing with X. The discontinuity at $X = c$ is positive, since the relative total cost changes from $\frac{Kc}{X} + \pi^L(c)$ to $\frac{Kc}{X} + \pi^H(c)$. The relative cost at $X = c$ is $K + \pi^L(c)$, assuming $K + \pi^L(c) < \pi^H(c)$, there is no decrease of the relative cost below the point it reaches at $c = X$. $\square$

An example illustrating Proposition 4.2.1 is given Fig. 4.1 and Fig. 4.2.

One property of interest derived from Proposition 4.2.1 is that the customer does not have any incentive to signal a capacity different from their consumption intent. This is in particular valuable for consuming units which are able to adjust their consumption through storage or flexible loads.

Another property of the proposed pricing is that the customer does not need to explicitly signal to the supplier that they do not wish to participate in TLOU at some given time frame. A customer can opt-out of the program simply by booking a capacity $c = 0$, for which the applied pricing matches TOU.

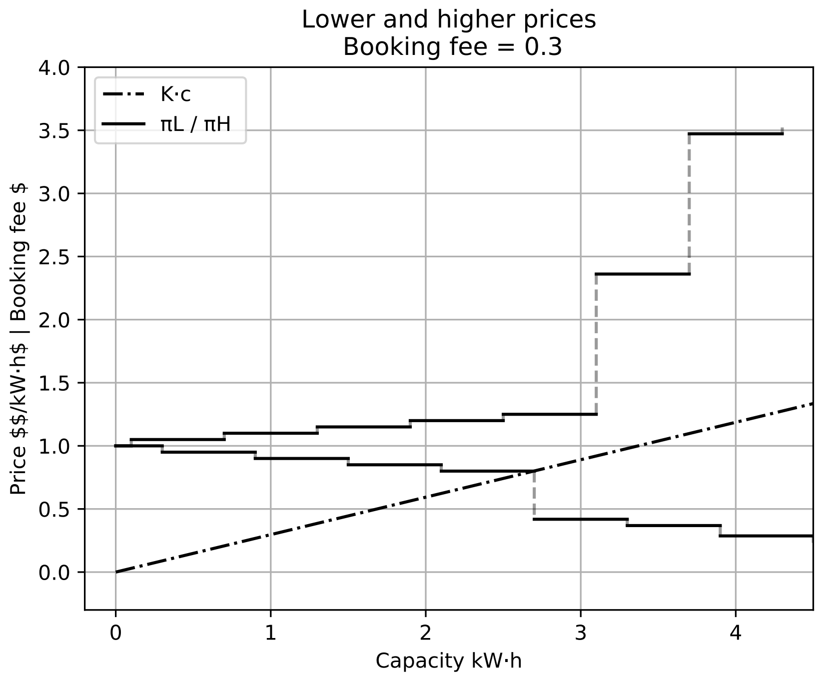

Figure 4.1 Example of TLOU pricing

In a realistic setting, the supplier will have multiple, potentially heterogeneous customers. The model presented here tackles the single-customer case, but also applies when the supplier is able to offer different price settings specific to each individual customer.

4.3 Bilevel model for the supplier decision

In this section, we present a mathematical optimization problem modeling the supplier decision when setting the parameters of the TLOU pricing. The sequential decision process, with one agent reacting to the decision of the other and taking it into account when solving their decision problem, is represented as a Stackelberg game, and can be modeled as a bilevel optimization problem. In this bilevel formulation, the upper-level represents the energy supplier deciding on the pricing components (K, π^L, π^H), while the lower-level represents the customer's response to the supplier's decision, in terms of booking a capacity c.

We consider the consumption of the customer to be unknown to both the supplier and themselves when taking both the pricing and the booking decision. Both players know the probability distribution of this consumption ahead of time before making their decision. This probability distribution is assumed to be discrete and its support Ω is a finite set, with each

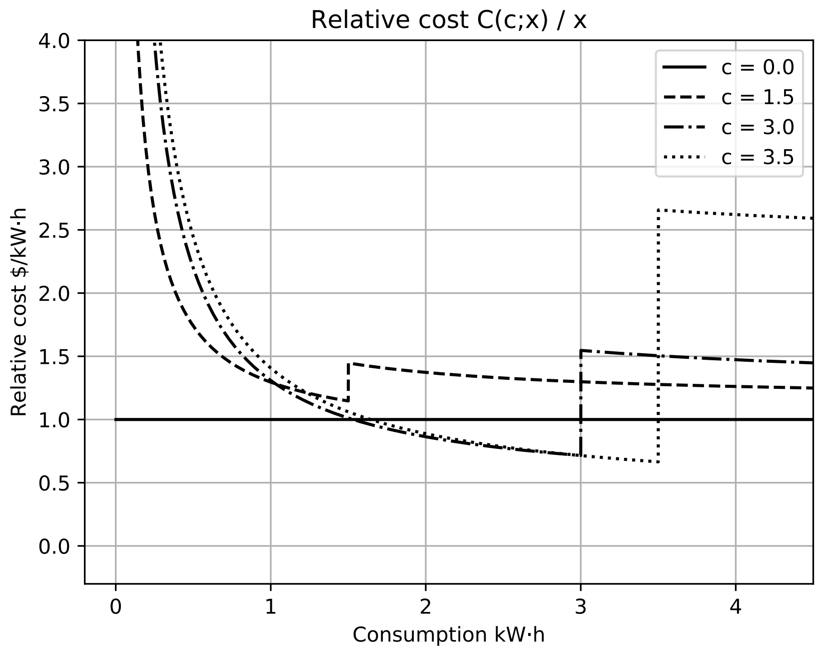

Figure 4.2 Relative cost of energy vs consumption for different capacities

element representing a consumption scenario $\omega \in \Omega$, with associated probability and value p_ω, x_ω respectively. The probability distribution and its support likely depend on the time frame considered t, we note it Ω_t when the time frame is specified. The expected cost of the customer, which is equivalent to the expected revenue of the supplier, is given as a function of both capacity and pricing:

$$\begin{aligned}\mathcal{C}(c, K, \pi^L, \pi^H) = K \cdot c + &\sum_{\omega \in \Omega^-(c)} x_\omega p_\omega \pi^L(c) + \\ &\sum_{\omega \in \Omega^+(c)} x_\omega p_\omega \pi^H(c),\end{aligned} \tag{4.2}$$

with any capacity booked defining a partition of the set of scenarios:

$$\Omega^-(c) = \{\omega \in \Omega, x_\omega \leq c\} \;\&\; \Omega^+(c) = \{\omega \in \Omega, x_\omega > c\}. \tag{4.3}$$

[94, Proposition 3.1] defines the subset of capacities respecting the necessary optimality conditions:

$$S_t = \{0\} \cup C^L \cup \Omega_t \tag{4.4}$$

where the steps of the lower and higher price functions are given at different breakpoints:

$$\begin{aligned} &\{c_0^L, c_1^L, c_2^L, ...\} = C^L && \& \ \{\pi_0^L, \pi_1^L, \pi_2^L, ...\} = \pi^L \\ &\{c_0^H, c_1^H, c_2^H, ...\} = C^H && \& \ \{\pi_0^H, \pi_1^H, \pi_2^H, ...\} = \pi^H \end{aligned}$$

The customer books the cost-minimizing capacity at each time frame, given the corresponding probability distribution. With the finite set of optimal candidates S_t, this constraint can be re-written as:

$$\mathcal{C}(c_t) \leq \mathcal{C}(c) \ \forall c \in S_t, \tag{4.5}$$

which corresponds to finitely many linear constraints. If multiple values of c reach a minimum cost, there is no unique choice the supplier can anticipate from the customer. In such situation, the supplier cannot correctly foresee the decision of the customer. They would want to ensure that the preferred solution of the customer is unique, by making one solution strictly lower in cost than all other capacity candidates. We re-formulate this requirement as the cost being lower than that of any other solution by at least a quantity $\delta > 0$. This quantity can be interpreted as the conservativeness of the customer (unwillingness to move to an optimal solution up to a difference of δ). It is a parameter of the decision-making process of the supplier, estimated a priori by the supplier based on its risk aversion and estimation of customer reactivity. Given this conservativeness parameter, the lower-level optimality conditions of a candidate $k \in S_t$ for a time frame t become:

$$\mathcal{C}(c_{tk}, K, \pi^L, \pi^H) \leq \mathcal{C}(c_{tl}, K, \pi^L, \pi^H) - \delta \ \forall l \in S_t \backslash \{k\}. \tag{4.6}$$

Given that the special case $c = 0$ matches the TOU pricing, the net difference in expected cost of δ is also ensuring that the effort of the customer committing to a capacity and engaging in the program should not be expected by the supplier below a net gain of δ for the customer.

The space of price parameters can be further restricted to include regularity constraints on the pricing curves, lower and upper bounds on the prices. All these constraints can be expressed as linear inequalities, and are summarized with the notation:

$$(K, \pi^L, \pi^H) \in \Phi \tag{4.7}$$

For each capacity candidate $k \in S_t \backslash \{0\}$, the supplier finds the optimal pricing parameters such that the candidate is the optimal capacity to book for the customer, and is at least better than any other candidate by a difference of δ. The k^{th} price-setting problem is expressed as:

$$\max_{K,\pi^L,\pi^H} \mathcal{C}(c_{tk}, K, \pi^L, \pi^H) \tag{4.8a}$$

$$\text{s.t.} \ (K, \pi^L, \pi^H) \in \Phi \tag{4.8b}$$

$$\mathcal{C}(c_{tk}, K, \pi^L, \pi^H) \leq \mathcal{C}(c_{lt}, K, \pi^L, \pi^H) - \delta \ \forall l \in S_t \backslash k, \tag{4.8c}$$

where $\mathcal{C}(c_{tk}, K, \pi^L, \pi^H)$ is defined in Eq. (4.2). In [94], a multi-objective version of the same model is developed, the second objective is experimentally found to be non-conflicting with the revenue maximization.

Formulation (4.8) fully captures the bilevel nature of the problem through Constraint (4.8c), which specifies that the specific k-th capacity candidate must be the optimal choice by at least δ for the customer (lower level). Given the discrete nature of the lower-level decision (choosing the lowest-cost capacity candidate among a discrete set), the Karush-Kuhn-Tucker primal-dual optimality conditions could not be used. Constraint (4.8c) is used as an optimality condition for the lower-level, leveraging the fact that the optimality candidates can all be known prior to the optimization phase.

4.4 Computational experiments

In this section, we present the setup and results for experiments with a supplier setting a TLOU pricing for 24 time frames, picking the pricing minimizing the deviation from the hourly. The data used and further experimentation are presented in [94, Section IV].

Given the probability distribution of the consumption at each time frame, the supplier can construct a TLOU price setting for each capacity candidate by solving Problem (4.8). Once these hourly options have been computed, the supplier can choose to create an incentive for the customer to follow a profile. To avoid unexpected deviations, the supplier can set the prices, as to have a non-zero capacity as close as possible to the average consumption of each time frame, as shown in Fig. 4.3. We use a $\delta = \$0.005$ and similar settings as [94, Section IV] for other parameters (with a reference TOU price of \$1.0).

Alternatively, a major objective of many DR programs is to smooth the demand curve by peak shaving and valley filling. The sum of variations from one time frame to the next can

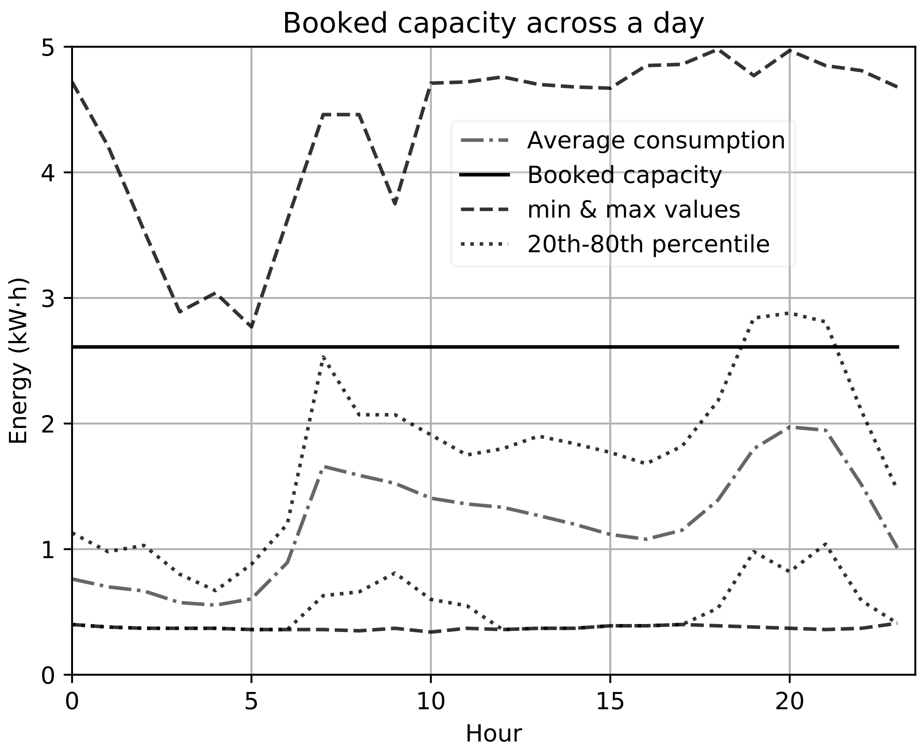

Figure 4.3 Booked capacity profile and load distribution bounds

be minimized with the following linear optimization model:

$$\min_{v,z} \sum_{t=1}^{|T|-1} v_t \tag{4.9a}$$

$$\text{s.t.} \quad v_t \geq \sum_{k \in S_t} c_t^k z_t^k - \sum_{k \in S_t} c_{t+1}^k z_{t+1}^k \qquad \forall\, t \in \{1..|T|-1\} \tag{4.9b}$$

$$v_t \geq \sum_{k \in S_t} c_{t+1}^k z_{t+1}^k - \sum_{k \in S_t} c_t^k z_t^k \qquad \forall\, t \in \{1..|T|-1\} \tag{4.9c}$$

$$\sum_{k \in S_t} z_t^k = 1 \qquad \forall t \in T \tag{4.9d}$$

In this model, $z_t^k = 1$ is equivalent to the supplier choosing the k^{th} candidate for the time frame t. v_t takes as value the absolute error between the capacity at t and $t+1$. The optimum of this model corresponds to a constant booked capacity over all the time frames. A trade-off can be set between this objective and the capacity averaged over each time frame:

$$\sum_{t=1}^{|T|} \sum_{k \in S_t} c_t^k z_t^k \,/\, |T|. \tag{4.10}$$

Different solutions corresponding to various trade-off are obtained by solving Problem (4.11) with a weight w varying as shown in Fig. 4.4. The weight parameter is the weight assigned to the average booked capacity, while the weight of the sum of variations remains at 1.

$$\min_{v,z} \sum_{t=1}^{|T|-1} v_t + w \sum_{t=1}^{|T|} \sum_{k \in S_t} c_t^k z_t^k \,/\, |T| \tag{4.11a}$$

$$\text{s.t } (4.9\text{b - } 4.9\text{d}) \tag{4.11b}$$

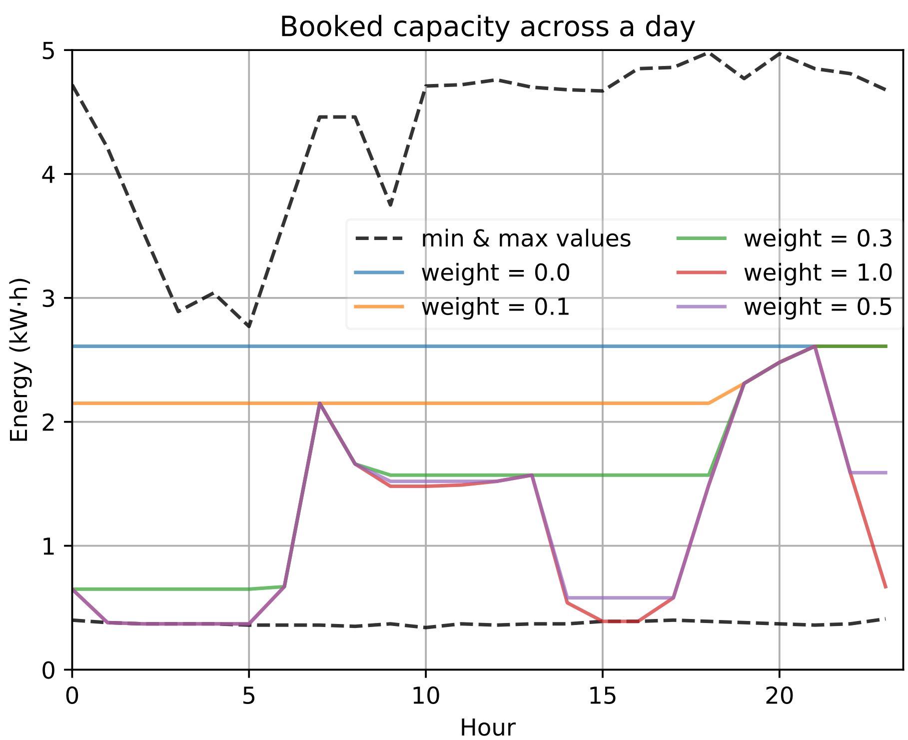

Figure 4.4 Booked capacity profile with the weighted bi-objective model

4.5 Conclusion

In this letter, we present a bilevel model for a supplier optimizing the TLOU pricing, selecting for each capacity candidate the revenue-maximizing price setting. We illustrate the application with a selection of capacities across the day to smooth the capacity curve or to stay close to the average, and thus create an incentive for the customer to consume at this level. Future research will consider a single setting for multiple customers and exploiting continuous probability distributions of the consumption.

CHAPTER 5 OPTIMAL MULTI-USER TIME-AND-LEVEL-OF-USE PRICING

Abstract

Price-based Demand Response leverages the flexibility of users' consumption to accommodate grid requirements and mitigate power generation uncertainty. The Time-and-Level-of-Use (TLOU) pricing extends price-based Demand Response programs with a capacity self-determined by users planning their consumption, providing the supplier with information on the intended consumption. We focus on the optimal pricing of TLOU for a supplier serving heterogeneous users, in the case where the forecast consumption follows a mixed continuous-discrete distribution. We propose bilevel optimization models for the optimal setting of TLOU parameters under different configurations. Under smoothness assumptions of the probability distributions, these models can be simplified as mixed-integer linear optimization problems. The impact of the number of users and configurations on the solution time and attained revenue highlights the scalability of the proposed method.

5.1 Introduction

In modern power grids, distributed renewable generation is a growing source of uncertainty for grid operators, because of power injections characterized by significant variability and limited ability to control. This increased uncertainty may lead to greater difficulty for maintaining the balance between supply and demand at any time, potentially putting the power grid at risk. System operators have been leveraging several mechanisms to cope with these fluctuations, with for instance load curtailment on the generation side or storage [96]. Demand-Side Management (DSM) has focused on the operations of the demand side of the power system [11], with several approaches including Demand Response (DR) to create and exploit changes in time of power consumption.

DR programs have been developed to let the demand-side decision makers support grid operation by creating value from their flexibility [12]. In incentive-based DR programs, changes in the consumption are achieved either from direct activation and deactivation of consuming devices controlled by the power supplier or through the lever of market mechanisms. Price-based DR programs use variations of prices to encourage users to change their consumption through time. These price variations can happen a few times during the day as in Time-of-

Use (TOU), separate a full day from the normal operation in terms of price as in Extreme Day Pricing, or happen on a quasi-continuous basis as in Real-Time Pricing. Price-based DR requires less frequent communication between users and suppliers, and little to no specific hardware [15]. In addition to their simplicity on the user side, these advantages increased the potential of price-based DR for the residential market.

In [27], a Time-and-Level-of-Use (TLOU) pricing policy was introduced, allowing users to determine and book in advance a consumption capacity for a time frame. In this policy, the price of energy depends not only on the time of consumption, similarly to a Time-of-Use (TOU) policy, but also on the capacity booked. In [90], a variant of the TLOU pricing is defined; given an energy consumption X_t at a time frame t and a capacity c_t booked by the user, the total cost for the user under TLOU is computed as:

$$\mathcal{C}(c_t, K, \pi^L, \pi^H) = Kc_t + \begin{cases} X_t \pi^L(c_t) & \text{if } X_t \leq c_t \\ X_t \pi^H(c_t) & \text{otherwise,} \end{cases}$$

where (K, π^L, π^H) are the decisions of the supplier which define the TLOU setting. K is a strictly positive booking fee which, for a booked capacity c, yields a booking cost Kc. $\pi^L(\cdot), \pi^H(\cdot)$ are piecewise constant, $\mathbb{R}_+ \to \mathbb{R}_+$ functions, decreasing and increasing respectively on predefined steps. They can equivalently be viewed as vectors of price steps with decreasing, respectively increasing successive components. The capacity is an amount of energy that can be consumed in the time frame by the user while benefitting from the lower price. However, the grid operator is concerned with power consumption more than energy, and in particular with the maximum power usage. If the time frames are short enough and assuming users do not use appliances consuming at high power for a short period, the energy consumed during a time frame is an acceptable proxy for the maximum power usage in the time frame.

In [90], a bilevel optimization model of the decision-making process of the supplier-user pair is designed, introducing the position of the energy supplier setting the values of the pricing components of TLOU, namely a lower and higher price curve, and a booking fee. Bilevel optimization problems integrate the optimality of a second optimization problem in the constraints. They are often used to represent Stackelberg or Leader-Follower games, where one player, the leader, makes a first decision, to which a second player, the follower, reacts by solving their own optimization problem parameterized by the leader's decision. We refer the interested reader to [97, 98] for recent reviews on the methods and applications of bilevel optimization.

Representing the decision-making process of the two actors as a Stackelberg game integrates the user as a strategic decision-maker with their own objective and set of constraints, avoiding assumptions such as demand elasticity. The supplier, acting as the leader, anticipates the optimal reaction of the user, acting as the follower, who then decides on a capacity to book for a given time frame, taking into account the TLOU parameters set by the supplier. Using bilevel optimization to build a mathematical formulation of the problem, the leader's and the followers' problems correspond to the upper and lower level respectively.

In a realistic setting, the supplier offers a TLOU pricing to multiple users who can be different households, buildings or commercial users within a neighborhood or micro-grid. The supplier can be an aggregator [16] setting the price for multiple users while buying electricity from the wholesale market. For legal or practical reasons, the supplier may want to offer the same price setting to all users at a given time. Indeed, offering different prices to different entities opens a range of issues, including fairness and transparency of the price determination process. A looser restriction on the pricing decision may be to be able to offer a single price setting for all users within a group of users. We consider three ways such groups are determined, they may be based on predefined criteria, such as geographic location, they can be a choice of the supplier (with an assignment of users to groups), or left to each user to choose a given group and the corresponding TLOU setting and then a capacity given this setting.

An assumption underlying the proposed model in [90] is that the probability distribution of the consumption at a given time frame is discrete with finite support. Some devices, however, such as electrical heating systems, consume energy in an intermittent fashion or in a continuous range. The probability distribution of the aggregated household consumption then follows a mixed continuous-discrete univariate probability distribution, i.e. a probability distribution over the real numbers with a continuous density and including mass points (singletons for which the probability measure is non-zero).

The contributions of this chapter are:

- Defining the TLOU pricing model with a consumption following a mixed continuous-discrete probability distribution, and defining conditions under which the optimal candidates of the lower level reduce are contained in a discrete finite set depending only on the problem data;

- Formulating the decision problem of the supplier determining a single TLOU price setting for multiple heterogeneous clients;

- Studying the impact of the clustering of clients in groups receiving a single TLOU price setting, with cluster assignments determined either by the supplier, each user, or given as fixed parameters.

This chapter is structured as follows. Section 5.2 provides the background on TLOU and related work in pricing for Demand Response. Section 5.3 introduces the mixed continuous-discrete probability distribution of the aggregated consumption and the expression of the user cost for given upper- and lower-level decisions. Section 5.4 introduces the formulation of the optimal price-setting problem with multiple users, each characterized by their consumption distribution, resulting in non-tractable bilevel Mixed-Integer Non-Linear Problems. In Section 5.5, these problems are reformulated as Mixed-Integer Linear Problems. In Section 5.6, computational experiments on various implementations of the MILP models are carried out. Section 5.7 concludes the chapter and lays out perspectives for future research.

5.2 Background on TLOU and related work

In this section, we provide the notation and previous work on TLOU DR programs and related research. TLOU was introduced under different angles in [27, 29, 99] which focus on the perspective of a smart home agent optimizing its booked capacity, consumption, and storage planning. The TLOU setting is considered fixed in advance. The studies show how a smart building aggregating multiple users can reduce its operating costs while meeting the user's consumption preferences.

In [90], completed by [94], a price-setting model is built from the supplier perspective to derive a set of TLOU pricing options, with each pricing maximizing the revenue for a given potentially optimal capacity level. The model incorporates the optimal user reaction in a bilevel formulation and anticipates a conservative behaviour of the user, who is assumed to choose a capacity only if it yields an expected cost lower than all alternatives by at least a difference of δ. The solution approach yields an optimal price setting for each capacity candidate which could be optimal for the lower level. The choice among these capacity profiles is further investigated, with a second bi-objective linear optimization model picking a TLOU setting at each time frame of the day, choosing a trade-off between the sum of variations from a time frame to the next (thus encouraging load shifting), and the average booked capacity.

Other approaches for the representation of user choices in the context of price-based DR are followed in [33, 34, 100] and consider the user decision as a trade-off between cost minimization happening through load shifting and inconvenience minimization (or limitation). The formulation of inconvenience depends on the nature of the electric load. A notion of inconvenience based on a comfort temperature zone is considered in [100], since thermal appliances are considered as the flexible consumption unit. In [33, 34], inconvenience is defined in a more generally applicable fashion as the displacement of electric loads in their schedule. In [101], competing retailers offer Time-of-Use prices to end-users, taking into consideration consumption uncertainty and buying electricity from multiple markets. An incentive-based DR program is proposed in [20], where power generators play a Cournot-Nash game. In [102], the authors develop a behaviour model for users making a multistage decision under a peak-time price-based DR program.

The TLOU DR program we focus on in this chapter is primarily price-based, but the dependence of the price on the consumption creates an incentive for the user to remain below the established capacity. In [93], multiple time-based DR programs are compared in a setting where the prosumer operates renewable generation and energy storage. Similarly to our approach, a default option is always offered to the prosumer in addition to the tariff designed by the supplier in order to ensure that the proposed pricing is at least as profitable for the prosumer as the baseline. In [30], a robust optimization model is developed for a multistage manufacturing unit for which electricity is priced in a TLOU scheme. The unit includes its own energy storage system renewable energy sources and can co-optimize its consumption, storage, and the price at which it buys electricity from the supplier from the time and level.

5.3 Single-user formulation with mixed continuous-discrete consumption distribution

In this section, we develop the formulation of the expected total cost for a single user booking a capacity in a TLOU program under a mixed continuous-discrete probability distribution of the consumption. From this formulation, we determine the set of candidates to optimality i.e. a set of capacity levels that contain the minimizer of the expected cost. This set of candidates can then be used in the following section to derive tractable optimality conditions for the user problem. For a non-empty interval $X \subseteq \mathbb{R}$ and a function $f : X \to \mathbb{R}$ continuous almost everywhere on X, we will qualify a point c in the interior of X as a positive discontinuity iff:

- $\lim_{x \to c, x<c} f(x)$ and $\lim_{x \to c, x>c} f(x)$ both exist and are finite;
- $\lim_{x \to c, x<c} f(x) < \lim_{x \to c, x>c} f(x)$.

A negative discontinuity is defined similarly.

Following the definition of the TLOU pricing from [90], the total expected cost at a time frame t for a user u picking a capacity c is:

$$\begin{aligned}\mathcal{C}_u(c, K, \pi^L, \pi^H) = & Kc + \\ & \pi^L(c)\mathbb{P}\left[X_t \le c\right]\mathbb{E}\left[X_t | X_t \le c\right] + \\ & \pi^H(c)\mathbb{P}\left[X_t > c\right]\mathbb{E}\left[X_t | X_t > c\right]\end{aligned}$$

where K is the booking fee, X_t is the random variable describing the consumption at time t and $\pi^L(\cdot), \pi^H(\cdot)$ are the lower and higher price curves respectively. They determine the price of energy if the consumption is below, respectively above the capacity. The ordered steps of the lower and higher price steps π^L, π^H respectively are denoted C^L, C^H respectively. Since we focus on the expected cost for the user in this section, $\mathcal{C}_u(c, K, \pi^L, \pi^H)$ will be shortened to $\mathcal{C}_u(c)$ when the pricing parameters are implied by the context. Since a single time frame is considered in this chapter, the time index t will be dropped when it is implied by the context.

If the probability density function (pdf) f_X is defined and has a support on $x \in [0, \bar{x}]$, with $\bar{x}$ the maximum consumption, the user cost $\mathcal{C}_u(c)$ can be computed as:

$$\mathcal{C}_u(c) = Kc + \int_0^c \pi^L(c) f_X(x) x dx + \int_c^{\bar{x}} \pi^H(c) f_X(x) x dx.$$

However, in the presence of mass points (positive discontinuities in the cumulative density function), the density function must be expressed as a generalized function. We separate the domain into segments where the probability density function is continuous, and mass points with a positive probability between these segments. Let $X_u^m = [x_0, x_1, ..., x_n]$ be the mass points of the consumption distribution X_u for user u, and define the following partition of the mass points created by a capacity $c \geq 0$:

$$
\begin{aligned}
&X_u^-(c) = \{x^m \in X_u^m, x^m \leq c\} = \{x_0, x_1, x_2 ... x_{B(c)}\} \\
&X_u^+(c) = \{x^m \in X_u^m, x^m > c\} = \{x_{B(c)+1}, x_{b+2} ... x_n\} \\
&\text{where: } B(c) = |X_u^-(c)|.
\end{aligned}
$$

We consider without loss of generality that $x_0 \equiv 0$. For a mixed continuous-discrete distribution containing mass points, the pdf can be integrated in the sense of Riemann in a piecewise fashion between the mass points. The user cost at capacity c given the pricing setting (K, π^L, π^H) is:

$$
\begin{aligned}
\mathcal{C}_u(c, K, \pi^L, \pi^H) = &Kc + \\
&\pi^L(c) \sum_{i=0}^{B(c)-1} \left(x_i p_i + \int_{x_i}^{x_{i+1}} f_X(x) x dx \right) + \\
&\pi^L(c) p_{B(c)} x_{B(c)} + \pi^L(c) \int_{x_{B(c)}}^{c} f_X(x) x dx + \\
&\pi^H(c) \int_{c}^{x_{B(c)+1}} f_X(x) x dx + \\
&\pi^H(c) \sum_{i=B(c)+1}^{n-1} \left(x_i p_i + \int_{x_i}^{x_{i+1}} f_X(x) x dx \right) + \\
&\pi^H(c) x_n p_n + \pi^H(c) \int_{x_n}^{\bar{x}} f_X(x) x dx.
\end{aligned}
$$

Considering that $f_X(\cdot)$ is continuous everywhere, $\mathcal{C}_u(\cdot, K, \pi^L, \pi^H)$ presents a discontinuity at a point x only if:

- $x \in C^L$: the point is a step change of the lower price curve $\pi^L(\cdot)$
- $x \in C^H$: the point is a step change of the higher price curve $\pi^H(\cdot)$
- $x \in X^m$: the point is a mass point, given that $B(c)$ increases by one on x.

We introduce the following quantities:

$$
\begin{aligned}
E_u^-(c) = & \sum_{i=0}^{B(c)-1} \left(x_i p_i + \int_{x_i}^{x_{i+1}} f_X(x) x dx \right) + \\
& p_{B(c)} x_{B(c)} + \int_{x_{B(c)}}^{c} f_X(x) x dx \\
E_u^+(c) = & \int_{c}^{x_{B(c)+1}} f_X(x) x dx + \\
& \sum_{i=B(c)+1}^{n-1} \left(x_i p_i + \int_{x_i}^{x_{i+1}} f_X(x) x dx \right) + \\
& x_n p_n + \int_{x_n}^{\bar{x}} f_X(x) x dx.
\end{aligned}
$$

They correspond to the two components of the expected value of the consumption:

$$
\begin{aligned}
& E_u^-(c) = P[X \leq c] \cdot \mathbb{E}[X | X \leq c] \\
& E_u^+(c) = P[X > c] \cdot \mathbb{E}[X | X > c] \\
& \mathbb{E}[X] = E_u^-(c) + E_u^+(c).
\end{aligned}
$$

The user cost $\mathcal{C}_u(c)$ computed at any c can be rewritten as:

$$
\mathcal{C}_u(c) = Kc + \pi^L(c) \cdot E_u^-(c) + \pi^H(c) \cdot E_u^+(c).
$$

Proposition 5.3.1. *The set of optimality candidates for a given user u is included in:*

$$
\hat{C} = \{0\} \cup X_u^m \cup C^L \cup C^H \cup S_{cont}
$$

where

$$
S_{cont} = \left\{ c > 0 \;\middle|\; f_X(c) = \frac{K}{(\pi^H(c) - \pi^L(c))c} \text{ and } \frac{df_X}{dc}(c) \leq -\frac{f_X(c)}{c} \right\}.
$$

Proof. The expected cost is piecewise continuous, with discontinuities on the mass points X_u^m and steps C^L, C^H of the price functions $\pi^L(\cdot), \pi^H(\cdot)$. A necessary condition for a capacity to be optimal in a continuous segment is a zero first derivative. Given that the point is assumed to be in the interior of a continuous segment, the price curves both have derivative 0 and

$$
\exists \epsilon,\ \pi^{L/H}(c+\epsilon) = \pi^{L/H}(c)
$$

for ϵ small enough. We can express the corresponding variation of the pricing components:

$$
\begin{aligned}
E_u^-(c+\epsilon) - E_u^-(c) &= \int_{x_b}^{c+\epsilon} f_X(x)xdx - \int_{x_b}^{c} f_X(x)xdx \;\Leftrightarrow \\
E_u^-(c+\epsilon) - E_u^-(c) &= \int_{x_b}^{c+\epsilon} f_X(x)xdx + \int_{c}^{x_b} f_X(x)xdx \;\Leftrightarrow \\
E_u^-(c+\epsilon) - E_u^-(c) &= \int_{c}^{c+\epsilon} f_X(x)xdx
\end{aligned}
$$

for the lower price component, and

$$
\begin{aligned}
E_u^+(c+\epsilon) - E_u^+(c) &= \int_{c+\epsilon}^{x_{B(c)}} f_X(x)xdx - \int_{c}^{x_{B(c)}} f_X(x)xdx \;\Leftrightarrow \\
E_u^+(c+\epsilon) - E_u^+(c) &= \int_{c}^{x_{B(c)}} f_X(x)xdx - \int_{c}^{c+\epsilon} f_X(x)xdx - \int_{c}^{x_{B(c)}} f_X(x)xdx \;\Leftrightarrow \\
E_u^+(c+\epsilon) - E_u^+(c) &= -\int_{c}^{c+\epsilon} f_X(x)xdx
\end{aligned}
$$

for the higher price component. We can deduce the expression of the derivative of the cost:

$$
\begin{aligned}
\frac{\partial \mathcal{C}}{\partial c}(c, K, \pi^L, \pi^H) &= K + \pi^L(c)\frac{\partial E_u^-}{\partial c}(c) + \pi^H(c)\frac{\partial E_u^+}{\partial c}(c) \\
&= K + \pi^L(c) f_X(c)c - \pi^H(c) f_X(c)c.
\end{aligned}
$$

Setting the derivative to 0 yields:

$$
0 = K + c\pi^L(c) f_X(c) - c\pi^H(c) f_X(c) \;\Leftrightarrow\; f_X(c) = \frac{K}{(\pi^H(c) - \pi^L(c))c}
$$

In addition to the first-order condition developed above, the necessary second-order condition is that the second derivative must be positive or zero for local optimality:

$$
\frac{\partial^2 \mathcal{C}}{\partial c^2}(c) = (\pi^L(c) - \pi^H(c)) \left(c\frac{\partial f_X}{\partial c}(c) + f_X(c) \right) \geq 0
$$

$$
\Leftrightarrow
$$

$$
(\pi^H(c) - \pi^L(c)) \left(c\frac{\partial f_X}{\partial c}(c) + f_X(c) \right) \leq 0.
$$

If $\pi^L(c) - \pi^H(c) = 0$, then all higher-order derivatives of $\mathcal{C}_u(\cdot)$ are zero. This condition is met only on the segment before the first steps of both the lower and higher prices. In this segment, booking any capacity different from 0 is suboptimal, since it does not change the price of energy from the baseline while inducing some booking fee. We therefore exclude this condition and focus on the second derivative when $\pi^L(c) < \pi^H(c)$, which also implies $c > 0$:

$$c\frac{df_X}{dc}(c) + f_X(c) \leq 0 \;\Leftrightarrow\; \frac{df_X}{dc}(c) \leq -\frac{f_X(c)}{c}.$$

The second-order necessary conditions for a candidate capacity c to be a local optimum are therefore:

$$\begin{cases} \frac{\partial \mathcal{C}_u}{\partial c}(c) = 0 \\ \frac{\partial^2 \mathcal{C}_u}{\partial^2 c}(c) \geq 0 \end{cases} \Leftrightarrow \begin{cases} f_X(c) = \frac{K}{c\,(\pi^H(c) - \pi^L(c))} \\ \frac{df_X}{dc}(c) \leq -\frac{f_X(c)}{c} \end{cases}$$

which correspond to the definition of S_{cont}. □

Proposition 5.3.1 defines two sets of candidate solutions for optimality. The candidates from the first set are independent of the price settings and can be computed using only problem data while the other candidates from S_{cont} in the continuous segments of the cost functions depend on the pricing decisions (K, π^L, π^H). However, under some assumptions on the density function, we can assume the necessary conditions for the existence of candidates in S_{cont} are never met. The first-order conditions require that $f_X(\cdot)$ reaches the quantity $\frac{K}{c\,(\pi^H(c) - \pi^L(c))}$ for some c which corresponds to a high-density concentration. If the density remains below this threshold, first-order optimality conditions are never met. The second-order optimality conditions require a "sharp-enough" decrease of the density at a point where first-order optimality conditions are satisfied, such that the derivative is below $-\frac{f_X(c)}{c}$. In the case where a sharp-enough spike in the density meeting the optimality condition is observed, it can also be viewed as a mass point and removed from the continuous segment.

In order to derive a tractable formulation of the pricing problem, we will rely in the rest of this chapter on the assumption that the necessary optimality conditions are never met on the continuous segments of the cost function, and that the optimum is located on one of the discontinuities. We denote this restricted set of capacity candidates for user u as:

$$S_u = \{0\} \cup C^L \cup C^H \cup X_u^m.$$

If a user books a given capacity c, the lower price is applied for a consumption below or equal to c. Given a point x such that $\mathcal{C}(\cdot, K, \pi^L, \pi^H)$ is discontinuous at x, booking a capacity of $x - \epsilon$ or a capacity $x + \epsilon$ with $\epsilon > 0$ arbitrarily small does not yield the same expected cost. For any discontinuity, the left and right limits are both optimality candidates. The following proposition however establishes which of the two limits is potentially optimal for a given discontinuity.

Proposition 5.3.2. *For a capacity candidate $k \in S_u$, the optimal capacity can only be the left limit $\lim_{c \to c_k, c < c_k} c$ if $k \in C^H$, and always c_k itself when $k \in C^L \cup X^m$.*

Proof. The pdf $f_X(\cdot)$ is differentiable at any point $x \notin X_u^m$. Given c and $\epsilon > 0$ such that

$$\begin{aligned}
&c + \epsilon \in C^L, \\
&X_u^m \cap [c, c+\epsilon] = \emptyset, \\
&C^H \cap [c, c+\epsilon] = \emptyset,
\end{aligned}$$

then the following holds true:

$$\begin{aligned}
&E^-(c+\epsilon) - E^-(c) = E^+(c) - E^+(c+\epsilon) \approx \epsilon\, c f_X(c) \\
&\pi^H(c+\epsilon) = \pi^H(c) \\
&\pi^L(c+\epsilon) < \pi^L(c) \\
&\mathcal{C}(c+\epsilon) = K\ (c+\epsilon) + \pi^L(c+\epsilon)E^-(c+\epsilon) + \pi^H(c+\epsilon)E^+(c+\epsilon) \\
&\mathcal{C}(c+\epsilon) - \mathcal{C}(c) \approx E^-(c)\left[\pi^L(c+\epsilon) - \pi^L(c)\right] < 0.
\end{aligned}$$

Given that the discontinuity is negative, the optimality candidate is on the right-hand side of the discontinuity. Similarly, given c and $\epsilon > 0$ such that:

$$\begin{aligned}
&c + \epsilon \in C^H, \\
&X_u^m \cap [c, c+\epsilon] = \emptyset, \\
&C^L \cap [c, c+\epsilon] = \emptyset,
\end{aligned}$$

then the following holds:

$$\begin{aligned}
&\pi^H(c+\epsilon) > \pi^H(c) \\
&\pi^L(c+\epsilon) = \pi^L(c) \\
&\mathcal{C}(c+\epsilon) = K\,(c+\epsilon) + \pi^L(c+\epsilon)E^-(c+\epsilon) + \pi^H(c+\epsilon)E^+(c+\epsilon) \\
&\mathcal{C}(c+\epsilon) - \mathcal{C}(c) \approx E^+(c)\left[\pi^H(c+\epsilon) - \pi^L(c)\right] > 0.
\end{aligned}$$

The discontinuity at $c \in C^H$ is positive, so the optimality candidate is the limit on the left-hand side of the discontinuity.

Lastly, on a mass point, a quantity is transferred from $E^+(c)$ to $E^-(c)$, while the lower and higher prices remain constant. The discontinuity is therefore negative, and the optimality candidate is on the right-hand side. $\square$

Based on Proposition 5.3.2, we consider the optimality candidates to be all the possible values of:

$$c_{uk} = \begin{cases} \lim_{c \to \bar{c}, c < \bar{c}} c & \text{if } \bar{c} \in C^H \\ \lim_{c \to \bar{c}, c > \bar{c}} \equiv \bar{c} & \text{if } \bar{c} \in C^L \cup X_u^m. \end{cases}$$

If

$$\exists x_0 \in C^H \cap (X_u^m \cup C^L), \tag{5.1}$$

then one candidate belonging to C^H is considered on the right side of x_0 and a second candidate is considered on the left side, corresponding to a point in $C^L \cup X_u^m$.

5.4 Supplier problem formulation with multiple users

In this section, we formulate the bilevel problem of the supplier setting the TLOU parameters with multiple users booking a capacity minimizing their expected cost as developed in Section 5.3. We establish three variants of the price-setting problem depending on the way users are clustered. Similarly to [90], the supplier maximizes the total revenue which is expressed as the sum of revenues from each user. Any user u involved in the TLOU program will book a capacity $c_u > 0$ at a time frame only if the resulting expected cost is better than the baseline. This induces a net financial loss through the program from the supplier perspective. The total revenue for the supplier is the sum of costs paid by users:

$$\sum_{u \in \mathcal{U}} \mathcal{C}_u(c; \pi^L, \pi^H, K),$$

with $\mathcal{U}$ the set of N_U users. If the supplier maximizes only their revenue, all users booking a capacity of 0 results in an optimal solution for the supplier. In any other case, one user u booking a positive capacity implies that their expected cost at this capacity is lower than the baseline of capacity 0, since a reduction in expected cost for a user implies a reduction in expected revenue. In order to smooth out the total demand curve or bring a certain guarantee on the consumption, the supplier may want to create an incentive for users to book a non-zero capacity, thus committing to consuming within a certain limit in the considered time frame. In [90], the supplier constructs a set of options at each capacity level $c \in S_u$, building for each one a set of pricing parameters such that each combination of capacity and pricing is optimal.

When serving multiple users, the supplier wants to influence the aggregated consumption of all users. Thus, we add the following constraint at the upper level capturing the willingness to bound the sum of capacities booked by all users:

$$\underline{c} \leq \sum_{u \in \mathcal{U}} c_u \leq \bar{c}.$$

Finally, the problem for each user is to minimize their expected cost with respect to the capacity they book:

$$\min_{c \geq 0} \; \mathcal{C}_u(c, \pi^L, \pi^H, K) \qquad \forall u \in \mathcal{U}.$$

The supplier problem is expressed as follows:

$$\max_{K,\pi^L,\pi^H} \; \sum_{u \in \mathcal{U}} \mathcal{C}_u(c_u, \pi^L, \pi^H, K) \tag{5.2a}$$

$$\text{s.t.} \;\; c_u \in \arg\min_{c \geq 0} \; \mathcal{C}_u(c, \pi^L, \pi^H, K) \qquad \forall u \in \mathcal{U} \tag{5.2b}$$

$$\underline{c} \leq \sum_{u \in \mathcal{U}} c_u \leq \bar{c} \tag{5.2c}$$

$$(K, \pi^L, \pi^H) \in \Phi, \tag{5.2d}$$

where Φ is a polytope restricting the possible set of values for (K, π^L, π^H) by defining lower and upper bounds on the values of the price parameters and on the difference between successive steps of π^L and π^H. This optimization model casts the TLOU pricing as a bilevel problem, by making c_u the optimal solution of the lower-level Problem (5.2b) for each user $u \in \mathcal{U}$.

On most power systems, users are not considered as a homogeneous group but are segmented using various criteria, with different segments being presented with different offers. We consider three variants of Problem (5.2). In the first variant, which we denote (Fixed-TLOU), given N_G groups, the user clustering is described as a set of boolean parameters:

$$p_{um} = \begin{cases} 1 \text{ if user } u \text{ in group } m, \\ 0 \text{ otherwise} \end{cases} \qquad \forall u \in \mathcal{U}, m \in \{1..N_G\},$$

with all users assigned to exactly one of the N_G disjoint sets, and the supplier choosing a price setting for each group $m \in \{1..N_G\}$. The resulting model is:

$$
\begin{aligned}
\text{(Fixed-TLOU)} \quad & \max_{K,\pi^L,\pi^H} \sum_{u\in\mathcal{U}} \mathcal{C}_u(c_u, \pi_u^L, \pi_u^H, K_u) && && (5.3a)\\
\text{s.t.} \quad & c_u \in \arg\min_{c\geq 0} \ \mathcal{C}_u(c, \pi_u^L, \pi_u^H, K_u) && \forall u \in \mathcal{U} && (5.3b)\\
& \pi_u^L = \sum_{m=1}^{N_G} \pi_m^L p_{um} && \forall u \in \mathcal{U} && (5.3c)\\
& \pi_u^H = \sum_{m=1}^{N_G} \pi_m^H p_{um} && \forall u \in \mathcal{U} && (5.3d)\\
& K_u = \sum_{m=1}^{N_G} K_m p_{um} && \forall u \in \mathcal{U} && (5.3e)\\
& \underline{c} \leq \sum_{u\in\mathcal{U}} c_u \leq \bar{c} && && (5.3f)\\
& (K_m, \pi_m^L, \pi_m^H) \in \Phi && \forall m \in \{1..N_G\}. && (5.3g)
\end{aligned}
$$

There is one set of pricing decisions for each of the N_G clusters, while the pricing assigned to a user is a direct linear expression of these decisions.

In a second variant denoted (Supplier-TLOU), the decisions of the problem include the segmentation and assignment of users in N_G groups, with users in each of those groups being offered different prices. It is equivalent to (Fixed-TLOU) with p_{um} becoming a decision variable for the assignment of user u to a group m_u:

$$
\begin{aligned}
\text{(Supplier-TLOU)} \quad & \max_{K,\pi^L,\pi^H,p_{um}} \ \sum_{u\in\mathcal{U}} \mathcal{C}_u(c_u, \pi_u^L, \pi_u^H, K_u) && &&\\
\text{s.t.} \quad & (5.3b) - (5.3g) && &&\\
& p_{um} \in \{0,1\}, && &&\\
& \sum_{m\in\{1..N_G\}} p_{um} = 1 && \forall u \in \mathcal{U}. && (5.4)
\end{aligned}
$$

Remark 5.4.1. *(Supplier-TLOU) is a relaxation of (Fixed-TLOU).*

Proof. Any solution feasible for (Fixed-TLOU) defines an assignment of users to clusters, and is also feasible for (Supplier-TLOU), with an identical objective value. □

In many configurations, the user will be able to pick the tariff group that best fits their needs, instead of being assigned based on fixed conditions or by the supplier. This corresponds to a situation identical to (Supplier-TLOU), but where the group assignment has to be optimal as a user choice. We denote this variant (Self-TLOU):

(Self-TLOU)

$$\begin{aligned}
\max_{K,\pi^L,\pi^H} \; & \sum_{u\in\mathcal{U}} \mathcal{C}_u(c_u, \pi^L_u, \pi^H_u, K_u) \\
\text{s.t.} \;\; & (c_u, p_{um}) \in \operatorname*{arg\,min}_{c\geq 0, p_{um}\in\{0,1\}} \{\mathcal{C}_u(c, \pi^L_u, \pi^H_u, K_u) \text{ s.t. } (5.3c) - (5.3e), (5.4)\} \quad \forall u \in \mathcal{U} \\
& \underline{c} \leq \sum_{u\in\mathcal{U}} c_u \leq \bar{c} \\
& (K_m, \pi^L_m, \pi^H_m) \in \Phi \qquad \forall m \in \{1..N_G\}
\end{aligned}$$

Remark 5.4.2. *(Supplier-TLOU) is a relaxation of (Self-TLOU).*

Proof. (Self-TLOU) has the additional constraint that the cluster assigned to a user must be optimal for the user, (Supplier-TLOU) is, therefore, a relaxation of (Self-TLOU). □

5.5 Mixed-integer formulation

The price-setting problems as presented in Section 5.4 cannot be solved directly using branch-and-bound based methods. More specifically, the lower-level optimality constraint (5.2b) includes an arg-min operator and is not in closed form. In this section, we reformulate the three price-setting variants as MILPs. Under the regularity conditions of the continuous density established in Section 5.3, the lower-level optimality candidates of each user are known in advance and are independent of decision variables. These can be used to reformulate the lower-level optimality as the lowest-cost choice in a discrete finite set.

(TLOU-Fixed) is parameterized by $(\underline{c}, \bar{c}, \delta_u, m_u)$ and noted $\mathcal{P}(\underline{c}, \bar{c}, \delta_u, m_u)$. Its MILP formulation (MFT) is:

$$
\begin{aligned}
\text{(MFT)} \quad & \max_{z,K,\pi^L,\pi^H} \sum_{u\in\mathcal{U}} \mathcal{C}_u && && \text{(5.5a)}\\
\text{s.t.} \quad & z_{uk} \Rightarrow \mathcal{C}_u \leq \mathcal{C}_u(C_{ul}, K_u, \pi_u^L, \pi_u^H) - \delta_u && \forall u\in\mathcal{U}, \forall k, l \in S_u, l \neq k && \text{(5.5b)}\\
& z_{uk} \Rightarrow \mathcal{C}_u \leq \mathcal{C}_u(0) - \delta_u && \forall u\in\mathcal{U}, \forall k \in S_u && \text{(5.5c)}\\
& (5.3c) - (5.3e) && && \\
& c_u = \langle z_u, C_u \rangle && \forall u\in\mathcal{U} && \text{(5.5d)}\\
& \sum_{j\in S_u} z_{uj} \leq 1 && \forall u\in\mathcal{U} && \text{(5.5e)}\\
& \underline{c} \leq \sum_{u\in\mathcal{U}} c_u \leq \overline{c} && && \text{(5.5f)}\\
& (K, \pi^L, \pi^H) \in \Phi && && \text{(5.5g)}\\
& z_{uj} \in \{0,1\} && \forall u\in\mathcal{U},\ \forall k \in S_u && \text{(5.5h)}\\
\text{where} & && && \text{(5.5i)}\\
& \mathcal{C}_u \equiv \mathcal{C}_u(c_u, \pi_u^L, \pi_u^H, K_u).
\end{aligned}
$$

Constraints (5.5b-5.5c) are indicator constraints expressing a logical implication, the notation used is:

$$y \Rightarrow \langle a, x \rangle \leq b$$

requiring the linear constraint $\langle a, x \rangle \leq b$ to be valid if the binary variable y is equal to one. Constraints (5.5b-5.5c) capture the optimality conditions of the lower level: a capacity, to be selected as a solution, must result in an expected cost lower than all other candidates S_u.

Many modern MILP solvers support indicator constraints out-of-the-box. Alternatively, those constraints can be reformulated using either Special Ordered Sets of type 1 (SOS1) or big-M. A big-M reformulation is possible because the quantities implied are bounded:

$$\langle a, x \rangle \leq b + M(1 - y).$$

The user cost is always positive and never higher than the baseline cost $C_u(0)$, this baseline can be used as an upper bound on the expected revenue generated by each user. c_u and C_u are linear expressions rather than variables, reused in different constraints.

The second model of interest includes the optimal clustering by the supplier, with the number of groups N_G as a parameter, introduced as (TLOU-Supplier) in the previous section. p_{um} is a variable in this model, which we formulate as follows:

$$
\begin{aligned}
\text{(MSupT)} \quad & \max_{z,K,\pi^L,\pi^H,p} \sum_{u\in\mathcal{U}} \mathcal{C}_u && && \text{(5.6a)}\\
\text{s.t.} \quad & z_{uk} \Rightarrow \mathcal{C}_u \le \mathcal{C}_u(0) - \delta_u && \forall u \in \mathcal{U}, \forall k \in S_u && \text{(5.6b)}\\
& z_{uk} \,\&\, p_{um} \Rightarrow \mathcal{C}_u \le \mathcal{C}_u(C_{ul}, K_m, \pi^L_m, \pi^H_m) - \delta_u && \forall u, \forall (k,l), l \neq k && \text{(5.6c)}\\
& c_u = \langle z_u, C_u \rangle && \forall u \in \mathcal{U} && \text{(5.6d)}\\
& \sum_{k\in S_u} z_{uk} \le 1 && \forall u \in \mathcal{U} && \text{(5.6e)}\\
& \sum_{m\in\mathcal{G}} p_{um} = 1 && \forall u \in \mathcal{U} && \text{(5.6f)}\\
& \underline{c} \le \sum_{u\in\mathcal{U}} c_u \le \overline{c} && && \text{(5.6g)}\\
& (K, \pi^L, \pi^H) \in \Phi && && \text{(5.6h)}\\
& z_{uk}, p_{um} \in \{0,1\} && \forall u, k, m && \text{(5.6i)}
\end{aligned}
$$

Constraint (5.6c) is activated only if $z_{uk} = p_{um} = 1$, and expresses the lower-level optimality constraint: the cost for the user at the chosen candidate k must be lower than other candidates l subject to the price of the same cluster m. Since most solvers require a single binary variable for indicator constraints, additional binary variables encoding $z_{uk} \,\&\, p_{um}$ must be added to the model.

Finally, we consider the pricing problem where the supplier creates a price setting for each user group $m \in N_G$, but each user is free to choose the group they belong to, as formulated in (Self-TLOU). This setting corresponds to the situation where each user can pick their own group, formulated in the following MILP denoted (MSelfT).

$$
\begin{aligned}
\text{(MSelfT)} \quad & \max_{z,K,\pi^L,\pi^H} \sum_{u\in\mathcal{U}} \mathcal{C}_u && && \text{(5.7a)}\\
\text{s.t.} \quad & z_{uk}^m \Rightarrow \mathcal{C}_u \le \mathcal{C}_u(0) - \delta_u && \forall u \in \mathcal{U}, \forall k \in S_u, m \in \mathcal{G} && \text{(5.7b)}\\
& z_{uk}^m \Rightarrow \mathcal{C}_u \le \mathcal{C}_u(C_{ul}, K_m, \pi_m^L, \pi_m^H) - \delta_u && \forall u,k,l,m; l \ne k && \text{(5.7c)}\\
& c_u = \sum_{m\in\mathcal{G}} \langle z_u^m, C_u \rangle && \forall u \in \mathcal{U} && \text{(5.7d)}\\
& \sum_{k\in S_u} z_{uk}^m \le 1 && \forall u \in \mathcal{U} && \text{(5.7e)}\\
& \underline{c} \le \sum_{u\in\mathcal{U}} c_u \le \overline{c} && && \text{(5.7f)}\\
& (K_m, \pi_m^L, \pi_m^H) \in \Phi && \forall m && \text{(5.7g)}\\
& z_{uk}^m \in \{0,1\} && \forall u,k,m && \text{(5.7h)}
\end{aligned}
$$

These three MILP formulations can be implemented and solved to proven global optimality using off-the-shelf branch and cut software as carried out in the next section.

5.6 Computational experiments

In this section, we present the implementation and computational experiments on the three models presented above. The original dataset [103] was collected from frequent consumption measurements of a household for several years. We generate hourly distributions from the experimental dataset by detecting mass points and using the residual consumption to estimate the continuous distribution. User consumption distributions are generated by randomly moving some probability masses and continuous observation points with a given budget. The continuous distribution is obtained in two steps. First, a Gaussian mixture is fitted with one initial Gaussian per data point, thus creating a Kernel Density Estimation of the distribution. Second, this mixture is truncated above 0, to account for the non-negative nature of the consumption. This data-generation process is implemented in Julia v1.4 [104], the construction of the truncated Gaussian mixture using the Distributions.jl library [105]. The mean of the continuous distribution and integrals over the different segments utilize the QuadGK.jl [106] library based on the Gauss-Kronrod quadrature. The different MILP problems are implemented using JuMP v0.21 [107, 108] with CPLEX v12.9 as the underlying MIP solver. We use a Debian 4.19.67 machine with 64G in RAM capacity and 30 Intel Xeon 2GHz CPUs. CPLEX is set on its single-threaded mode with default parameters.

We investigate the evolution of the time taken to build and solve the instances, and the number of feasible instances when both the number of clusters and the number of users vary. A timeout limit is set at one hour. The lower-level optimality constraints are introduced either in their big-M formulation (*standard*), as lazy constraints activated through a callback (*lazy*), or using indicator constraints (*indicator*). The detailed runtimes for each variant, number of users and clusters are presented in Fig. 5.1,Fig. 5.2, and Fig. 5.3, while the aggregated performance profiles are highlighted in Fig. 5.4.

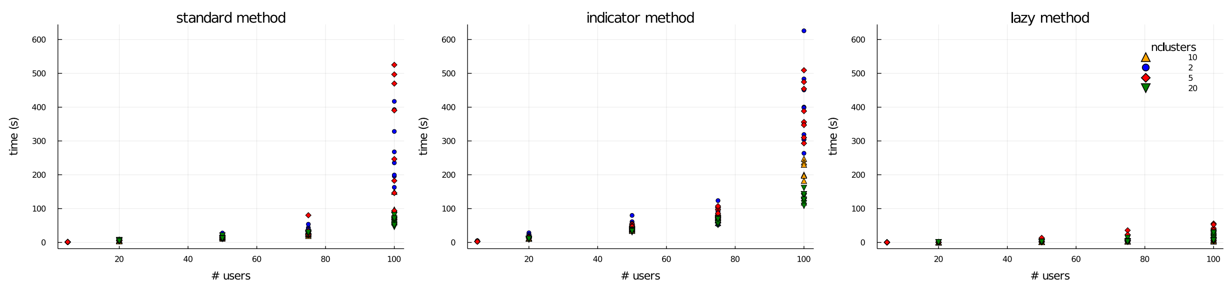

Figure 5.1 Computation time for predetermined clusters

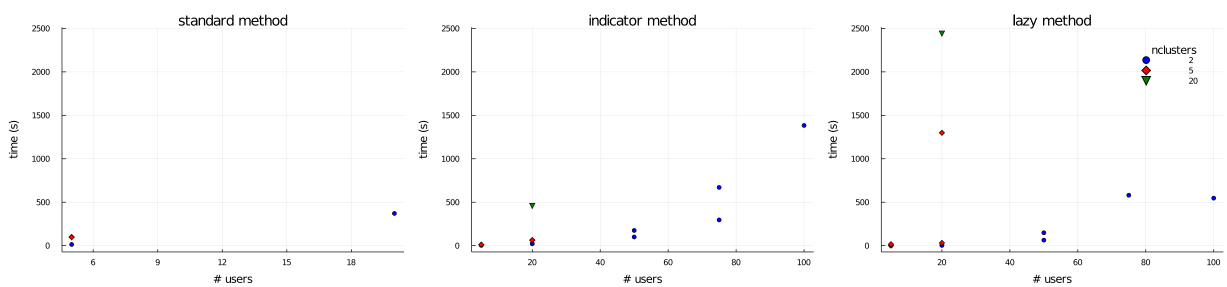

Figure 5.2 Computation time for clusters decided by the supplier

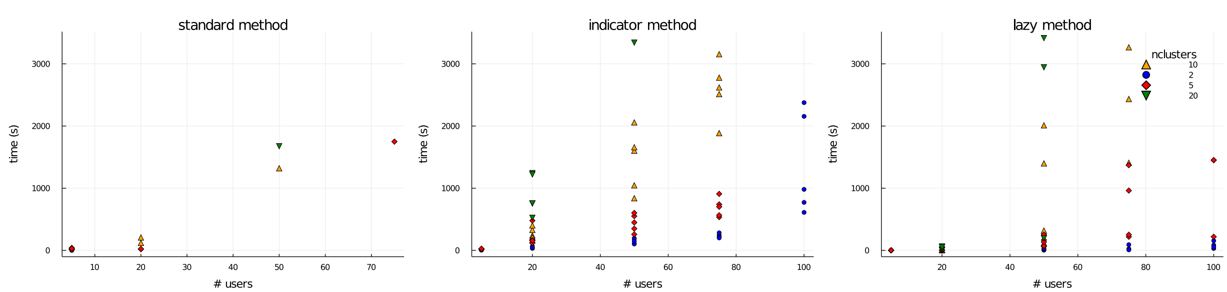

Figure 5.3 Computation time for cluster decided by each user

A larger number of users not only makes the model computationally more challenging, but also reduces the fraction of feasible problems for all configurations and solution methods.

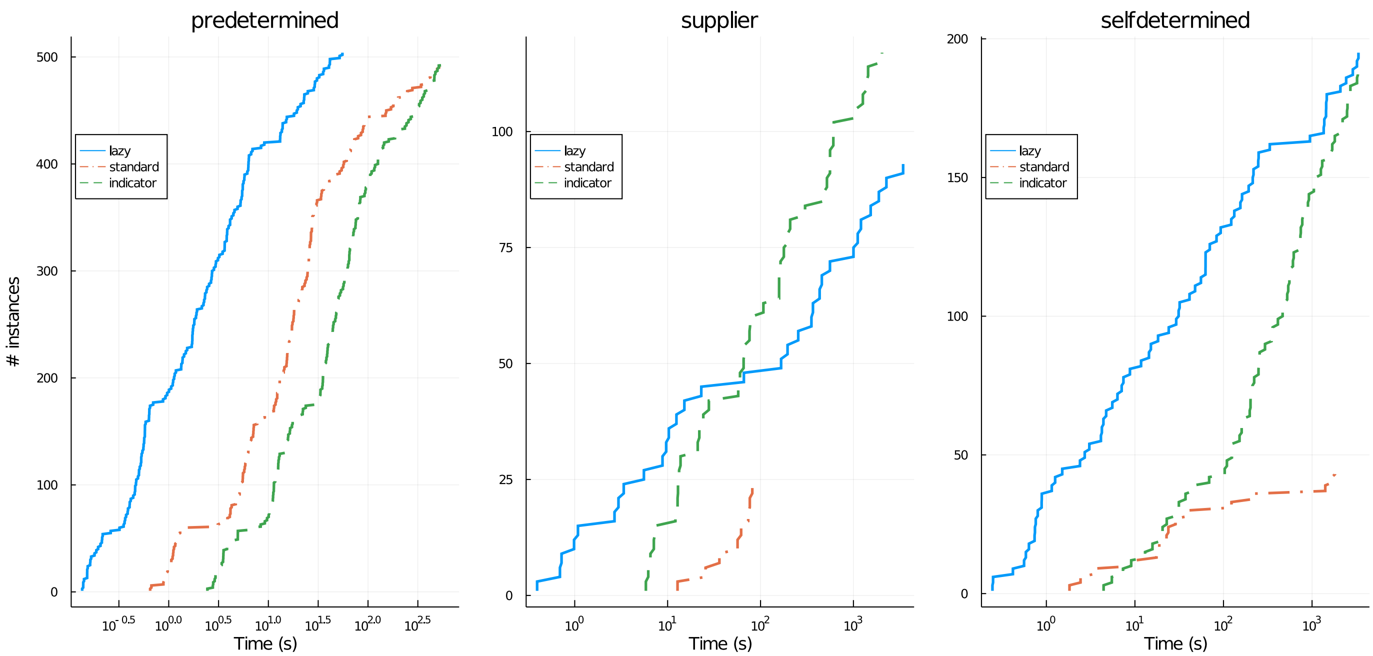

Figure 5.4 Performance profiles of each technique by variant

Indeed, even when the assignment of users to a cluster is done by the supplier, a valid solution must include a capacity choice for each additional user, which is optimal for them while accommodating the supplier constraints. We also observe that models with predetermined clusters are solved in a shorter time across all methods, that can be explained by the reduced number of variables and constraints, since the assignment is a parameter and does not need to be encoded as a set of binary variables. The lazy constraint approach scales best as the number of users increases but exhibits outliers for which the solving time is comparatively higher, as in Fig. 5.2 where the indicator formulation solves the problems at 20 users in less than 500 seconds, while the indicator approach takes up to 2500 seconds for 20 clusters. Since all lazy constraints correspond to an upper bound on the user cost, higher computational time may imply that the callback was called multiple times and added many constraints to the model before convergence.

Tables 5.1-5.3 show the number of successfully optimized, infeasible, and timed-out instances for different numbers of clusters, users, and for the three methods. The experiments are run only for a number of users greater or equal to the number of clusters. The number of timed out instances increases with both the number of users and clusters. The solving time for the configuration where user groups are pre-assigned is much shorter at an equivalent problem size than the configurations where the assignment is a user or supplier decision.

Table 5.1 Feasible/Infeasible/Timed-out instances, predetermined case (MFT)

	method	standard				indicator				lazy			
	nclusters	2	5	10	20	2	5	10	20	2	5	10	20
nusers	5	10/0/0	10/0/0	-	-	10/0/0	10/0/0	-	-	10/0/0	10/0/0	-	-
	20	10/0/0	10/0/0	8/2/0	10/0/0	9/1/0	10/0/0	10/0/0	10/0/0	10/0/0	10/0/0	10/0/0	10/0/0
	50	9/1/0	10/0/0	10/0/0	9/1/0	9/1/0	7/3/0	10/0/0	10/0/0	10/0/0	9/1/0	9/1/0	9/1/0
	75	9/1/0	10/0/0	9/1/0	7/3/0	9/1/0	10/0/0	9/1/0	8/2/0	9/1/0	10/0/0	6/4/0	9/1/0
	100	9/1/0	10/0/0	6/4/0	8/2/0	10/0/0	9/1/0	7/3/0	8/2/0	8/2/0	10/0/0	10/0/0	9/1/0

Table 5.2 Feasible/Infeasible/Timed-out instances, supplier case (MSupT)

	method	standard				indicator				lazy			
	nclusters	2	5	10	20	2	5	10	20	2	5	10	20
nusers	5	3/2/0	5/0/0	-	-	5/0/0	5/0/0	-	-	5/0/0	5/0/0	-	-
	20	0/5/0	0/3/2	0/3/2	0/3/2	4/1/0	5/0/0	4/1/0	5/0/0	5/0/0	5/0/0	0/0/5	0/1/4
	50	0/5/0	0/5/0	0/4/1	0/3/2	5/0/0	0/0/5	0/0/5	0/1/4	5/0/0	0/0/5	0/0/5	0/0/5
	75	0/5/0	0/4/1	0/5/0	0/3/2	3/1/1	0/0/5	0/0/5	0/0/5	3/1/1	0/1/4	0/0/5	0/1/4
	100	0/5/0	0/5/0	0/5/0	0/4/1	3/0/2	0/0/5	0/0/5	0/0/5	3/0/2	0/1/4	0/0/5	0/1/4

Table 5.3 Feasible/Infeasible/Timed-out instances, selfdetermined case (MSelfT)

	method	standard				indicator				lazy			
	nclusters	2	5	10	20	2	5	10	20	2	5	10	20
nusers	5	3/2/0	5/0/0	-	-	5/0/0	5/0/0	-	-	5/0/0	5/0/0	-	-
	20	0/5/0	2/3/0	2/3/0	0/4/1	4/1/0	5/0/0	4/1/0	5/0/0	5/0/0	5/0/0	5/0/0	4/1/0
	50	0/5/0	0/5/0	1/4/0	1/3/1	5/0/0	5/0/0	5/0/0	1/1/3	5/0/0	5/0/0	4/0/1	3/0/2
	75	0/5/0	1/4/0	0/5/0	0/5/0	4/1/0	5/0/0	5/0/0	0/0/5	4/1/0	4/1/0	3/0/2	0/1/4
	100	0/5/0	0/5/0	0/5/0	0/5/0	5/0/0	0/0/5	0/0/5	0/0/5	5/0/0	3/1/1	0/0/5	0/1/4

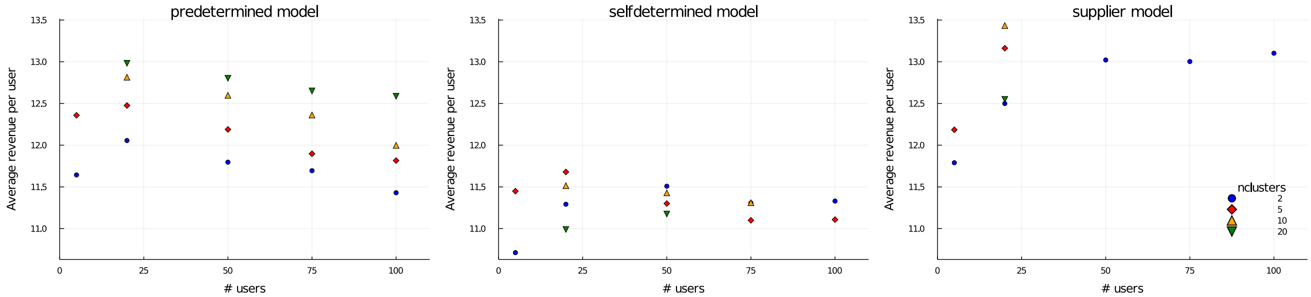

Figure 5.5 Supplier revenue per user

The evolution of the supplier revenue per user is reported in Fig. 5.5. In the self-determined configuration, the supplier revenue decreases as the number of users or the number of clusters increase. In the self-determined configuration, the average revenue decreases with the number of clusters but does not evolve much with the number of users. The average revenue increases with the number of clusters and increases with the number of users for two clusters. We cannot draw any conclusion on the evolution of the revenue for greater numbers of clusters and users for this configuration since all instances timed out or were found to be infeasible.

An observation that can be made for all three models is that the solution maintains a sparsity of capacities, in the sense that only a small part of users have a non-zero capacity, as illustrated Fig. 5.6. Most solutions maintain a number of active users below 4, even in the cases where $N_u = 100$.

We investigate the evolution of optimal user capacities with the upper and lower bounds with a single TLOU pricing for five users. The results are depicted in Fig. 5.7 and also highlight a sparsity of the solutions, with most users remaining at a zero capacity except for instances where the interval between the lower and upper bounds is at its tightest, such as instances 9 and 14. Overall, the sum of capacities remains close to or equal to the lower bound.

In the last experiment, we investigate the effectiveness of symmetry-breaking techniques for Problems (5.6) and (5.7). Symmetry-breaking is a technique developed for combinatorial optimization and satisfiability problems in constraint programming, consisting in adding constraints to a problem to remove redundant solutions [109–113], i.e. solutions implying an identical objective value and feasibility. For Problems (5.6) and (5.7) in particular, permuting the prices and user assignments of two different clusters results in the same objective function and feasibility. Having such a symmetric structure of the problem implies that a branch-and-bound algorithm will explore multiple nodes that are effectively equivalent. Imposing an

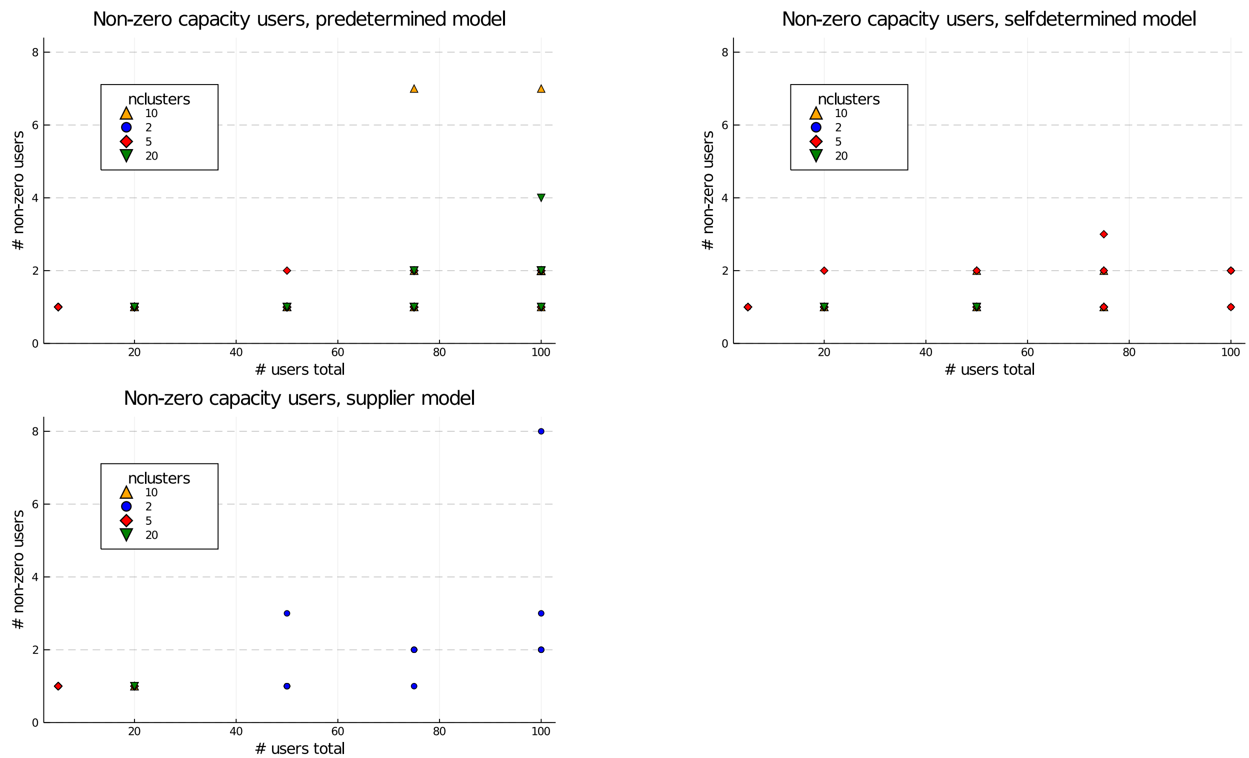

Figure 5.6 Number of users with non-zero capacity in optimal solutions

ordering on the clusters breaks the symmetry by imposing exactly one of the symmetric solutions. In a first symmetry-breaking formulation, we add a set of constraints to order clusters by increasing booking fee K:

$$K_m \leq K_{m+1} \qquad \forall m \in \{1 \mathinner{..} N_G - 1\}. \tag{5.8}$$

Constraint (5.8) is added to the two models to compare the runtime for the different number of users and clusters. Optimality of the lower level uses lazy constraints, which resulted in the fastest resolution on average.

In a second attempt, we impose symmetry-breaking constraints on binary variables. More specifically, for the supplier model, we add the constraint:

$$\sum_{u \in U} p_{u,m} \leq \sum_{u \in U} p_{u,(m+1)} \qquad \forall m \in \{1 ... N_G - 1\} \tag{5.9}$$

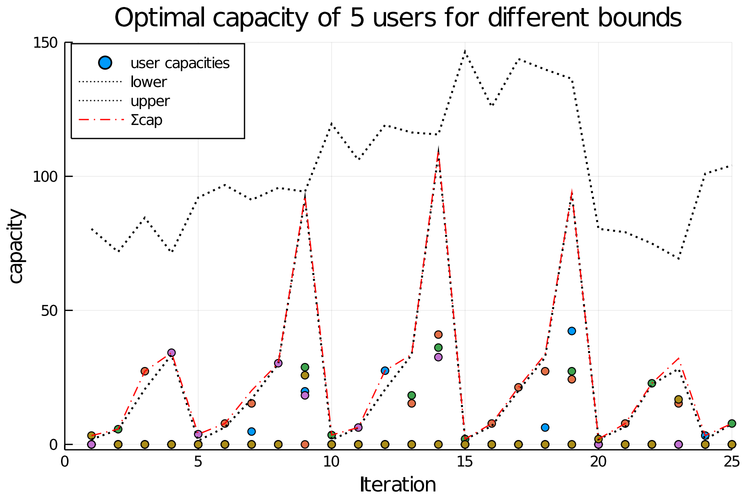

Figure 5.7 Evolution of solutions with various total lower and upper bounds

which is added to Problem (5.6). For the self-determined model, we impose:

$$\sum_{u \in U, k \in S_u} z_{uk}^1 \geq \sum_{u \in U, k \in S_u} z_{uk}^m \qquad \forall m \in \mathcal{G} \tag{5.10}$$

which is added to Problem (5.7).

The average runtimes are computed and reported Table 5.4 on the instances where both models successfully terminated across a range of numbers of users and numbers of clusters. We also report the number of successfully optimized instances for each setting. In Fig. 5.8, we also report the runtime of the formulation without symmetry breaking (*simple*) and with the various symmetry-breaking formulations (*f9, f10, f11* for the corresponding constraints), in the form of performance profiles.

Table 5.4 Runtime and solved instances for the symmetry-breaking constraints

Problem	Method	Runtime (s)	# solved	Constraint
Supplier	Simple	387.3	27	(5.8)
Supplier	Symmetry	353.8	26	(5.8)
Self-determined	Simple	14.1	55	(5.8)
Self-determined	Symmetry	354.5	26	(5.8)
Supplier	Simple	420	27	(5.10)
Supplier	Symmetry	307.7	27	(5.10)
Self-determined	Simple	308.6	55	(5.9)
Self-determined	Symmetry	271	53	(5.9)

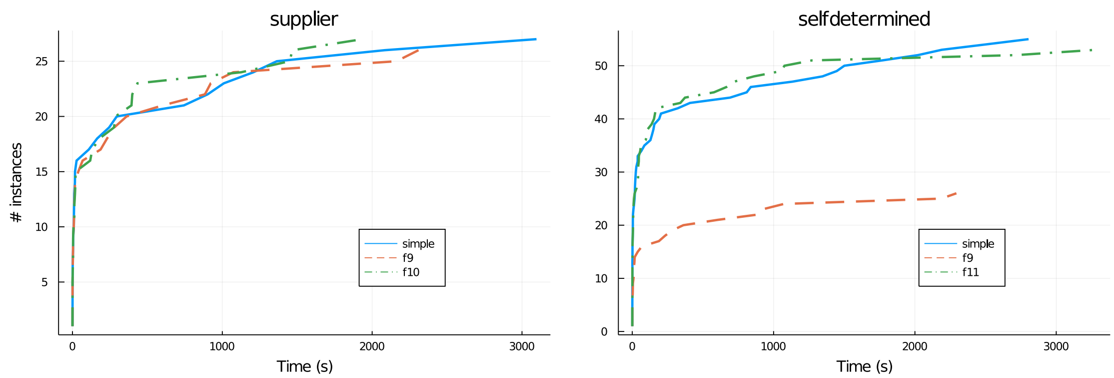

Figure 5.8 Performance profiles comparisons for symmetry-breaking constraints

For Constraint (5.8), the results highlight that the symmetry-breaking constraints increase the computation time and reduce the number of solved instances for the instances with self-determined capacity. Using an ANOVA model from the *GLM.jl* package [114], we can determine that the increase in runtime is significant with a 95% confidence interval when using the symmetry-breaking constraint on the self-determined model. The difference in runtime for the supplier model is inconclusive under the same confidence interval. The average runtime for the model without the symmetry-breaking constraints appears as very low for the self-determined model (14 seconds) because the model with symmetry-breaking fails to solve most instances of significant size.

The addition of Constraints (5.10) and (5.9) improve their respective formulation and reduce the average runtime by 112.3 and 37.6 seconds respectively. We must nonetheless note that for the self-determined model, two instances were solved by the simple formulation but not by the formulation with the symmetry-breaking constraints.

5.7 Conclusion

In this chapter, we defined the problem of optimal price-setting for TLOU tariffs in the context of a supplier interacting with multiple users under a mixed continuous-discrete probability distribution of the consumption. The problem is formulated as a MINLP including bilevel constraints and non-linear non-closed constraints from the integration of the probability density function. Under regularity assumptions on the distribution, the discrete structure can be exploited to linearize the problem, obtaining a MILP. Several variants were introduced, with the user clusters as a fixed parameter, a decision of the supplier, or a decision of each user. These three clustering variants allow the supplier to integrate TLOU in power markets with different economic constraints, such as uniform prices in geographic areas, or by user category. Using a regularity assumption on the continuous part of the probability distributions, one can reformulate the problems as single-level MILPs, which can be solved to global optimality using standard solvers. Computational experiments highlight the practicality of the method and superior efficiency of lazy constraints based on callbacks to express lower-level optimality constraints.

Future research will consider the price-setting problem under uncertainty on the probability distribution of the consumption itself and more complex user clustering structures.

CHAPTER 6 ARTICLE 2: NEAR-OPTIMAL ROBUST BILEVEL OPTIMIZATION

This chapter was submitted to Mathematical Programming in October 2020.

Authors: Mathieu Besançon, Miguel F. Anjos, Luce Brotcorne

Abstract

Bilevel optimization problems embed the optimality conditions of a subproblem into the constraints of another optimization problem. We introduce the concept of near-optimality robustness for bilevel problems, protecting the upper-level solution feasibility from limited deviations at the lower level. General properties and necessary conditions for the existence of solutions are derived for near-optimal robust versions of generic bilevel problems. A duality-based solution method is defined when the lower level is convex, leveraging the methodology from the robust and bilevel literature. Numerical results assess the efficiency of the proposed algorithm and the impact of valid inequalities on the solution time.

6.1 Introduction

Bilevel optimization problems embed the optimality conditions of a subproblem into the constraints of another one. They can model various decision-making problems such as Stackelberg or leader-follower games, market equilibria, or pricing and revenue management. A review of methods and applications of bilevel problems is presented in [98]. In the classical bilevel setting, when optimizing its objective function, the upper level anticipates an optimal reaction of the lower level to its decisions. However, in many practical cases, the lower level can make near-optimal decisions [48]. An important issue in this setting is the definition of the robustness of the upper-level decisions with respect to near-optimal lower-level solutions.

In some engineering applications [37, 40, 41], the decision-maker optimizes an outcome over a dynamical system (modelled as the lower level). For stable systems, the rate of change of the state variables decreases as the system converges towards the minimum of a potential function. If the system is stopped before reaching the minimum, the designer of the system would require that the upper-level constraints be feasible for near-optimal lower-level solutions.

The concept of bounded rationality initially proposed in [91], sometimes referred as ε-rationality [92], defines an economic and behavioural interpretation of a decision-making process where an agent aims to take *any* solution associated with a "satisfactory" objective value instead of the optimal one.

Protecting the upper level from a violation of its constraints by deviations of the lower level is a form of robust optimization. Indeed, it corresponds to a protection of some constraints against uncertain parameters of the problem. Therefore, we use the terms "near-optimality robustness" and "near-optimal robust bilevel problem" or *NORBiP* in the rest of the paper.

The introduction of uncertainty and robustness in games has been approached from different points of view in the literature. In [115], the authors prove the existence of robust counterparts of Nash equilibria under standard assumptions for simultaneous games without the knowledge of probability distributions associated with the uncertainty. In [116], the robust version of a network congestion problem is developed. Users are assumed to make decisions under bounded rationality, leading to a robust Wardrop equilibrium. Robust versions of bilevel problems modelling specific Stackelberg games have been studied in [117, 118], using robust formulations to protect the leader against non-rationality or partial rationality of the follower. A stochastic version of the pessimistic bilevel problem is studied in [88], where the realization of the random variable occurs after the upper level and before the lower level. The authors then derive lower and upper bounds on the pessimistic and optimistic versions of the stochastic bilevel problem as MILPs, leveraging an exact linearization by assuming the upper-level variables are all binary. The models developed in [80] and [79] explore different forms of bounded or partial rationality of the lower level, where the lower level either makes a decision using a heuristic or approximation algorithm algorithm, or may deviate from its optimal value in a way that penalizes the objective of the upper level.

Solving bilevel problems under limited deviations of the lower-level response was introduced in [48] under the term "ε-approximation" of the pessimistic bilevel problem. The authors focus on the independent case, *i.e.* cases where the lower-level feasible set is independent of the upper-level decisions. Problems in such settings are shown to be simpler to handle than the dependent case and can be solved in polynomial time when the lower-level problem is linear under the optimistic and pessimistic assumptions. A custom algorithm is designed for the independent case, solving a sequence of non-convex non-linear problems relying on global optimization solvers.

We consider bilevel problems involving upper- and lower-level variables in the constraints and objective functions at both levels, thus more general than the independent "ε-approximation" from [48]. Unlike the independent case, the dependent bilevel problem is $\mathcal{NP}$-hard even when the constraints and objectives are linear. By defining the uncertainty in terms of a deviation from optimality of the lower level, our formulation offers a novel interpretation of robustness for bilevel problems and Stackelberg games. In the case of a linear lower level, we derive an exact MILP reformulation while not requiring the assumption of pure binary upper-level variables.

The main contributions of the paper are:

1. The definition and formulation of the dependent near-optimal robust bilevel problem, resulting in a generalized semi-infinite problem and its interpretation as a special case of robust optimization applied to bilevel problems.

2. The study of duality-based reformulations of *NORBiP* where the lower-level problem is convex conic or linear in Section 6.3, resulting in a finite-dimensional single-level optimization problem.

3. An extended formulation for the linear-linear *NORBiP* in Section 6.4, linearizing the bilinear constraints of the single-level model using disjunctive constraints.

4. A solution algorithm for the linear-linear *NORBiP* in Section 6.5 using the extended formulation and its implementation with several variants.

The paper is organized as follows. In Section 6.2, we define the concepts of near-optimal set and near-optimal robust bilevel problem. We study the near-optimal bilevel problems with convex and linear lower-level problems in Section 6.3 and Section 6.4 respectively. In both cases, the near-optimal robust bilevel problem can be reformulated as a single level. For a linear lower level, an extended formulation can be derived from the single-level problem. A solution algorithm is provided and computational experiments are conducted for the linear case in Section 6.5, comparing the extended formulation to the compact one and studying the impact of valid inequalities. Finally, in Section 6.6 we draw some conclusions and highlight research perspectives on near-optimality robustness.

6.2 Near-optimal set and near-optimal robust bilevel problem

In this section, we first define the near-optimal set of the lower level and near-optimality robustness for bilevel problems. Next, we illustrate the concepts on an example and highlight several properties of general near-optimal robust bilevel problems before focusing on the convex and linear cases in the following sections.

The generic bilevel problem is classically defined as:

$$\min_{x} \; F(x, v) \tag{6.1a}$$
$$\text{s.t. } G_k(x, v) \leq 0 \qquad \forall k \in [\![m_u]\!] \tag{6.1b}$$
$$x \in \mathcal{X} \tag{6.1c}$$
$$v \in \arg\min_{y \in \mathcal{Y}} \{f(x, y) \text{ s.t. } g_i(x, y) \leq 0 \, \forall i \in [\![m_l]\!]\}. \tag{6.1d}$$

The upper- and lower-level objective functions are noted $F, f : \mathcal{X} \times \mathcal{Y} \mapsto \mathbb{R}$ respectively. Constraint (6.1b) and $g_i(x, y) \leq 0 \, \forall i \in [\![m_l]\!]$ are the upper- and lower-level constraints respectively. In this section, we assume that $\mathcal{Y} = \mathbb{R}^{n_l}$ in order that the lower-level feasible set can be only determined by the g_i functions. The optimal value function $\phi(x)$ is defined as follows:

$$\begin{aligned} &\phi : \mathbb{R}^{n_u} \to \{-\infty\} \cup \mathbb{R} \cup \{+\infty\} \\ &\phi(x) = \min_{y} \{f(x, y) \text{ s.t. } g(x, y) \leq 0\}. \end{aligned} \tag{6.2}$$

To keep the notation succinct, the indices of the lower-level constraints g_i are omitted when not needed as in Constraint (6.2). Throughout the paper, it is assumed that the lower-level problem is feasible and bounded for any given upper-level decision.

When, for a feasible upper-level decision, the solution to the lower-level problem is not unique, the bilevel problem is not well defined and further assumptions are required [98]. In the optimistic case, we assume that the lower level selects the optimal solution favouring the upper level and the optimal solution disfavouring them the most in the pessimistic case. We refer the reader to [47, Chapter 1] for further details on these two approaches. The near-optimal set of the lower level $\mathcal{Z}(x; \delta)$ is defined for a given upper-level decision x and tolerance δ as:

$$\mathcal{Z}(x; \delta) = \{y \mid g(x, y) \leq 0, \; f(x, y) \leq \phi(x) + \delta\}.$$

A Near-Optimal Robust Bilevel Problem, *NORBiP*, of parameter δ is defined as a bilevel problem where the upper-level constraints are satisfied for any lower-level solution z in the near-optimal set $\mathcal{Z}(x;\delta)$.

$$\min_{x,v}\ F(x,v) \tag{6.3a}$$
$$\text{s.t. } G_k(x,v) \leq 0 \qquad \forall k \in [\![m_u]\!] \tag{6.3b}$$
$$f(x,v) \leq \phi(x) \tag{6.3c}$$
$$g(x,v) \leq 0 \tag{6.3d}$$
$$G_k(x,z) \leq 0\ \forall z \in \mathcal{Z}(x;\delta) \qquad \forall k \in [\![m_u]\!] \tag{6.3e}$$
$$x \in \mathcal{X}. \tag{6.3f}$$

Each k constraint in (6.3b) is satisfied if the corresponding constraint set in (6.3e) holds and is therefore redundant, since $v \in \mathcal{Z}(x;\delta)$. However, we mention Constraint (6.3b) in the formulation to highlight the structure of the initial bilevel problem in the near-optimal robust formulation.

The special case $\mathcal{Z}(x;0)$ is the set of optimal solutions to the original lower-level problem, NORBiP with $\delta = 0$ is therefore equivalent to the pessimistic bilevel problem as formulated in [48]:

$$f(x,y) \leq \phi(x)\ \forall y \in \mathcal{Z}(x;0).$$

For $\delta < 0$, $\mathcal{Z}(x;\delta)$ is the empty set, in which case Problem (6.3) is equivalent to the original optimistic bilevel problem while the set $\mathcal{Z}(x;\infty)$ corresponds to the complete lower-level feasible set, assuming the lower-level optimal solution is not unbounded for the given upper-level decision x.

Unlike the constraint-based pessimistic bilevel problem presented in [48], the upper-level objective $F(x,v)$ depends on both the upper- and lower-level variables, but is only evaluated with the optimistic lower-level variable v and not with a worst-case near-optimal solution. This implies the upper level chooses the best optimistic decision which protects its feasibility from near-optimal deviations. One implication for the modeller is that a near-optimal robust problem can be constructed directly from a bilevel instance where the objective function often depends on the variables of the two levels. Alternatively, the near-optimal robust formulation can protect both the upper-level objective value and constraints from near-optimal deviations of the lower level using an epigraph formulation introducing an additional variable:

$$\min_{x,v,\tau} \tau \tag{6.4a}$$
$$\text{s.t. } G_k(x,v) \leq 0 \qquad \forall k \in [\![m_u]\!] \tag{6.4b}$$
$$f(x,v) \leq \phi(x) \tag{6.4c}$$
$$g(x,v) \leq 0 \tag{6.4d}$$
$$F(x,z) \leq \tau \qquad \forall z \in \mathcal{Z}(x;\delta) \tag{6.4e}$$
$$G_k(x,z) \leq 0 \qquad \forall z \in \mathcal{Z}(x;\delta)\ \forall k \in [\![m_u]\!] \tag{6.4f}$$
$$x \in \mathcal{X}. \tag{6.4g}$$

The two models define different levels of conservativeness and risk. Indeed:

$$\text{opt}(6.1\text{a-}6.1\text{d}) \leq \text{opt}(6.3\text{a-}6.3\text{f}) \leq \text{opt}(6.4\text{a-}6.4\text{g}),$$

where $\text{opt}(P)$ denotes the optimal value of problem P. Both near-optimal robust formulations can be of interest to model decision-making applications. It can also be noted that Problem 6.3 includes the special case of opposite objectives between the two levels, i.e. problems for which $F(x,v) = -f(x,v)$. The two models offer different levels of conservativeness and risk and can both be of interest when modelling decision-making applications.

Constraint (6.3e) is a generalized semi-infinite constraint, based on the terminology from [78]. The dependence of the set of constraints $\mathcal{Z}(x;\delta)$ on the decision variables leads to the characterization of Problem (6.3) as a robust problem with decision-dependent uncertainty [77]. Each constraint in the set (6.3e) can be replaced by the corresponding worst-case second-level decision z_k obtained as the solution of the *adversarial problem*, parameterized by (x,v,δ):

$$z_k \in \arg\max_y G_k(x,y) \tag{6.5a}$$
$$\text{s.t. } f(x,y) \leq f(x,v) + \delta \tag{6.5b}$$
$$g(x,y) \leq 0. \tag{6.5c}$$

Finally, the near-optimal robust bilevel optimization problem can be expressed as:

$$\min_{x,v}\ F(x,v) \tag{6.6a}$$
$$\text{s.t. } f(x,v) \leq \phi(x) \tag{6.6b}$$
$$g(x,v) \leq 0 \tag{6.6c}$$
$$0 \geq \max_{y}\{G_k(x,y) \text{ s.t. } y \in \mathcal{Z}(x;\delta)\} \qquad \forall k \in [\![m_u]\!] \tag{6.6d}$$
$$x \in \mathcal{X}. \tag{6.6e}$$

In the robust optimization literature, models can present uncertainty on the constraints and/or on the objective function [119]. In bilevel optimization, the first case corresponds to *NORBiP*, where the impact of near-optimal lower-level solutions on the upper-level constraints is studied. The second case corresponds to the impact of near-optimal lower-level decisions on the upper-level objective value.

We next prove that the model including uncertainty on the objective, named Objective-Robust Near-Optimal Bilevel Problem (*ORNOBiP*), is a special case of *NORBiP*.

ORNOBiP is defined as:

$$\min_{x \in X}\ \sup_{z \in \mathcal{Z}(x;\delta)}\ F(x,z) \tag{6.7a}$$
$$\text{s.t. } X = \{x \in \mathcal{X}, G_k(x) \leq 0 \forall k \in [\![m_u]\!]\} \tag{6.7b}$$
$$\text{and where: } \mathcal{Z}(x;\delta) = \{y \text{ s.t. } g(x,y) \leq 0, f(x,y) \leq \phi(x) + \delta\}.$$

In contrast to most objective-robust problem formulations, the uncertainty set $\mathcal{Z}$ depends on the upper-level solution x, qualifying Problem (6.7) as a problem with decision-dependent uncertainty.

Proposition 6.2.1. *ORNOBiP is a special case of NORBiP.*

Proof. The reduction of the objective-uncertain robust problem to a constraint-uncertain robust formulation is detailed in [76]. In particular, Problem (6.7) is equivalent to:

$$\begin{aligned}
\min_{x,\tau}\ & \tau \\
\text{s.t. } & x \in X \\
& \tau \geq F(x,z) && \forall z \in \mathcal{Z}(x,\delta),
\end{aligned}$$

this formulation is a special case of *NORBiP*. □

The pessimistic bilevel optimization problem defined in [120] is both a special case and a relaxation of *ORNOBiP*. For $\delta = 0$, the inner problem of *ORNOBiP* is equivalent to finding the worst lower-level decision with respect to the upper-level objective amongst the lower-level-optimal solutions. For any $\delta > 0$, the inner problem can select the worst solutions with respect to the upper-level objective that are not optimal for the lower level. The pessimistic bilevel problem is therefore a relaxation of *ORNOBiP*.

We illustrate the concept of near-optimal set and near-optimal robust solution with the following linear bilevel problem, represented in Fig. 6.1.

$$\begin{aligned}
\min_{x,v}\ & x \\
\text{s.t. } & x \geq 0 \\
& v \geq 1 - \frac{x}{10} \\
& v \in \arg\max_{y} \{ y \text{ s.t. } y \leq 1 + \frac{x}{10} \}.
\end{aligned} \tag{6.8}$$

The high-point relaxation of Problem (6.8), obtained by relaxing the optimality constraint of the lower-level, while maintaining feasibility, is:

$$\begin{aligned}
\min_{x,v}\ & x \\
\text{s.t. } & x \geq 0 \\
& v \geq 1 - \frac{x}{10} \\
& v \leq 1 + \frac{x}{10}.
\end{aligned}$$

The shaded area in Fig. 6.1 represents the interior of the polytope, which is feasible for the high-point relaxation. The induced set, resulting from the optimal lower-level reaction, is given by:

$$\{(x, y) \in (\mathbb{R}_+, \mathbb{R}) \,\text{s.t.}\, y = 1 + \frac{x}{10}\}.$$

The unique optimal point is $(\hat{x}, \hat{y}) = (0, 1)$.

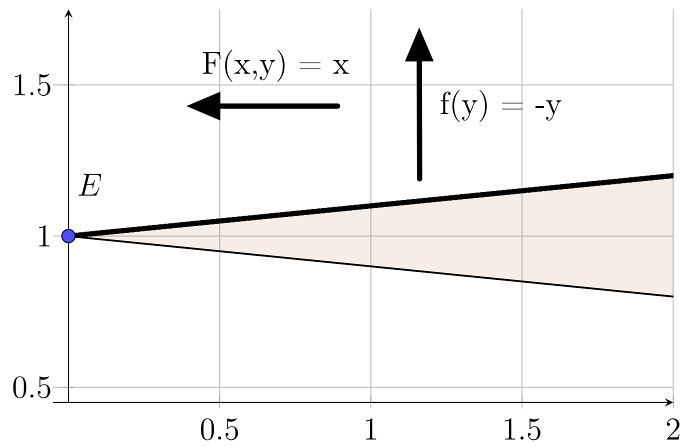

Figure 6.1 Linear bilevel problem

Let us now consider a near-optimal tolerance of the follower with $\delta = 0.1$. If the upper-level decision is $\hat{x}$, then the lower level can take any value between $1 - \delta = 0.9$ and 1. All these values except 1 lead to an unsatisfied upper-level constraint problem. The problem can be reformulated as:

$$\begin{aligned}
\min_{x,v}\ & x \\
\text{s.t.}\ & x \geq 0 \\
& v \geq 1 - \frac{x}{10} \\
& v \in \arg\max_{y} \{y \text{ s.t. } \ y \leq 1 + \frac{x}{10}\} \\
& z \geq 1 - \frac{x}{10} \ \forall z \text{ s.t. } \{z \leq 1 + \frac{x}{10}, z \geq v - \delta\}.
\end{aligned}$$

Fig. 6.2 illustrates the near-optimal equivalent of the problem with an additional constraint ensuring the satisfaction of the upper-level constraint for all near-optimal responses of the lower level.

This additional constraint is represented by the dashed line. The optimal upper-level decision is $x = 0.5$, for which the optimal lower-level reaction is $y = 1 + 0.1 \cdot 0.5 = 1.05$. The boundary of the near-optimal set is $y = 1 - 0.1 \cdot 0.5 = 0.95$.

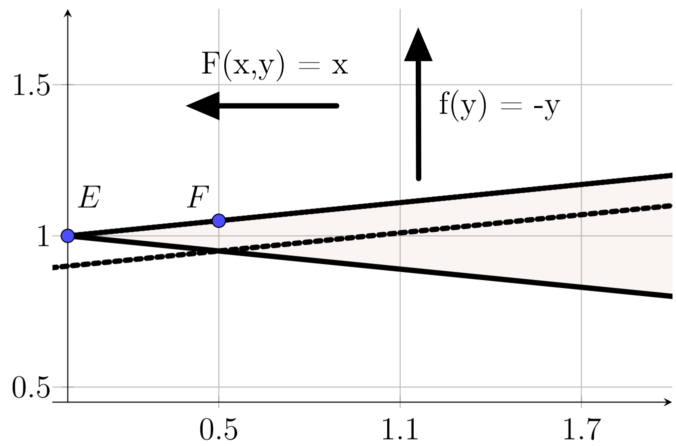

Figure 6.2 Linear bilevel problem with a near-optimality robustness constraint

In the rest of this section, we establish properties of the near-optimal set and near-optimal robust bilevel problems. If the lower-level optimization problem is convex, then the near-optimal set $\mathcal{Z}(x;\theta)$ is convex as the intersection of two convex sets:

- $\{y \mid g(x,y) \leq 0\}$
- $\{y \mid f(x,y) \leq \phi(x) + \delta\}$.

In robust optimization, the characteristics of the uncertainty set sharply impact the difficulty of solving the problem. The near-optimal set of the lower-level is not always bounded; this can lead to infeasible or ill-defined near-optimal robust counterparts of bilevel problems. In the next proposition, we define conditions under which the uncertainty set $\mathcal{Z}(x;\delta)$ is bounded.

Proposition 6.2.2. *For a given pair (x,δ), any of the following properties is sufficient for $\mathcal{Z}(x;\delta)$ to be a bounded set:*

1. *The lower-level feasible domain is bounded.*
2. *$f(x,\cdot)$ is radially unbounded, i.e. $\|y\| \to \infty \Rightarrow f(x,y) \to \infty$.*
3. *$f(x,\cdot)$ is radially bounded such that:*

$$\lim_{r\to+\infty} f(x, rs) > f(x,v) + \delta \;\forall s \in \mathcal{S},$$

with $\mathcal{S}$ the unit sphere in the space of lower-level variables.

Proof. The first case is trivially satisfied since $\mathcal{Z}(x;\delta)$ is the intersection of sets including the lower-level feasible set. If $f(x,\cdot)$ is radially unbounded, for any finite $\delta > 0$, there is a maximum radius around v beyond which any value of the objective function is greater than $f(x,v)+\delta$. The third case follows the same line of reasoning as the second, with a lower bound in any direction $\|y\| \to \infty$, such that this lower bound is above $f(x,v)+\delta$. □

The radius of robust feasibility is defined as the maximum "size" of the uncertain set [121,122], such that the robust problem remains feasible. In the case of near-optimality robustness, the radius can be interpreted as the maximum deviation of the lower-level objective from its optimal value, such that the near-optimal robust bilevel problem remains feasible.

Definition 6.2.1. *For a given optimization problem BiP, let $\mathcal{NO}(BiP;\delta)$ be the optimum value of the near-optimal robust problem constructed from BiP with a tolerance δ. The radius of near-optimal feasibility $\hat{\delta}$ is defined by:*

$$\hat{\delta} = \arg\max_{\delta} \{\delta \text{ s.t. } \mathcal{NO}(BiP;\delta) < \infty\}. \tag{6.9}$$

It is interesting to note that the radius as defined in Definition 6.2.1 can be interpreted as a maximum robustness budget in terms of the objective value of the lower level. It represents the maximum level of tolerance of the lower level on its objective, such that the upper level remains feasible.

Proposition 6.2.3. *The standard optimistic bilevel problem BiP is a relaxation of the equivalent near-optimal robust bilevel problem for any $\delta > 0$.*

Proof. By introducing additional variables $z_{jk}, j \in [\![n_l]\!], k \in [\![m_u]\!]$ in the optimistic bilevel problem, we obtain:

$$\begin{aligned}
\min_{x,v,z}\ & F(x,v) && \qquad (6.10)\\
\text{s.t. } & G_k(x,v) \leq 0 && \forall k \in [\![m_u]\!]\\
& f(x,v) \leq \phi(x) &&\\
& g(x,v) \leq 0 &&\\
& x \in \mathcal{X}, v \in \mathbb{R}^{n_l}, z \in \mathbb{R}^{n_l \times m_u}. &&
\end{aligned}$$

Problem (6.10) is equivalent to the optimistic bilevel problem with additional variables z that are not used in the objective nor constraints. Furthermore, Problem (6.10) is a relaxation of Problem (6.6), which has similar variables but additional constraints (6.6d). At each point where the bilevel problem is feasible, either the objective value of the two problems are the same or *NORBiP* is infeasible. □

Proposition 6.2.4. *If the bilevel problem is feasible, then the adversarial problem* (6.5) *is feasible.*

Proof. If the bilevel problem is feasible, then the solution $z = v$ is feasible for the primal adversarial problem. $\square$

Proposition 6.2.5. *If* $(\hat{x}, \hat{y})$ *is a bilevel-feasible point, and* $G_k(\hat{x}, \cdot)$ *is* K_k*-Lipschitz continuous for a given* $k \in [\![m_u]\!]$ *such that:*

$$G_k(\hat{x}, \hat{y}) < 0,$$

then the constraint $G_k(\hat{x}, y) \leq 0$ *is satisfied for all* $y \in \mathcal{F}_L^{(k)}$ *such that:*

$$\mathcal{F}_L^{(k)}(\hat{x}, \hat{y}) = \{y \in \mathbb{R}^{n_l} \mid \|y - \hat{y}\| \leq \frac{|G_k(\hat{x}, \hat{y})|}{K_k}\}.$$

Proof. As $G_k(\hat{x}, \hat{y}) < 0$, and $G_k(\hat{x}, \cdot)$ is continuous, there exists a ball $\mathcal{B}_r(\hat{y})$ in $\mathbb{R}^{n_l}$ centered on $(\hat{y})$ of radius $r > 0$, such that

$$G(\hat{x}, y) \leq 0 \;\forall y \in \mathcal{B}_r(\hat{y}).$$

Let us define:

$$\begin{aligned} r_0 = \arg\max_{r} \;\; & r \\ \text{s.t.} \;\; & G(\hat{x}, y) \leq 0 \qquad\qquad \forall y \in \mathcal{B}_r(\hat{y}). \end{aligned} \tag{6.11}$$

By continuity, Problem (6.11) always admits a feasible solution. If the feasible set is bounded, there exists a point y_0 on the boundary of the ball, such that $G_k(\hat{x}, y_0) = 0$. It follows from Lipschitz continuity that:

$$\begin{aligned} & |G_k(\hat{x}, \hat{y}) - G_k(\hat{x}, y_0)| \leq K_k \|y_0 - \hat{y}\| \\ & \frac{|G_k(\hat{x}, \hat{y})|}{K_k} \leq \|y_0 - \hat{y}\|. \end{aligned}$$

$G_k(\hat{x}, y) \leq G_k(\hat{x}, y_0) \;\forall y \in \mathcal{B}_{r_0}(\hat{y})$, therefore all lower-level solutions in the set

$$\mathcal{F}_L^{(k)}(\hat{x}, \hat{y}) = \{y \in \mathbb{R}^{n_l} \text{ s.t. } \|y - \hat{y}\| \leq \frac{|G_k(\hat{x}, \hat{y})|}{K_k}\}$$

satisfy the k-th constraint. $\square$

Corollary 6.2.1. *Let $(\hat{x}, \hat{y})$ be a bilevel-feasible solution of a near-optimal robust bilevel problem of tolerance δ, and*

$$\mathcal{F}_L(\hat{x}, \hat{y}) = \bigcap_{k=1}^{m_u} \mathcal{F}_L^{(k)}(\hat{x}, \hat{y}),$$

then $\mathcal{Z}(x;\delta) \subseteq \mathcal{F}_L(\hat{x}, \hat{y})$ is a sufficient condition for the solution $(\hat{x}, \hat{y})$ to be near-optimal robust.

Proof. Any lower-level solution $y \in \mathcal{F}_L(\hat{x}, \hat{y})$ satisfies all m_u upper-level constraints, thus $\mathcal{Z}(x;\delta) \subseteq \mathcal{F}_L(\hat{x}, \hat{y})$ is a sufficient condition for the near-optimality robustness of $(\hat{x}, \hat{y})$. □

Corollary 6.2.2. *Let $(\hat{x}, \hat{y})$ be a bilevel-feasible solution of a near-optimal robust bilevel problem of tolerance δ, R be the radius of the lower-level feasible set and $G_k(\hat{x}, \cdot)$ be K_k-Lipschitz for a given k, then the k-th constraint is robust against near-optimal deviations if:*

$$|G_k(\hat{x}, \hat{y})| \leq K_k R.$$

Proof. The inequality can be deduced from the fact that $\|y - \hat{y}\| \leq R$. □

Corollary 6.2.2 can be used when the lower level feasible set is bounded to verify near-optimality robustness of incumbent solutions.

6.3 Near-optimal robust bilevel problems with a convex lower level

In this section, we study near-optimal robust bilevel problems where the lower-level problem (6.1d) is a parametric convex optimization problem with both a differentiable objective function and differentiable constraints. If Slater's constraint qualifications hold, the KKT conditions are necessary and sufficient for the optimality of the lower-level problem and strong duality holds for the adversarial subproblems. These two properties are leveraged to reformulate *NORBiP* as a single-level closed-form problem.

Given a bilevel solution (x, v), the adversarial problem associated with constraint k can be formulated as:

$$\max_{y} \; G_k(x, y) \tag{6.12a}$$

$$\text{s.t.} \; g(x, y) \leq 0 \tag{6.12b}$$

$$f(x, y) \leq f(x, v) + \delta. \tag{6.12c}$$

Even if the upper-level constraints are convex with respect to y, Problem (6.12) is in general non-convex since the function to maximize is convex over a convex set. First-order optimality conditions may induce several non-optimal critical points and the definition of a solution method needs to rely on global optimization techniques [123, 124].

By assuming that the constraints of the upper-level problem $G_k(x, y)$ can be decomposed and that the projection onto the lower variable space is affine, the upper-level constraint can be re-written as:

$$G_k(x, y) \leq 0 \Leftrightarrow G_k(x) + H_k^T y \leq q_k. \tag{6.13}$$

The k-th adversarial problem is then expressed as:

$$\begin{aligned}
&\max_{y} \; \langle H_k, y \rangle && && (6.14a)\\
&\text{s.t.} \;\; g_i(x, y) \leq 0 \quad \forall i \in [\![m_l]\!] && (\alpha_i) && (6.14b)\\
&\qquad f(x, y) \leq f(x, v) + \delta && (\beta) && (6.14c)
\end{aligned}$$

and is convex for a fixed pair (x, v). Satisfying the upper-level constraint in the worst-case requires that the objective value of Problem (6.14) is lower than $q_k - G_k(x)$. We denote by $\mathcal{A}_k$ and $\mathcal{D}_k$ the objective values of the adversarial problem (6.14) and its dual respectively. $\mathcal{D}_k$ takes values in the extended real set to account for infeasible and unbounded cases. Proposition 6.2.4 holds for Problem (6.14). The feasibility of the upper-level constraint with the dual adversarial objective value as formulated in Constraint (6.15) is, by weak duality of convex problems, a sufficient condition for the feasibility of a near-optimal solution. If Slater's constraint qualifications hold, it is also a necessary condition [125] by strong duality:

$$\mathcal{A}_k \leq \mathcal{D}_k \leq q_k - G_k(x). \tag{6.15}$$

The generic form for the single-level reformulation of the near-optimal robust problem can then be expressed as:

$$
\begin{aligned}
\min_{x,v,\alpha,\beta} \quad & F(x,v) && && (6.16a)\\
\text{s.t.} \quad & G(x) + Hv \leq q && && (6.16b)\\
& f(x,v) \leq \phi(x) && && (6.16c)\\
& g(x,v) \leq 0 && && (6.16d)\\
& \mathcal{D}_k \leq q_k - G_k(x) && \forall k \in [\![m_u]\!] && (6.16e)\\
& x \in \mathcal{X}. && && (6.16f)
\end{aligned}
$$

In order to write Problem (6.16) in a closed form, the lower-level problem (6.16c-6.16d) is reduced to its KKT conditions:

$$
\begin{aligned}
& \nabla_v f(x,v) - \sum_{i=1}^{m_l} \lambda_i \nabla_v g_i(x,v) = 0 && && (6.17a)\\
& g_i(x,v) \leq 0 && \forall i \in [\![m_l]\!] && (6.17b)\\
& \lambda_i \geq 0 && \forall i \in [\![m_l]\!] && (6.17c)\\
& \lambda_i g_i(x,v) = 0 && \forall i \in [\![m_l]\!]. && (6.17d)
\end{aligned}
$$

Constraint (6.17d) derived from the KKT conditions cannot be tackled directly by non-linear solvers [126]. Specific reformulations, such as relaxations of the equality Constraints (6.17d) into inequalities or branching on combinations of variables (as developed in [62,64]) are often used in practice.

In the rest of this section, we focus on bilevel problems such that the lower level is a conic convex optimization problem. Unlike the convex version developed above, the dual of a conic optimization problem can be written in closed form.

$$
\begin{aligned}
\min_{y} \quad & \langle d, y \rangle \\
\text{s.t.} \quad & Ax + By = b \\
& y \in \mathcal{K},
\end{aligned}
\tag{6.18}
$$

where $\langle \cdot, \cdot \rangle$ is the inner product associated with the space of the lower-level variables and $\mathcal{K}$ is a proper cone [125, Chapter 2]. This class of problems encompasses a broad class of convex optimization problems of practical interest [127, Chapter 4], while the dual problem can be written in a closed-form if the dual cone is known, leading to a closed-form single-level reformulation. The $k-$th adversarial problem is given by:

$$\begin{aligned}
\max_{y,r} \quad & \langle H_k, y\rangle && (6.19a)\\
\text{s.t.} \quad & By = b - Ax && (6.19b)\\
& \langle d, y\rangle + r = \langle d, v\rangle + \delta && (6.19c)\\
& y \in \mathcal{K} && (6.19d)\\
& r \geq 0, && (6.19e)
\end{aligned}$$

with the introduction of a slack variable r. With the following change of variables:

$$\hat{y} = \begin{bmatrix} y \\ r \end{bmatrix} \;\; \hat{B} = \begin{bmatrix} B & 0 \end{bmatrix} \;\; \hat{d} = \begin{bmatrix} d & 1 \end{bmatrix} \;\; \hat{H}_k = \begin{bmatrix} H_k \\ 0 \end{bmatrix},$$

$$\hat{\mathcal{K}} = \{(y, r),\; y \in \mathcal{K},\; r \geq 0\},$$

$\hat{\mathcal{K}}$ is a cone as the Cartesian product of $\mathcal{K}$ and the nonnegative orthant. Problem (6.19) is reformulated as:

$$\begin{aligned}
\max_{\hat{y}} \quad & \langle \hat{H}_k, \hat{y}\rangle \\
\text{s.t.} \quad & (\hat{B}\hat{y})_i = b_i - (Ax)_i \quad \forall i \in [\![m_l]\!] && (\alpha_i)\\
& \langle \hat{d}, \hat{y}\rangle = \langle d, v\rangle + \delta && (\beta)\\
& \hat{y} \in \hat{\mathcal{K}},
\end{aligned}$$

which is a conic optimization problem, for which the dual problem is:

$$\begin{aligned}
\min_{\alpha,\beta,s_k} \quad & \langle (b - Ax), \alpha\rangle + (\langle d, v\rangle + \delta)\beta && (6.20a)\\
\text{s.t.} \quad & \hat{B}^T\alpha + \beta\hat{d} + s = \hat{H}_k && (6.20b)\\
& s \in -\hat{\mathcal{K}}^*, && (6.20c)
\end{aligned}$$

with $\hat{\mathcal{K}}^*$ the dual cone of $\hat{\mathcal{K}}$. In the worst case (maximum number of non-zero coefficients), there are $(m_l \cdot n_u + n_l)$ bilinear terms in m_u non-linear non-convex constraints. This number of bilinear terms can be reduced by introducing the following variables (p, o), along with the corresponding constraints:

$$
\begin{aligned}
\min_{\alpha,\beta,s,p,o} \quad & \langle p, \alpha \rangle + (o+\delta)\beta && \text{(6.21a)}\\
\text{s.t.} \quad & p = b - Ax && \text{(6.21b)}\\
& o = \langle d, v \rangle && \text{(6.21c)}\\
& \hat{B}^T \alpha + \beta \hat{d} + s = \hat{H}_k && \text{(6.21d)}\\
& s \in -\hat{\mathcal{K}}^*. && \text{(6.21e)}
\end{aligned}
$$

The number of bilinear terms in the set of constraints is thus reduced from $n_u m_l + n_l$ to $m_l + 1$ terms in (6.21a). Problem (6.20) or equivalently Problem (6.21) have a convex feasible set but a bilinear non-convex objective function. The KKT conditions of the follower problem (6.18) are given for the primal-dual pair (y, λ):

$$
\begin{aligned}
& By = b - Ax && \text{(6.22a)}\\
& y \in \mathcal{K} && \text{(6.22b)}\\
& d - B^T \lambda \in \mathcal{K}^* && \text{(6.22c)}\\
& \langle d - B^T \lambda, y \rangle = 0. && \text{(6.22d)}
\end{aligned}
$$

The single-level problem is:

$$
\begin{aligned}
\min_{x,v,\lambda,\alpha,\beta,s} \quad & F(x, v) && && \text{(6.23a)}\\
\text{s.t.} \quad & G(x) + Hv \leq q && && \text{(6.23b)}\\
& Ax + Bv = b && && \text{(6.23c)}\\
& d - B^T \lambda \in \mathcal{K}^* && && \text{(6.23d)}\\
& \langle d - B^T \lambda, v \rangle = 0 && && \text{(6.23e)}\\
& \langle Ax - b, \alpha_k \rangle + \beta_k \left(\langle v, d \rangle + \delta \right) \leq q_k - (Gx)_k && \forall k \in [\![m_u]\!] && \text{(6.23f)}\\
& \hat{B}^T \alpha_k + \hat{d} \beta_k + s_k = \hat{H}_k && \forall k \in [\![m_u]\!] && \text{(6.23g)}\\
& x \in \mathcal{X}, v \in \mathcal{K} && && \text{(6.23h)}\\
& s_k \in -\hat{\mathcal{K}}^* && \forall k \in [\![m_u]\!]. && \text{(6.23i)}
\end{aligned}
$$

The Mangasarian-Fromovitz constraint qualification is violated at every feasible point of Constraint (6.23e) [128]. In non-linear approaches to complementarity constraints [62, 126], parameterized successive relaxations of the complementarity constraints are used:

$$\langle d - B^T\lambda, v\rangle \leq \varepsilon \tag{6.24a}$$

$$-\langle d - B^T\lambda, v\rangle \leq \varepsilon. \tag{6.24b}$$

Constraints (6.23f) and (6.24) are both bilinear non-convex inequalities, the other ones added by the near-optimal robust model are conic and linear constraints.

In conclusion, near-optimal robustness has only added a finite number of constraints of the same nature (bilinear inequalities) to the reformulation proposed in [126]. Solution methods used for bilevel problems with convex lower-level thus apply to their near-optimal robust counterpart.

6.4 Linear near-optimal robust bilevel problem

In this section, we focus on near-optimal robust linear-linear bilevel problems. More precisely, the structure of the lower-level problem is exploited to derive an extended formulation leading to an efficient solution algorithm. We consider that all vector spaces are subspaces of $\mathbb{R}^n$, with appropriate dimensions. The inner product of two vectors $\langle a, b\rangle$ is equivalently written $a^T b$.

The linear near-optimal robust bilevel problem is formulated as:

$$\min_{x,v} \; c_x^T x + c_y^T v \tag{6.25a}$$

$$\text{s.t.} \; Gx + Hv \leq q \tag{6.25b}$$

$$d^T v \leq \phi(x) \tag{6.25c}$$

$$Ax + Bv \leq b \tag{6.25d}$$

$$Gx + Hz \leq q \qquad \forall z \in \mathcal{Z}(x;\delta) \tag{6.25e}$$

$$v \in \mathbb{R}_+^{n_l} \tag{6.25f}$$

$$x \in \mathcal{X}. \tag{6.25g}$$

For a given pair (x, v), each semi-infinite robust constraint (6.25e) can be reformulated as the objective value of the following adversarial problem:

$$\max_{y} \; H_k^T y \tag{6.26a}$$
$$\text{s.t.} \; (By)_i \leq b_i - (Ax)_i \;\; \forall i \in [\![m_l]\!] \qquad (\alpha_i) \tag{6.26b}$$
$$d^T y \leq d^T v + \delta \qquad (\beta) \tag{6.26c}$$
$$y \in \mathbb{R}_+^{n_l}. \tag{6.26d}$$

Let (α, β) be the dual variables associated with each group of constraints (6.26b-6.26c). The near-optimal robust version of Problem (6.25) is feasible only if the objective value of each k-th adversarial subproblem (6.26) is lower than $q_k - (Gx)_k$. The dual of Problem (6.26) is defined as:

$$\min_{\alpha,\beta} \alpha^T (b - Ax) + \beta \, (d^T v + \delta) \tag{6.27a}$$
$$\text{s.t.} \; B^T \alpha + \beta d \geq H_k \tag{6.27b}$$
$$\alpha \in \mathbb{R}_+^{m_l} \beta \in \mathbb{R}_+. \tag{6.27c}$$

Based on Problem (6.2.4) and weak duality results, the dual problem is either infeasible or feasible and bounded. By strong duality, the objective value of the dual and primal problems are equal. This value must be smaller than $q_k - (Gx)_k$ to satisfy Constraint (6.25e). This is equivalent to the existence of a feasible dual solution (α, β) certifying the feasibility of (x, v) within the near-optimal set $\mathcal{Z}(x; \delta)$. We obtain one pair of certificates (α, β) for each upper-level constraint in $[\![m_u]\!]$, resulting in the following problem:

$$\min_{x,v,\alpha,\beta} \; c_x^T x + c_y^T v \tag{6.28a}$$
$$\text{s.t.} \; Gx + Hv \leq q \tag{6.28b}$$
$$d^T v \leq \phi(x) \tag{6.28c}$$
$$Ax + Bv \leq b \tag{6.28d}$$
$$\alpha_k^T (b - Ax) + \beta_k \, (d^T v + \delta) \leq q_k - (Gx)_k \; \forall k \in [\![m_u]\!] \tag{6.28e}$$
$$B^T \alpha_k + \beta_k d \geq H_k \; \forall k \in [\![m_u]\!] \tag{6.28f}$$
$$\alpha_k \in \mathbb{R}_+^{m_l} \beta_k \in \mathbb{R}_+ \; \forall k \in [\![m_u]\!] \tag{6.28g}$$
$$v \in \mathbb{R}_+^{n_l} \tag{6.28h}$$
$$x \in \mathcal{X}. \tag{6.28i}$$

Lower-level optimality is guaranteed by the corresponding KKT conditions:

$$d_j + \sum_i B_{ij}\lambda_i - \sigma_j = 0 \qquad \forall j \in [\![n_l]\!] \tag{6.29a}$$

$$0 \leq b_i - (Ax)_i - (Bv)_i \perp \lambda_i \geq 0 \qquad \forall i \in [\![m_l]\!] \tag{6.29b}$$

$$0 \leq v_j \perp \sigma_j \geq 0 \qquad \forall j \in [\![n_l]\!] \tag{6.29c}$$

$$\sigma \geq 0, \lambda \geq 0 \tag{6.29d}$$

where $\perp$ defines a complementarity constraint. A common technique to linearize Constraints (6.29b-6.29c) is the "big-M" reformulation, introducing auxiliary binary variables with primal and dual upper bounds. The resulting formulation has a weak continuous relaxation. Furthermore, the correct choice of bounds is itself an NP-hard problem [61], and the incorrect choice of these bounds can lead to cutting valid and potentially optimal solutions [60]. Other modelling and solution approaches, such as special ordered sets of type 1 (SOS1) or indicator constraints avoid the need to specify such bounds in a branch-and-bound procedure.

The aggregated formulation of the linear near-optimal robust bilevel problem is:

$$\min_{x,v,\lambda,\sigma,\alpha,\beta} \quad c_x^T x + c_y^T v \tag{6.30a}$$

$$\text{s.t.} \quad Gx + Hv \leq q \tag{6.30b}$$

$$Ax + Bv \leq b \tag{6.30c}$$

$$d_j + \sum_i \lambda_i B_{ij} - \sigma_j = 0 \qquad \forall j \in [\![n_l]\!] \tag{6.30d}$$

$$0 \leq \lambda_i \perp A_i x + B_i v - b_i \leq 0 \qquad \forall i \in [\![m_l]\!] \tag{6.30e}$$

$$0 \leq \sigma_j \perp v_j \geq 0 \qquad \forall j \in [\![n_l]\!] \tag{6.30f}$$

$$x \in \mathcal{X} \tag{6.30g}$$

$$\alpha_k \cdot (b - Ax) + \beta_k(d^T v + \delta) \leq q_k - (Gx)_k \qquad \forall k \in [\![m_u]\!] \tag{6.30h}$$

$$\sum_{i=1}^{m_l} B_{ij}\alpha_{ki} + \beta_k d_j \geq H_{kj} \qquad \forall k \in [\![m_u]\!],\ \forall j \in [\![n_l]\!] \tag{6.30i}$$

$$\alpha_k \in \mathbb{R}_+^{m_l}, \beta_k \in \mathbb{R}_+ \qquad \forall k \in [\![m_u]\!]. \tag{6.30j}$$

Problem (6.30) is a single-level problem and has a closed form. However, constraints (6.30h) contain bilinear terms, which cannot be tackled as efficiently as convex constraints by branch-and-cut based solvers. Therefore, we exploit the structure of the dual adversarial problem and its relation to the primal lower level to design a new efficient reformulation and solution algorithm.

6.4.1 Extended formulation

The bilinear constraints (6.30h) involve products of variables from the upper and lower level (x, v) as well as dual variables of each of the m_u dual-adversarial problems. For fixed values of (x, v), m_u dual adversarial subproblems (6.27) are defined. The optimal value of each k-th subproblem must be lower than $q_k - (Gx)_k$. The feasible region of each subproblem is defined by (6.30h-6.30j) and is independent of (x, v). The objective functions are linear in (α, β). Following Proposition 6.2.4, Problem (6.27) is bounded. If, moreover, Problem (6.27) is feasible, a vertex of the polytope (6.30h-6.30j) is an optimal solution. Following these observations, Constraints (6.30h-6.30j) can be replaced by disjunctive constraints, such that for each k, at least one extreme vertex of the k-th dual polyhedron is feasible. This reformulation of the bilinear constraints has, to the best of our knowledge, never been developed in the literature. We must highlight that disjunctive formulations are well established in the bilevel literature to express the complementarity constraints from the lower-level KKT conditions [129–131]. However, the bilinear reformulation of near-optimality robustness constraints does not possess the same structure and thus cannot leverage similar techniques.

Let $\mathcal{V}_k$ be the number of vertices of the k-th subproblem and α_k^l, β_k^l be the l-th vertex of the k-th subproblem. Constraints (6.30h-6.30j) can be written as:

$$\bigvee_{l=1}^{\mathcal{V}_k} \sum_{i=1}^{m_l} \alpha_{ki}^l (b - Ax)_i + \beta_k^l \cdot (d^T v + \delta) \leq q_k - (Gx)_k \qquad \forall k \in [\![m_u]\!], \tag{6.31}$$

where $\bigvee_{i=1}^N \mathcal{C}_i$ is the disjunction (logical "OR") operator, expressing the constraint that at least one of the constraints $\mathcal{C}_i$ must be satisfied. These disjunctions are equivalent to indicator constraints [132].

This reformulation of bilinear constraints based on the polyhedral description of the (α, β) feasible space is similar to the Benders decomposition [133]. Indeed in the near-optimal robust extended formulation, at least one of the vertices must satisfy a constraint (a disjunction) while Benders decomposition consists in satisfying a set of constraints for all extreme vertices and rays of the dual polyhedron (a constraint described with a universal quantifier). Disjunctive constraints (6.31) are equivalent to the following formulation, using set cover and SOS1 constraints:

$$\theta_k^l \in \{0,1\} \qquad \forall k, \forall l \tag{6.32a}$$

$$\omega_k^l \geq 0 \qquad \forall k, \forall l \tag{6.32b}$$

$$(b - Ax)^T \alpha_k^l + \beta_k^l (d^T v + \delta) - \omega_k^l \leq q_k - (Gx)_k \qquad \forall k, \forall l \tag{6.32c}$$

$$\sum_{l=1}^{\mathcal{V}_k} \theta_k^l \geq 1 \qquad \forall k \tag{6.32d}$$

$$SOS1(\theta_k^l, \omega_k^l) \qquad \forall k,\ \forall l, \tag{6.32e}$$

where $SOS1(a, b)$ expresses a SOS1-type constraint between the variables a and b.

In conclusion, using disjunctive constraints over the extreme vertices of each dual polyhedron and SOS1 constraints to linearize the complementarity constraints leads to an equivalent reformulation of Problem (6.30). The finite solution property holds even though the boundedness of the dual feasible set is not required. This single-level extended reformulation can be solved by any off-the-shelf MILP solver. Nevertheless, to decrease the computation time, we design a specific algorithm based on necessary conditions for the existence of a solution.

First, we illustrate the extended formulation on the following example.

6.4.2 Bounded example

Consider the bilevel linear problem defined by the following data:

$$x \in \mathbb{R}_+, y \in \mathbb{R}_+,$$

$$G = \begin{bmatrix} -1 \\ 1 \end{bmatrix} \; H = \begin{bmatrix} 4 \\ 2 \end{bmatrix} \; q = \begin{bmatrix} 11 \\ 13 \end{bmatrix} \; c_x = \begin{bmatrix} 1 \end{bmatrix} \; c_y = \begin{bmatrix} -10 \end{bmatrix},$$

$$A = \begin{bmatrix} -2 \\ 5 \end{bmatrix} \; B = \begin{bmatrix} -1 \\ -4 \end{bmatrix} \; b = \begin{bmatrix} -5 \\ 30 \end{bmatrix} \; d = \begin{bmatrix} 1 \end{bmatrix}.$$

The optimal solution of the high-point relaxation $(x, v) = (5, 4)$ is not bilevel-feasible. The optimal value of the optimistic bilevel problem is reached at $(x, v) = (1, 3)$. These two points are respectively represented by the blue diamond and red cross in Fig. 6.3. The dotted segments represent the upper-level constraints and the solid lines represent the lower-level constraints.

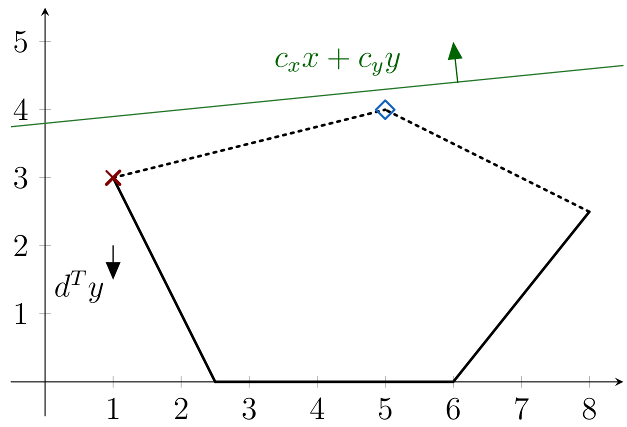

Figure 6.3 Representation of the bilevel problem.

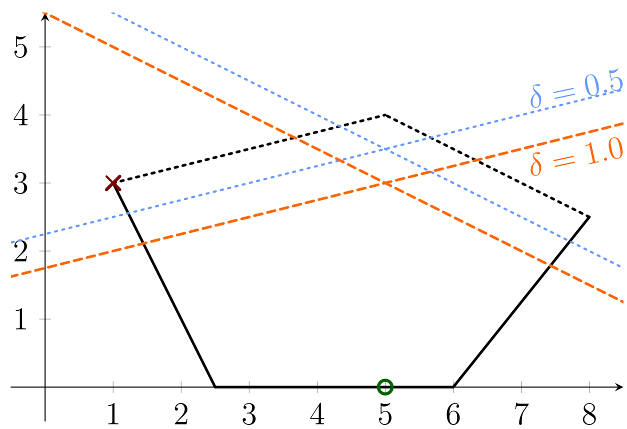

Figure 6.4 Near-optimal robustness constraints.

The feasible space for (α, β) is given by:

$$
\begin{aligned}
&-1\alpha_{11} - 4\alpha_{12} + \beta_1 \geq 4 \\
&-1\alpha_{21} - 4\alpha_{22} + \beta_2 \geq 2 \\
&\alpha_{ki} \geq 0, \beta_k \geq 0.
\end{aligned}
$$

This feasible space can be described as a set of extreme points and rays. It consists in this case of one extreme point $(\alpha_{ki} = 0, \beta_1 = 4, \beta_2 = 2)$ and 4 extreme rays. The (x, v) solution needs to be valid for the corresponding near-optimality conditions:

$$
\begin{aligned}
&\beta_1 \ (v + \delta) \leq 11 + x \\
&\beta_2 \ (v + \delta) \leq 13 - x.
\end{aligned}
$$

This results in two constraints in the (x, v) space, represented in Fig. 6.4 for $\delta = 0.5$ and $\delta = 1.0$ in dotted blue and dashed orange respectively. The radius of near-optimal feasibility $\hat{\delta} = 5$ can be computed using the formulation provided in Definition 6.2.1, for which the feasible domain at the upper-level is reduced to the point $x = 5$, for which $v = 0$, represented as a green circle at $(5, 0)$ in Fig. 6.4.

6.4.3 Solution algorithm

The solution procedure is defined as follows based on the structure of the extended formulation. The main goal of the algorithm is to prove infeasibility early in the resolution process and to solve the extended formulation only in the last step. Let $\mathcal{P}_0(BiP)$, $\mathcal{P}_1(BiP)$, feas_k, $\mathcal{P}_{no}(BiP;\delta)$ be the high-point relaxation, optimistic bilevel problem, feasibility of the dual adversarial problem and near-optimal robust problem respectively. Let C_k be the list of extreme vertices of the k-th dual adversarial polyhedron.

Algorithm 6.1 Near-Optimal Robust Vertex Enumeration Procedure (NORVEP)

function *near_optimal_bilevel*(BiP, δ)
$\diamond$ **Step 1**: dual subproblems expansion & pre-solving
for $k \in [\![m_u]\!]$ **do**
    Solve dual adversarial problem
    if feas_k = Infeasible **then**
        Terminate: k-th dual adversarial infeasible
    else
        $C_k \leftarrow (\alpha_k^l, \beta_k^l)_{l \in \mathcal{V}_k}$
    end if
end for
$\diamond$ **Step 2**: high-point relaxation $\mathcal{P}_0(BiP)$
if $\mathcal{P}_0(BiP)$ infeasible **then**
    Terminate: high-point relaxation infeasible
end if
$\diamond$ **Step 3**: optimistic relaxation $\mathcal{P}_1(BiP)$
if $\mathcal{P}_1(BiP)$ infeasible **then**
    Terminate: optimistic bilevel infeasible
end if
$\diamond$ **Step 4**: extended formulation $\mathcal{P}_{no}(BiP;\delta)$
Solve extended formulation with vertex list $(C_k)_{k \in [\![m_u]\!]}$, return solution information
end function

Each step consists in solving a problem that must be feasible for the feasibility of NORBiP to hold. The algorithm terminates without proceeding to the subsequent steps if an infeasibility is detected.

6.4.4 Valid inequalities

The extended formulation and Algorithm 6.1 can be directly applied. Nevertheless, we propose two groups of valid inequalities to tighten the formulation.

The first group of inequalities consists of the primal upper-level constraints:

$$(Gx)_k + (Hv)_k \leq q_k \qquad\qquad \forall z \in [\![m_u]\!].$$

These constraints are necessary for the optimistic formulation but not for the near-optimal robust one since they are always redundant with and included in the near-optimal robust constraints. However, their addition can strengthen the linear relaxation of the extended formulation and lead to faster convergence.

The second group of inequalities is defined in [134] and based on strong duality of the lower level. We only implement the valid inequalities for the root node, which are the primary focus of [134]:

$$\langle \lambda, b \rangle + \langle v, d \rangle \leq \langle A^+, \lambda \rangle, \tag{6.33}$$

where A_i^+ is an upper bound on $\langle A_i, x \rangle$. The computation of each upper bound A_i^+ relies on solving an auxiliary problem:

$$\begin{aligned}
A_i^+ = \max_{x,v,\lambda}\ & \langle A_i, x \rangle && (6.34a)\\
\text{s.t. }\ & Gx + Hv \leq q && (6.34b)\\
& Ax + Bv \leq b && (6.34c)\\
& d + B^T\lambda \geq 0 && (6.34d)\\
& x \in \mathcal{X}, v \geq 0, \lambda \geq 0 && (6.34e)\\
& (x, v, \lambda) \in \Upsilon, && (6.34f)
\end{aligned}$$

where Υ is the set containing all valid inequalities (6.33).

The method proposed in [134] relies on solving each i-th auxiliary problem once and using the resulting bound A^+. We define a new iterative procedure to improve the bounds computed at the root node:

1. Solve Problem (6.34a) $\forall i \in [\![m_l]\!]$ and obtain A^+;
2. If $\exists i,\ A_i^+$ is unbounded, terminate;
3. Otherwise, add Constraint (6.33) to (6.34f) and go to step 1;
4. Stopping criterion: when an iteration does not improve any of the bounds, terminate and return the last inequality with the sharpest bound.

This procedure allows tightening the bound as long as improvement can be made in one of the A_i^+. If the procedure terminates with one A_i^+ unbounded, the right-hand side of (6.33) is $+\infty$, the constraint is trivial and cannot be improved upon. Otherwise, each iteration improves the bound until the convergence of A^+.

6.5 Computational experiments

In this section, we demonstrate the applicability of our approach through numerical experiments on instances of the linear-linear near-optimal robust bilevel problem. We first describe the sets of test instances and the computational setup and then the experiments and their results.

6.5.1 Instance sets

Two sets of data are considered. For the first one, a total number of 1000 small, 200 medium and 100 large random instances are generated and characterized as follows:

$$\begin{aligned} (m_u, m_l, n_l, n_u) &= (5, 5, 5, 5) && \text{(small)} \\ (m_u, m_l, n_l, n_u) &= (10, 10, 10, 10) && \text{(medium)} \\ (m_u, m_l, n_l, n_u) &= (20, 10, 20, 20) && \text{(large).} \end{aligned}$$

All matrices are randomly generated with each coefficient having a 0.6 probability of being 0 and uniformly distributed on $[0, 1]$ otherwise. High-point feasibility and the vertex enumeration procedures are run after generating each tuple of random parameters to discard infeasible instances. Collecting 1000 small instances required generating 10532 trials, the 200 medium-sized instances were obtained with 18040 trials and the 100 large instances after

90855 trials. A second dataset is created from the 50 MIPS/Random instances of the Bilevel Problem library [135], where integrality constraints are dropped. All of these instances contain 20 lower-level constraints and no upper-level constraints. For each of them, two new instances are built by moving either the first 6 or the last 6 constraints from the lower to the upper level, resulting in 100 instances. We will refer to the first set of instances as the small/medium/large instances and the second as the MIPS instances. All instances are available in [136] in JLD format, along with a reader to import them in Julia programs.

6.5.2 Computational setup

Algorithm 6.1 is implemented in Julia [137] using the JuMP v0.21 modelling framework [107,108]; the MILP solver is SCIP 6.0 [138] with SoPlex 4.0 as the inner LP solver, both with default solving parameters. SCIP handles indicator constraints in the form of linear inequality constraints activated only if a binary variable is equal to one. *Polyhedra.jl* [139] is used to model the dual subproblem polyhedra with CDDLib [140] as a solver running the double-description algorithm, computing the list of extreme vertices and rays from the constraint-based representation. The exact rational representation of numbers is used in CDDLib instead of floating-point types to avoid rounding errors. Moreover, CDDLib fails to produce the list of vertices for some instances when set in floating-point mode. All experiments are performed on a consumer-end laptop with 15.5GB of RAM and an Intel i7 1.9GHz CPU running Ubuntu 18.04LTS.

6.5.3 Bilinear and extended formulation

To assess the efficiency of the extended formulation, we compare its solution time to that of the non-extended formulation including bilinear constraints (6.25). The bilinear formulation is implemented with SCIP using SoPlex as the linear optimization solver and Ipopt as the non-linear solver. SCIP handles the bilinear terms through bound computations and spatial branching. We test the two methods on 100 small instances. The bilinear version only manages to solve the small random instances and runs out of time or memory for all other instance sets. A time limit of 3600 seconds and a memory limit of 5000MB were fixed. The distribution of runtimes is presented in Fig. 6.5.

The extended formulation dominates at almost any time the bilinear formulation that uses spatial branching. The latter runs out of time or memory for most instances.

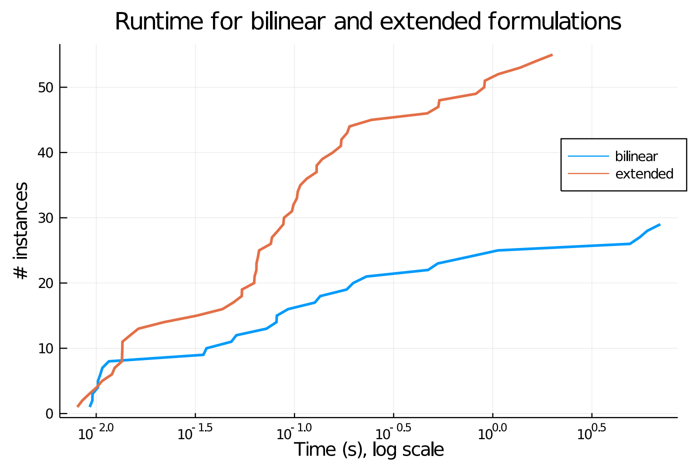

Figure 6.5 Runtime of the two methods on 100 of the small instances

6.5.4 Robustness of optimistic solutions and influence of δ

We solve the MIPS instances to bilevel optimality and verify the near-optimal robustness of the obtained solutions. We use various tolerance values:

$$\delta = \max(0.05, \delta_r \times \text{opt}(L))$$

with $\text{opt}(L)$ the lower-level objective value at the found solution and

$$\delta_r \in \{0.01, 0.05, 0.1, 0.5, 3.0\}.$$

Out of the 100 instances, 57 have canonical solutions that are not robust to even the smallest near-optimal deviation $0.01\text{opt}(L)$. Twelve more instances that have a near-optimal robust solution with the lowest tolerance are not near-optimal robust when the tolerance is increased to $3\text{opt}(L)$. Out of the 57 instances that are not near-optimal robust with the lowest tolerance, 40 have exactly one upper-level constraint that is violated by near-optimal deviations of the lower level and 17 that have more than one. Finally, we observe that the number of violated constraints changes across the range of tolerance values for 31 out of 100 instances. For the other 69 instances, the number of violated upper-level constraints remains identical for all tolerance values.

Table 6.1 summarizes the number of infeasible instances for different values of δ. As δ increases, so does the proportion of infeasible problems. This is due to the increase in the left-hand side in constraints (6.31).

In Fig. 6.6, we present the runtime difference between the canonical bilevel problem and its

Table 6.1 Number of infeasible problems for various tolerance levels δ

δ	0.01	0.1	0.2	1	3	5	7	10	12
Small (/1000)	366	423	466	595	658	670	672	674	676
Medium (/200)	78	88	95	118	122	123	123	123	123

near-optimal robust counterpart.

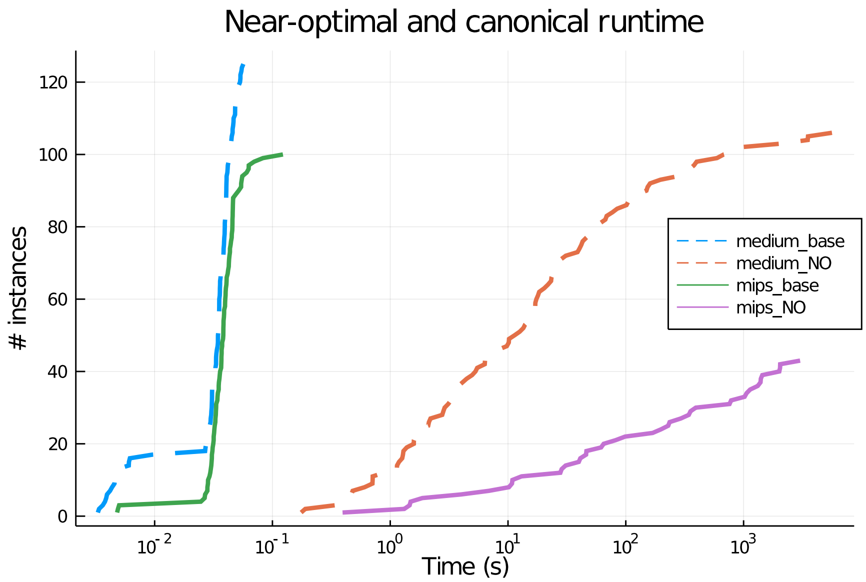

Figure 6.6 Runtime cost of adding near-optimality robustness constraints.

Scaling up the dimension of the tackled problems is limited not only due to high computation time but also due to memory requirements since the formulation of the problem requires allocating binary variables and a disjunctive constraint over all vertices of the dual polyhedron of each of the $k \in [\![m_u]\!]$ subproblems.

These runtime profiles highlight the fact that near-optimality robustness implemented using the extended formulation adds a significant runtime cost to the resolution of linear-linear bilevel problems. Nonetheless, the study of near-optimal lower-level decisions on optimistic solutions shows that these optimistic solutions are not robust, even for small tolerance values. This time difference also motivates the design of Algorithm 6.1. Indeed, since the optimistic bilevel problem is solved in a much shorter time, it is interesting to verify its feasibility before solving the near-optimal robust version.

6.5.5 Computational time of Algorithm 6.1

Statistics on the computation times of the two phases of Algorithm 6.1 for each instance size are provided in Table 6.2 and Table 6.3.

Table 6.2 Runtime statistics for the vertex enumeration (s).

Size	mean	10% quant.	50% quant.	90% quant.
Small	0.023	0.014	0.019	0.046
Medium	1.098	0.424	0.956	2.148
MIPS	21.061	0.231	3.545	65.004

Table 6.3 Runtime statistics for the optimization phase (s).

Instance type	# optimized	mean	10% quant.	50% quant.	90% quant.
Small	577	0.205	0.004	0.064	0.596
Medium	106	207.399	0.797	14.451	317.624
MIPS	70	909.302	57.592	344.202	2613.404

The solution time, corresponding to Steps 2-4 of Algorithm 6.1, is greater than the vertex enumeration phase, corresponding to Step 1, but does not dominate it completely for any of the problem sizes.

Figure 6.7 shows the distribution of the upper-level objective values across small and medium-sized instances. The number of problems solved to optimality monotonically decreases when δ increases (Table 6.1); greater δ values indeed make reduce the set of feasible solutions to *NORBiP*. The optimal values only slightly increase with δ and the lower-level objective value does not vary significantly with δ.

Even though more instances become infeasible as δ increases, the degradation of the objective value is in general insignificant for the optimal near-optimal robust solution compared to the optimistic solution.

6.5.6 Implementation of valid inequalities

In the last group of experiments, we implement and investigate the impact of the valid inequalities defined in Section 6.4.4.

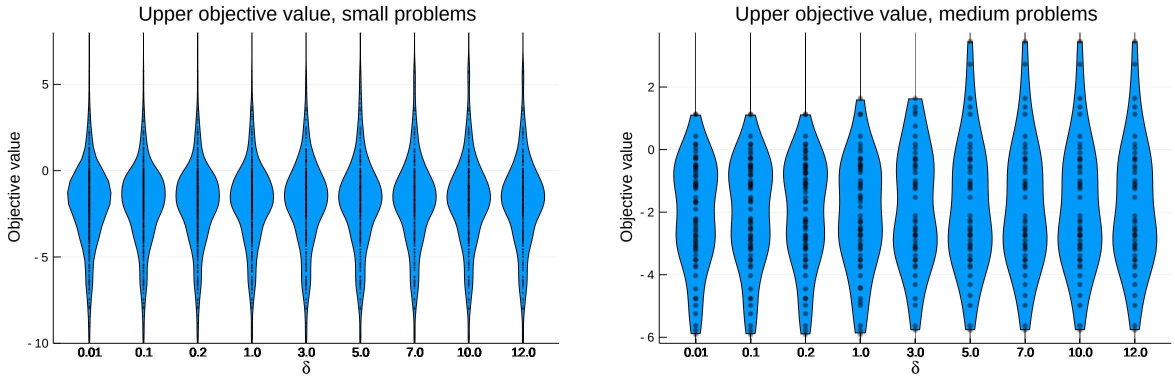

Figure 6.7 Violin plots of the upper objective value distributions versus δ.

On the 200 medium-sized instances, adding the valid inequality (6.33) is enough to prove the infeasibility of 61 instances out of 68 that are infeasible but possess a feasible high-point relaxation. On 100 large instances, adding the valid inequality proves the infeasibility of 29 out of 45 infeasible instances for which the high-point relaxation is feasible. For all medium and large instances, a non-trivial valid inequality i.e. where all A_i^+ are finite was computed.

These results highlight the improvement of the model tightness with the addition of the valid inequalities, compared to the high-point relaxation where primal and dual variables are subject to distinct groups of constraints. These inequalities thus discard infeasible instances without the need to solve the complete MILP reformulation. In Fig. 6.8, the distribution of the number of iterations of the inequality-finding procedure is presented for the medium and large instances. For the majority of instances of both sizes (about 80% and 60% of instances for the medium and large instances), a single iteration is sufficient to find the best valid inequality (6.33). The number of iteration, however, goes up to 40 and 50 for the medium and large instances respectively (truncated on the graph for clarity).

In Fig. 6.9, we compare the total runtime for MIPS and medium instances under near-optimality robustness constraints using $\delta = 0.1$ with and without valid inequalities for all instances solved to optimality. The runtime for instances with valid inequalities includes the runtime of the inequality computation.

Valid inequalities do not improve the runtime for NORBiP in either group of instances. This result is similar to the observations in [134] for instances of the canonical bilevel linear problem without near-optimality robustness.

We next study the inequalities based on the upper-level constraints on the small, medium

Figure 6.8 Distribution of the number of iterations for the computation of valid inequalities.

and MIPS instances.

As shown in Fig. 6.10, the addition of primal upper-level constraints accelerates the resolution of the MIPS and medium instances and dominates the standard extended formulation. For the small instances, we observe smaller runtimes for the first instances solved. This can be due to the upper-level constraints making the linear relaxation larger by adding constraints, thus creating overhead for smaller problems. This overhead is compensated for instances that are harder to optimize, i.e. that require more than 0.02 seconds to solve.

6.6 Conclusion

In this work, we introduce near-optimal robust bilevel optimization, a specific formulation of the bilevel optimization problem where the upper-level constraints are protected from deviations of the lower level from optimality. Near-optimality robustness challenges the assumption that the lower-level problem is solved to optimality, resulting in a generalized, more conservative formulation including the optimistic and pessimistic bilevel problems as special cases. We formulate NORBiP in the dependent case, i.e. where the upper- and lower-level constraints depend on both upper- and lower-level variables, thus offering a framework applicable to many bilevel problems of practical interest.

We derive a closed-form, single-level expression of NORBiP for convex lower-level problems,

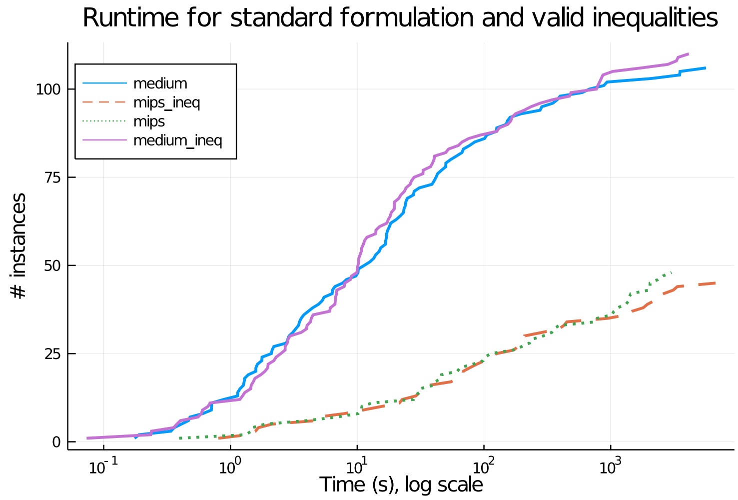

Figure 6.9 Runtime for MIPS and medium instances with and without valid inequalities.

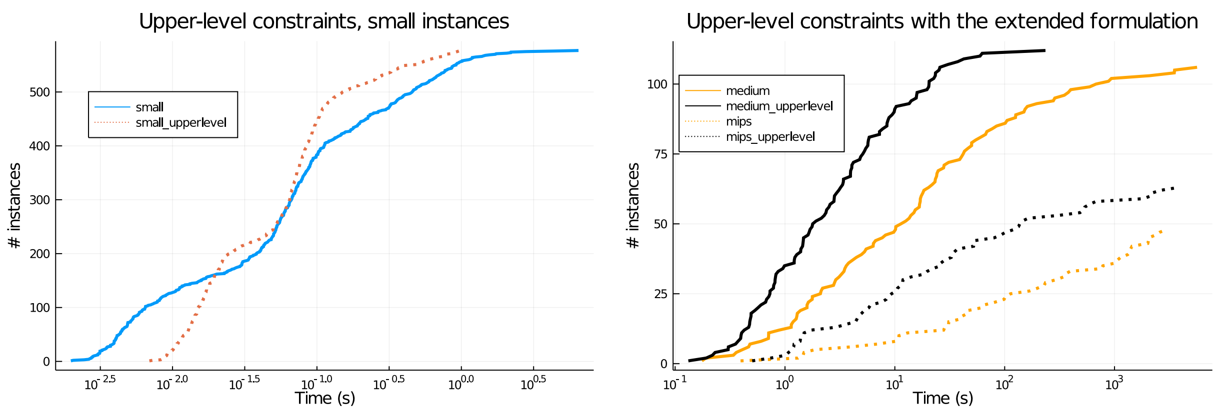

Figure 6.10 Runtime for small, medium and MIPS instances with and without upper-level constraints.

based on dual adversarial certificates to guarantee near-optimality robustness. In the linear case, we derive an extended formulation that can be represented as a MILP with indicator constraints. Numerical experiments highlight the efficiency of the extended method compared to the compact bilinear formulation and the impact of some valid inequalities on both solution time and tightness of the linear relaxation.

CHAPTER 7 ARTICLE 3: COMPLEXITY OF NEAR-OPTIMAL ROBUST VERSIONS OF MULTILEVEL OPTIMIZATION PROBLEMS

This chapter was submitted to Optimization Letters in October 2020.

Authors: Mathieu Besançon, Miguel F. Anjos, Luce Brotcorne

Abstract

Near-optimality robustness extends multilevel optimization with a limited deviation of a lower level from its optimal solution, anticipated by higher levels. We analyze the complexity of near-optimal robust multilevel problems, where near-optimal robustness is modelled through additional adversarial decision-makers. Near-optimal robust versions of multilevel problems are shown to remain in the same complexity class as the problem without near-optimality robustness under general conditions.

Multilevel optimization is a class of mathematical optimization problems where other problems are embedded in the constraints. They are well suited to model sequential decision-making processes, where a first decision-maker, the leader intrinsically integrates the reaction of another decision-maker, the follower, into their decision-making problem.

In recent years, most of the research focuses on the study and design of efficient solution methods for the case of two levels, namely bilevel problems [98], which fostered a growing range of applications.

Near-optimal robustness, defined in [141], is an extension of bilevel optimization. In this setting, the upper level anticipates limited deviations of the lower level from an optimal solution and aims at a solution that remains feasible for any feasible and near-optimal solution of the lower level. This protection of the upper level against uncertain deviations of the lower-level has led to the characterization of near-optimality robustness as a robust optimization approach for bilevel optimization. The lower-level response corresponds to the uncertain parameter and the maximum deviation of the objective value from an optimal solution to the uncertainty budget. Because the set of near-optimal lower-level solutions potentially has infinite cardinality and depends on the upper-level decision itself, near-optimality robustness adds generalized semi-infinite constraints to the bilevel problem. The additional constraint can also be viewed as a form of robustness under decision-dependent uncertainty.

In this paper, we prove complexity results on multilevel problems to which near-optimality robustness constraints are added under various forms. We show that under fairly general conditions, the near-optimal robust version of a multilevel problem remains on the same level of the polynomial hierarchy as the canonical problem. These results are non-trivial assuming that the polynomial hierarchy does not collapse and open the possibility of solution algorithms for near-optimal robust multilevel problems as efficient as for their canonical counterpart. Even though we focus on near-optimal robust multilevel problems, the complexity results we establish hold for all multilevel problems that present the same hierarchical structure, i.e. the same anticipation and parameterization between levels as the near-optimal formulation with the adversarial problems, as defined in Section 7.2.

The rest of this paper is organized as follows. Section 7.1 introduces the notation and the background on near-optimality robustness and existing complexity results in multilevel optimization. Section 7.2 presents complexity results for the near-optimal robust version of bilevel problems, where the lower level belongs to $\mathcal{P}$ and $\mathcal{NP}$. These results are extended in Section 7.3 to multilevel optimization problems, focusing on integer multilevel linear problems with near-optimal deviations of the topmost intermediate level. Section 7.4 provides complexity results for a generalized form of near-optimal robustness in integer multilevel problems, where multiple decision-makers anticipate near-optimal reactions of a lower level. Finally, we draw some conclusions in Section 7.5.

7.1 Background on near-optimality robustness and multilevel optimization

In this section, we introduce the notation and terminology for bilevel optimization and near-optimality robustness, and highlight prior complexity results in multilevel optimization. Let us define a bilevel problem as:

$$\min_{x} \; F(x,v) \tag{7.1a}$$

$$\text{s.t.} \; G_k(x,v) \leq 0 \qquad \forall k \in [\![m_u]\!] \tag{7.1b}$$

$$x \in \mathcal{X} \tag{7.1c}$$

$$\text{where } v \in \arg\min_{y\in\mathcal{Y}} \{f(x,y) \text{ s.t. } g_i(x,y) \leq 0 \; \forall i \in [\![m_l]\!]\}. \tag{7.1d}$$

We denote by $\mathcal{X}$ and $\mathcal{Y}$ the domain of upper- and lower-level variables respectively. We use the convenience notation $[\![n]\!] = \{1, \ldots, n\}$ for a natural n.

Problem (7.1) is ill-posed, since multiple solutions to the lower level may exist [142, Ch. 1]. Models often rely on additional assumptions to alleviate this ambiguity, the two most common being the optimistic and pessimistic approaches. In the optimistic case (BiP), the lower level selects an optimal decision that most favours the upper level. In this setting, the lower-level decision can be taken by the upper level, as long as it is optimal for the lower-level problem. The upper level can thus optimize over both x and v, leading to:

$$
\begin{aligned}
\text{(BiP):}\ \min_{x,v}\ & F(x,v) && && \text{(7.2a)}\\
\text{s.t.}\ & G_k(x,v) \leq 0 && \forall k \in [\![m_u]\!] && \text{(7.2b)}\\
& x \in \mathcal{X} && && \text{(7.2c)}\\
& v \in \arg\min_{y\in\mathcal{Y}}\{f(x,y) \text{ s.t. } g_i(x,y) \leq 0\ \forall i \in [\![m_l]\!]\}. && && \text{(7.2d)}
\end{aligned}
$$

Constraint (7.2d) implies that v is feasible for the lower level and that $f(x,v)$ is the optimal value of the lower-level problem, parameterized by x.

The pessimistic approach assumes that the lower level chooses an optimal solution that is the worst for the upper-level objective as in [98] or with respect to the upper-level constraints as in [48].

The near-optimal robust version of (BiP) considers that the lower-level solution may not be optimal but near-optimal with respect to the lower-level objective function. The tolerance for near-optimality, denoted by δ is expressed as a maximum deviation of the objective value from optimality. The problem solved at the upper level must integrate this deviation and protects the feasibility of its constraints for any near-optimal lower-level decision. The problem is formulated as:

$$
\begin{aligned}
\text{(NORBiP):}\ \min_{x,v}\ & F(x,v) && && \text{(7.3a)}\\
\text{s.t.}\ & \text{(7.2b)} - \text{(7.2d)} && && \text{(7.3b)}\\
& G_k(x,z) \leq 0\ \forall z \in \mathcal{Z}(x;\delta) && \forall k \in [\![m_u]\!] && \text{(7.3c)}\\
& \text{where } \mathcal{Z}(x;\delta) = \{y \in \mathcal{Y} \mid f(x,y) \leq f(x,v) + \delta, g(x,y) \leq 0\}. && && \text{(7.3d)}
\end{aligned}
$$

$\mathcal{Z}(x;\delta)$ denotes the near-optimal set, i.e. the set of near-optimal lower-level solutions, depending on both the upper-level decision x and δ. (NORBiP) is a generalization of the pessimistic bilevel problem since the latter is both a special case and a relaxation of (NOR-

BiP) [141]. We refer to (BiP) as the **canonical** problem for (NORBiP) (or equivalently Problem (7.4)) and (NORBiP) as the near-optimal robust version of (BiP). In the formulation of (NORBiP), the upper-level objective depends on decision variables of both levels, but is not protected against near-optimal deviations. A more conservative formulation also protecting the objective by moving it to the constraints in an epigraph formulation [141] is given by:

$$
\begin{aligned}
\text{(NORBiP-Alt):}\; \min_{x,v,\tau}\; & \tau \\
\text{s.t.}\; & (7.2\text{b}) - (7.2\text{d}) \\
& G_k(x,z) \leq 0 \;\forall z \in \mathcal{Z}(x;\delta) \qquad\qquad \forall k \in [\![m_u]\!] \\
& F(x,z) \quad \leq \tau \;\forall z \in \mathcal{Z}(x;\delta),
\end{aligned}
$$

The optimal values of the three problems are ordered as:

$$\text{opt(BiP)} \leq \text{opt(NORBiP)} \leq \text{opt(NORBiP-Alt)}.$$

We next provide a review of complexity results for bilevel and multilevel optimization problems. Bilevel problems are $\mathcal{NP}$-hard in general, even when the objective functions and constraints at both levels are linear [44]. When the lower-level problem is convex, a common solution approach consists in replacing it with its KKT conditions [143, 144], which are necessary and sufficient if the problem satisfies certain constraint qualifications. This approach results in a single optimization problem with complementarity constraints, of which the decision problem is $\mathcal{NP}$-complete [145]. A specific form of the three-level problem is investigated in [146], where only the objective value of the bottom-level problem appears in the objective functions of the first and second levels. If these conditions hold and all objectives and constraints are linear, the problem can be reduced to a single level one with complementarity constraints of polynomial size. This model is similar to the worst-case reformulation of (NORBiP) presented in Section 7.2.

Pessimistic bilevel problems for which no upper-level constraint depends on lower-level variables are studied in [80]. The problem of finding an optimal solution to the pessimistic case is shown to be $\mathcal{NP}$-hard, even if a solution to the optimistic counterpart of the same problem is provided. A variant is also defined, where the lower level may pick a suboptimal response only impacting the upper-level objective. This variant is comparable to the Objective-Robust Near-Optimal Bilevel Problem defined in [141]. In [48], the independent case of the pessimistic bilevel problem is studied, corresponding to a special case of (NOR-BiP) with $\delta = 0$ and all lower-level constraints independent of the upper-level variables. It is shown that the linear independent pessimistic bilevel problem, and consequently the linear near-optimal robust bilevel problem, can be solved in polynomial time while it is strongly $\mathcal{NP}$-hard in the non-linear case.

When the lower-level problem cannot be solved in polynomial time, the bilevel problem is in general Σ_2^P-hard. The notion of Σ_2^P-hardness and classes of the polynomial hierarchy are recalled in Section 7.2. Despite this complexity result, new algorithms and corresponding implementations have been developed to solve these problems and in particular, mixed-integer linear bilevel problems [69,73,147]. Variants of the bilevel knapsack were investigated in [148], and proven to be Σ_2^P-hard as the generic mixed-integer bilevel problem.

Multilevel optimization was initially investigated in [46] in the case of linear constraints and objectives at all levels. In this setting, the problem is shown to be in Σ_s^P, with $s+1$ being the number of levels. The linear bilevel problem corresponds to $s = 1$ and is in $\Sigma_1^P \equiv \mathcal{NP}$. If, on the contrary, at least the bottom-level problem involves integrality constraints (or more generally belongs to $\mathcal{NP}$ but not $\mathcal{P}$), the multilevel problem with s levels belongs to Σ_s^P. A model unifying multistage stochastic and multilevel problems is defined in [87], based on a risk function capturing the component of the objective function which is unknown to a decision-maker at their stage, either because of a stochastic component or of another decision-maker that acts after the current one.

As highlighted in [87, 149], most results in the literature on complexity of multilevel optimization use $\mathcal{NP}$-hardness as the sole characterization. This only indicates that a given problem is at least as hard as all problems in $\mathcal{NP}$ and that no polynomial-time solution method should be expected unless $\mathcal{NP} = \mathcal{P}$.

We characterize near-optimal robust multilevel problems not only on the hardness or "lower bound" on complexity, i.e. being at least as hard as all problems in a given class but through their complexity "upper bound", i.e. the class of the polynomial hierarchy they belong to. The linear optimistic bilevel problem is for instance strongly $\mathcal{NP}$-hard, but belongs to $\mathcal{NP}$ and is therefore not Σ_2^P-hard.

7.2 Complexity of near-optimal robust bilevel problems

We establish in this section complexity results for near-optimal robust bilevel problems for which the lower level $\mathcal{L}$ is a single-level problem parameterized by the upper-level decision. (NORBiP) can be reformulated by replacing each k-th semi-infinite Constraint (7.3c) with the lower-level solution z_k in $\mathcal{Z}(x;\delta)$ that yields the highest value of $G_k(x, z_k)$:

$$\min_{x,v} \; F(x,v) \tag{7.4a}$$
$$\text{s.t. } (7.2\text{b}) - (7.2\text{d}) \tag{7.4b}$$
$$G_k(x, z_k) \leq 0 \qquad \forall k \in [\![m_u]\!] \tag{7.4c}$$
$$z_k \in \arg\max_{y\in\mathcal{Y}} \{G_k(x,y) \text{ s.t. } f(x,y) \leq f(x,v) + \delta, g(x,y) \leq 0\} \qquad \forall k \in [\![m_u]\!] \tag{7.4d}$$

From a game-theoretical perspective, the near-optimal robust version of a bilevel problem can be seen as a three-player hierarchical game. The upper level $\mathcal{U}$ and lower level $\mathcal{L}$ are identical to the canonical bilevel problem. The third level is the adversarial problem $\mathcal{A}$ and selects the worst near-optimal lower-level solution with respect to upper-level constraints, as represented by the embedded maximization in Constraint (7.4d). If the upper-level problem has multiple constraints, the adversarial problem can be decomposed into problems $\mathcal{A}_k, k \in [\![m_u]\!]$, where m_u is the number of upper-level constraints. The interaction among the three players is depicted in Fig. 7.1. The blue dashed arcs represent a parameterization of the source vertex by the decisions of the destination vertex, and the solid red arcs represent an anticipation of the destination vertex decisions in the problem of the source vertex. This convention will be used in all figures throughout the paper.

The adversarial problem can be split into m_u adversarial problems as done in [141], each finding the worst-case with respect to one of the upper-level constraints. The canonical problem refers to the optimistic bilevel problem without near-optimal robustness constraints. We refer to the variable v as the canonical lower-level decision.

The complexity classes of the polynomial hierarchy are only defined for decision problems. We consider that an optimization problem belongs to a given class if that class contains the decision problem of determining if there exists a feasible solution for which the objective value at least as good as a given bound.

Definition 7.2.1. *The decision problem associated with an optimization problem is in $\mathcal{P}^*[\mathcal{H}]$, with $\mathcal{H}$ a set of real-valued functions on a vector space $\mathcal{Y}$, iff:*

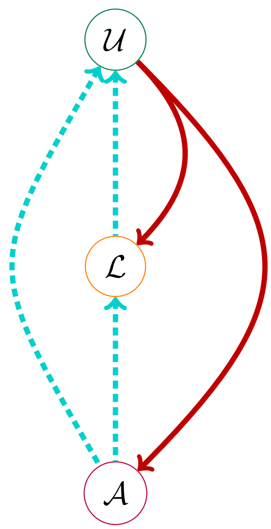

Figure 7.1 Near-optimal robust bilevel problem

1. *it belongs to $\mathcal{P}$;*
2. *for any $h \in \mathcal{H}$, the problem with an additional linear constraint and an objective function set as $h(\cdot)$ is also in $\mathcal{P}$.*

A broad range of problems in $\mathcal{P}$ are also in $\mathcal{P}^*[\mathcal{H}]$ for certain sets of functions $\mathcal{H}$ (see Example 7.2.1 for linear problems and linear functions and Example 7.2.2 for some combinatorial problems in $\mathcal{P}$). $\mathcal{NP}^*[\mathcal{H}]$ and $\Sigma_s^{P*}[\mathcal{H}]$ are defined in a similar way. We next consider two examples illustrating these definitions.

Example 7.2.1. *Denoting by $\mathcal{H}_L$ the set of linear functions from the space of lower-level variables to $\mathbb{R}$, linear optimization problems are in $\mathcal{P}^*[\mathcal{H}_L]$, since any given problem with an additional linear constraint and a different linear objective function is also a linear optimization problem.*

Example 7.2.2. *Denoting by $\mathcal{H}_L$ the set of linear functions from the space of lower-level variables to $\mathbb{R}$, combinatorial optimization problems in $\mathcal{P}$ which can be formulated as linear optimization problems with totally unimodular matrices are not in $\mathcal{P}^*[\mathcal{H}_L]$ in general. Indeed, adding a linear constraint may break the integrality of solutions of the linear relaxation of the lower-level problem.*

Σ_s^P is the complexity class at the s-th level of the polynomial hierarchy [45, 46], defined recursively as $\Sigma_0^P = \mathcal{P}$, $\Sigma_1^P = \mathcal{NP}$, and problems of the class Σ_s^P, $s > 1$ being solvable in non-deterministic polynomial time, provided an oracle for problems of class Σ_{s-1}^P. In particular, a positive answer to a decision problem in $\mathcal{NP}$ can be verified, given a certificate, in polynomial time. If the decision problem associated with an optimization problem is in $\mathcal{NP}$, and given a potential solution, the objective value of the solution can be compared to a given bound and the feasibility can be verified in polynomial time. We reformulate these statements in the following proposition:

Proposition 7.2.1. *[46] An optimization problem is in Σ_{s+1}^P if verifying that a given solution is feasible and attains a given bound can be done in polynomial time, when equipped with an oracle solving problems in Σ_s^P in a single step.*

Proposition 7.2.1 is the main property of the classes of the polynomial hierarchy used to determine the complexity of near-optimal robust bilevel problems in various settings throughout this paper.

Lemma 7.2.1. *Given a bilevel problem in the form of Problem* (7.2)*, if the lower-level problem is in $\mathcal{P}^*[\mathcal{H}]$, and*

$$-G_k(x,\cdot) \in \mathcal{H} \;\forall x, \forall k \in [\![m_u]\!],$$

then the adversarial problem (7.4d) *is in $\mathcal{P}$.*

Proof. The lower-level problem can equivalently be written in an epigraph form:

$$\begin{aligned}(v,w) \in \arg\min_{y,u}\ & u\\ \text{s.t. }\ & f(x,y) - u \leq 0\\ & g(x,y) \leq 0.\end{aligned}$$

Given a solution of the lower-level problem (v,w) and an upper-level constraint $G_k(x,y) \leq 0$, the adversarial problem is defined by:

$$\begin{aligned}\min_{y,u}\ & -G(x,y)\\ \text{s.t. }\ & f(x,y) - u \leq 0\\ & g(x,y) \leq 0\\ & u \leq w.\end{aligned}$$

Compared to the lower-level problem, the adversarial problem contains an additional linear constraint $u \leq w$ and an objective function updated to $-G(x,\cdot)$. $\square$

Theorem 7.2.1. *Given a bilevel problem (P), if there exists $\mathcal{H}$ such that the lower-level problem is in $\mathcal{NP}^*[\mathcal{H}]$ and*

$$-G_k(x,\cdot) \in \mathcal{H}\ \forall x \in \mathcal{X}, \forall k \in [\![m_u]\!],$$

then the near-optimal robust version of the bilevel problem is in Σ_2^P like the canonical bilevel problem.

Proof. The proof relies on the ability to verify that a given solution (x, v) results in an objective value at least as low as a bound Γ according to Proposition 7.2.1. This verification can be carried out with the following steps:

1. Compute the upper-level objective value $F(x, v)$ and verify that $F(x, v) \leq \Gamma$;
2. Verify that upper-level constraints are satisfied;
3. Verify that lower-level constraints are satisfied;
4. Compute the optimum value $\mathcal{L}(x)$ of the lower-level problem parameterized by x and check if:
$$f(x, v) \leq \min_y \mathcal{L}(x);$$
5. Compute the worst case:
Find
$$z_k \in \arg\max_{y \in \mathcal{Y}} \mathcal{A}_k(x, v)\ \ \forall k \in [\![m_u]\!];$$
where $\mathcal{A}_k(x, v)$ is the k-th adversarial problem parameterized by (x, v);
6. Verify near-optimal robustness: $\forall k \in [\![m_u]\!]$, verify that the k-th upper-level constraint is feasible for the worst-case z_k.

Steps 1 and 2 can be carried out in polynomial time by assumption. Step 3 requires to check the feasibility of a solution to a problem in $\mathcal{NP}$. This can be done in polynomial time. Step 4 consists in solving the lower-level problem, while Step 5 corresponds to solving m_u problems similar to the lower level, with the objective function modified and an additional linear constraint ensuring near-optimality. □

Theorem 7.2.2. *Given a bilevel problem (P), if the lower-level problem is convex and in $\mathcal{P}^*[\mathcal{H}]$ with $\mathcal{H}$ a set of convex functions, and if the upper-level constraints are such that $-G_k(x,\cdot) \in \mathcal{H}$, then the near-optimal robust version of the bilevel problem is in $\mathcal{NP}$. If the* *upper-level constraints are convex non-affine with respect to the lower-level constraints, the near-optimal robust version is in general not in $\mathcal{NP}$.*

Proof. If the upper-level constraints are concave with respect to the lower-level variables, the adversarial problem defined as:

$$\max_{y \in \mathcal{Y}} \; G_k(x, y) \tag{7.5a}$$

$$\text{s.t.} \;\; g(x, y) \leq 0 \tag{7.5b}$$

$$f(x, y) \leq f(x, v) + \delta \tag{7.5c}$$

is convex. Furthermore, by definition of $\mathcal{P}^* [\mathcal{H}]$, the adversarial problem is in $\mathcal{P}$.

Applying the same reasoning as in the proof of Theorem 7.2.1, Steps 1-3 are identical and can be carried out in polynomial time. Step 4 can be performed in polynomial time since $\mathcal{L}$ is in $\mathcal{P}$. Step 5 is also performed in polynomial time since $\forall k \in [\![m_u]\!]$, each k-th adversarial problem (7.5) is a convex problem that can be solved in polynomial time since $\mathcal{L}$ is in $\mathcal{P}^* [\mathcal{H}]$. Step 6 simply is a simple comparison of two quantities.

If the upper-level constraints are convex non-affine with respect to the lower-level variables, Problem (7.5) maximizes a convex non-affine function over a convex set. Such a problem is $\mathcal{NP}$-hard in general. Therefore, the verification that a given solution is feasible and satisfies a predefined bound on the objective value requires solving the m_u $\mathcal{NP}$-hard adversarial problems. If $\mathcal{L}$ is in $\mathcal{NP}^* [\mathcal{H}]$, then these adversarial problems are in $\mathcal{NP}$ by Eq. (7.3), and the near-optimal robust problem is in Σ_2^P according to Proposition 7.2.1. □

7.3 Complexity of near-optimal robust mixed-integer multilevel problems

In this section, we study the complexity of a near-optimal robust version of mixed-integer multilevel linear problems (MIMLP), where the lower level itself is a s-level problem and is Σ_s^P-hard. The canonical multilevel problem is, therefore, Σ_{s+1}^P-hard [46]. For some instances of mixed-integer bilevel, the optimal value can be approached arbitrarily but not reached [150]. To avoid such pathological cases, we restrict our attention to multilevel problems satisfying the criterion for mixed-integer bilevel problems from [69]:

Property 7.3.1. *The continuous variables at any level s do not appear in the problems at levels that are lower than s (the levels deciding after s).*

More specifically, we will focus on mixed-integer multilevel linear problems where the uppermost lower level $\mathcal{L}_1$ may pick a solution deviating from the optimal value, while we ignore deviations of the levels $\mathcal{L}_{i>1}$. This problem is noted (NOMIMLP$_s$) and depicted in Fig. 7.1.

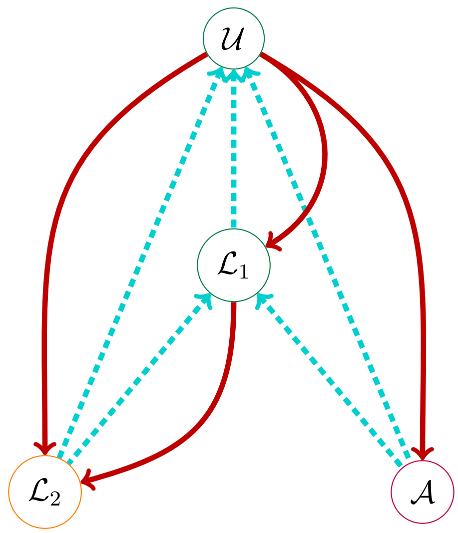

Figure 7.2 Near-optimality robustness for multilevel problems

The adversarial problem corresponds to a decision of the level $\mathcal{L}_1$ different from the canonical decision. This decision induces a different reaction from the subsequent levels $\mathcal{L}_2$, $\mathcal{L}_3$. Since the top-level constraints depend on the joint reaction of all following levels, we will note $z_{ki} = (z_{k1}, z_{k2}, z_{k3})$ the worst-case joint near-optimal solution of all lower levels with respect to the top-level constraint k.

Theorem 7.3.1. *If $\mathcal{L}_1$ is in $\Sigma_s^{P*}[\mathcal{H}_L]$, the decision problem associated with (NOMIMLP$_s$) is in Σ_{s+1}^P as the canonical multilevel problem.*

Proof. Given a solution to all levels $(x_U, v_1, v_2, ...v_s)$ and a bound Γ, verifying that this solution is (i) feasible, (ii) near-optimal robust of parameter δ, and (iii) has an objective value at least as good as the bound Γ can be done through the following steps:

1. Compute the objective value and verify that it is lower than Γ;
2. Verify variable integrality;
3. Solve the problem $\mathcal{L}_1$, parameterized by x_U, and verify that the solution $(v_1, v_2, ...)$ is optimal;
4. $\forall k \in [\![m_u]\!]$, solve the k-th adversarial problem. Let $z_k = (z_{k1}, z_{k2}...z_{ks})$ be the solution;

5. $\forall k \in [\![m_u]\!]$, verify that the k-th upper-level constraint is feasible for the adversarial solution z_k.

Steps 1, 2, and 5 can be performed in polynomial time. Step 3 requires solving a problem in Σ_s^P, while step 4 consists in solving m_u problems in Σ_s^P, since $\mathcal{L}_1$ is in $\Sigma_s^{P*}[\mathcal{H}_L]$. Checking the validity of a solution thus requires solving problems in Σ_s^P and is itself in Σ_{s+1}^P, similarly to the canonical problem. □

7.4 Complexity of a generalized near-optimal robust multilevel problem

In this section, we study the complexity of a variant of the problem presented in Section 7.3 with $s+1$ decision-makers at multiple top levels $\mathcal{U}_1$,$\mathcal{U}_2$, ...$\mathcal{U}_s$ and a single bottom level $\mathcal{L}$. We denote by $\mathcal{U}_1$ the top-most level. We assume that the bottom-level entity may choose a solution deviating from optimality. This requires that the entities at all $\mathcal{U}_i \, \forall i \in \{1..s\}$ levels anticipate this deviation, thus solving a near-optimal robust problem to protect their feasibility from it. The variant, coined (GNORMP$_s$), is illustrated in Fig. 7.3. We assume throughout this section that Property 7.3.1 holds in order to avoid the unreachability problem previously mentioned. The decision variables of all upper levels are denoted by $x_{(i)}$, and the objective functions by $F_{(i)}(x_{(1)}, x_{(2)}...x_{(s)})$. The lower-level canonical decision is denoted v as in previous sections.

If the lowest level $\mathcal{L}$ belongs to $\mathcal{NP}$, $\mathcal{U}_s$ belongs to Σ_2^P and the original problem is in Σ_{s+1}^P. In a more general multilevel case, if the lowest level $\mathcal{L}$ solves a problem in Σ_r^P, $\mathcal{U}_s$ solves a problem in Σ_{r+1}^P and $\mathcal{U}_1$ in Σ_{r+s}^P.

We note that for all fixed decisions $x_{(i)} \forall i \in \{1..s-1\}$, $\mathcal{U}_s$ is a near-optimal robust bilevel problem. This differs from the model presented in Section 7.3 where, for a fixed upper-level decision, the top-most lower level $\mathcal{L}_1$ is the same parameterized problem as in the canonical setting. Furthermore, as all levels $\mathcal{U}_i$ anticipate deviations of the lower-level decision in the near-optimal set, the worst case can be formulated with respect to the constraints of each of these levels. In conclusion, distinct adversarial problems $\mathcal{A}_i \, \forall i \in \{1..s\}$ can be formulated. Each upper level $\mathcal{U}_i$ integrates the reaction of the corresponding adversarial problem in its near-optimality robustness constraint. This formulation of (GNORMP$_s$) is depicted Fig. 7.4.

Theorem 7.4.1. *Given a $s+1$-level problem (GNORMP$_s$), if the bottom-level problem parameterized by all upper-level decisions $\mathcal{L}(x_{(1)}, x_{(2)}...x_{(s)})$ is in $\Sigma_r^{P*}[\mathcal{H}_L]$, then (GNORMP$_s$) is in Σ_{r+s}^P like the corresponding canonical bilevel problem.*

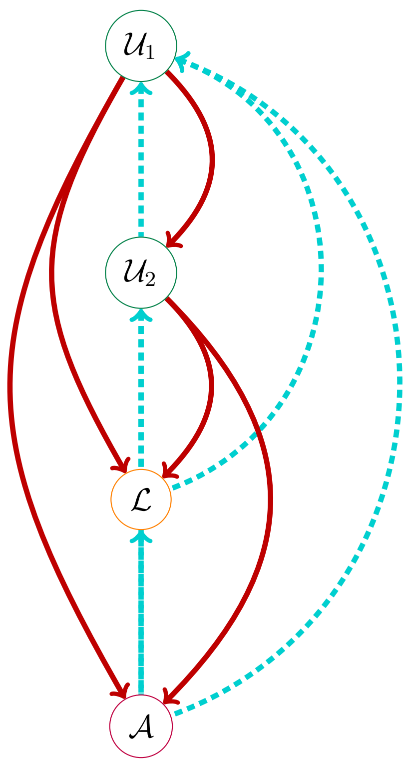

Figure 7.3 Generalized NOR multi-level

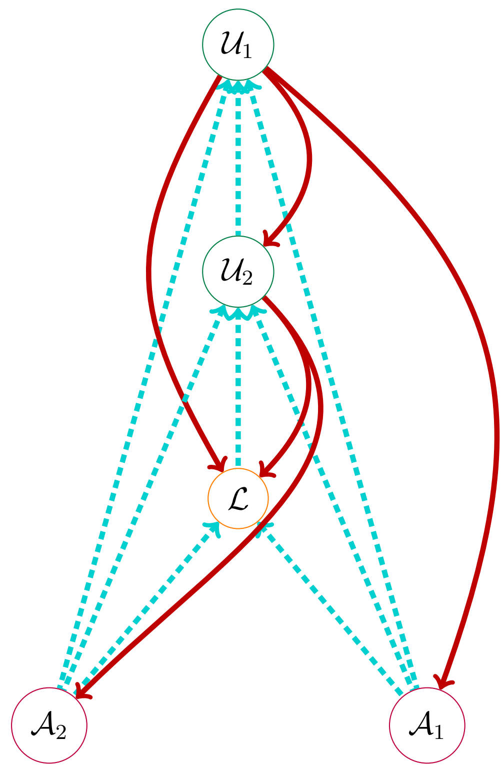

Figure 7.4 Decoupling of the two adversarial problems

Proof. We denote by $x_U = (x_{(1)}, x_{(2)}, ...x_{(s)})$ and m_{U_i} the number of constraints of problem $\mathcal{U}_i$. As for Theorem 7.2.1, this proof is based on the complexity of verifying that a given solution (x_U, v) is feasible and results in an objective value below a given bound. The verification requires the following steps:

1. Compute the top-level objective value and assert that it is below the bound;

2. Verify feasibility of (x_U, v) with respect to the constraints at all levels;

3. Verify optimality of v for $\mathcal{L}$ parameterized by x_U;

4. Verify optimality of $x_{(i)}$ for the near-optimal robust problem solved by the i-th level $\mathcal{U}_i(x_{(1)}, x_{(2)}...x_{(i-1)}; \delta)$ parameterized by all the decisions at levels above and the near-optimality tolerance δ;

5. Compute the near-optimal lower-level solution z_k which is the worst-case with respect to the k-th constraint of the top-most level $\forall k \in [\![m_{U_1}]\!]$;

6. Verify that each $k \in [\![m_{U_1}]\!]$ top-level constraint is satisfied with respect to the corresponding worst-case solution z_k.

Steps 1-2 are performed in polynomial time. Step 3 requires solving Problem $\mathcal{L}(x_U)$, belonging to Σ_r^P. Step 4 consists in solving a generalized near-optimal robust multilevel problem GNORMP$_{s-1}$ with one level less than the current problem. Step 5 requires the solution of m_{U_1} adversarial problems belonging to Σ_r^P since $\mathcal{L}$ is in $\Sigma_r^{P*}[\mathcal{H}_L]$. Step 6 is an elementary comparison of two quantities for each $k \in [\![m_u]\!]$. The step of highest complexity is Step 4. If it requires to solve a problem in Σ_{r+s-1}^P, then GNORMP$_s$ is in Σ_{r+s}^P similarly to its canonical problem.

Let us assume that Step 4 requires to solve a problem outside Σ_{r+s-1}^P. Then GNORMP$_{s-1}$ is not in Σ_{r+s-1}^P as the associated canonical problem, and that Step 4 requires to solve a problem not in Σ_{r+s-2}^P. By recurrence, GNORMP$_1$ is not in Σ_{r+1}^P. However, GNORMP$_1$ is a near-optimal robust bilevel problem where the lower level itself in Σ_r^P; this corresponds to the setting of Section 7.3. This contradicts Theorem 7.3.1, GNORMP$_{s-1}$ is therefore in Σ_{r+s-1}^P. Verifying the feasibility of a given solution to GNORMP$_s$ requires solving a problem at most in Σ_{r+s-1}^P. Based on Proposition 7.2.1, GNORMP$_s$ is in Σ_{r+s}^P as its canonical multilevel problem. $\square$

In conclusion, Theorem 7.4.1 shows that adding near-optimality robustness at an arbitrary level of the multilevel problem does not increase its complexity in the polynomial hierarchy. By combining this property with the possibility to add near-optimal deviation at an intermediate level as in Theorem 7.3.1, near-optimality robustness can be added at multiple levels of a multilevel model without changing its complexity class in the polynomial hierarchy.

7.5 Conclusion

In this paper, we have shown that for many configurations of bilevel and multilevel optimization problems, adding near-optimality robustness to the canonical problem does not increase its complexity in the polynomial hierarchy. This result is obtained even though near-optimality robustness constraints add another level to the multilevel problem, which in general would change the complexity class. We defined the class $\Sigma_s^{P*}[\mathcal{H}]$ that is general enough to capture many non-linear multilevel problems but avoids corner cases where the modified objective or additional linear constraint makes the problem harder to solve.

Future work will consider specialized solution algorithms for some classes of near-optimal robust bilevel and multilevel problems.

CHAPTER 8 NEW ALGORITHMS FOR NEAR-OPTIMAL ROBUST BILEVEL LINEAR OPTIMIZATION

8.1 Introduction

Chapter 6 introduces the near-optimal robust bilevel problem (NORBiP), defined as a bilevel optimization problem where the upper level protects the feasibility of its constraints against deviations of the lower level solution. In the linear case, an extended single-level closed-form formulation is derived, leading to a problem with complementarity constraints and disjunctive constraints:

$$\min_{x,v} \; c_x^T x + c_y^T v \tag{8.1a}$$

$$\text{s.t. } G_k^T x + H_k^T v \leq q_k \qquad \forall k \in [\![m_u]\!] \tag{8.1b}$$

$$d_j + \sum_i \lambda_i B_{ij} - \sigma_j = 0 \qquad \forall j \in [\![n_l]\!] \tag{8.1c}$$

$$0 \leq \lambda_i \perp b_i - A_i x - B_i v \geq 0 \qquad \forall i \in [\![m_l]\!] \tag{8.1d}$$

$$0 \leq \sigma_j \perp v_j \geq 0 \qquad \forall j \in [\![n_l]\!] \tag{8.1e}$$

$$\bigvee_{l=1}^{|\mathcal{V}_k|} \sum_{i=1}^{m_l} \alpha_{ki}^l (b - Ax)_i + \beta_k^l \left(d^T v + \delta\right) \leq q_k - G_k^T x \qquad \forall k \in [\![m_u]\!] \tag{8.1f}$$

$$x \in \mathcal{X}. \tag{8.1g}$$

Constraint (8.1b) ensures upper-level feasibility, Constraints (8.1d-8.1e) capture the lower-level primal feasibility and complementarity constraints between the primal and dual constraints and variables. They can be reformulated with SOS1 or indicator constraints in a MILP modelling framework. Constraint (8.1f) is the disjunctive constraint of the extended formulation, enforcing one vertex of the dual adversarial polyhedron to respect the upper-level constraint with the specified level of near-optimality robustness, with $\mathcal{V}_k$ the list of extreme vertices of the dual adversarial polyhedron for the k-th upper-level constraint. The length of this list grows exponentially with the size of the lower-level problem (and hence of its dual). In a MILP framework, it can be reformulated using auxiliary binary variables and indicator constraints:

$$\begin{aligned}
&\sum_{l=1}^{|\mathcal{V}_k|} w_{kl} \geq 1 && \forall k \in [\![m_u]\!] \\
&w_{kl} = 1 \Rightarrow \alpha_{ki}^l (b - Ax)_i + \beta_k^l \cdot (d^T v + \delta) \leq q_k - G_k^T x && \forall k \in [\![m_u]\!],\ \forall l \in [\![|\mathcal{V}_k|]\!] \\
&w_{kl} \in \{0, 1\} && \forall k \in [\![m_u]\!],\ \forall l \in [\![|\mathcal{V}_k|]\!]
\end{aligned}$$

Computational experiments highlighted the difficulty of solving bilevel linear problems with added near-optimal robust constraints to global optimality compared to their canonical counterpart. In this chapter, we design new exact and heuristic solution methods for the linear NORBiP and evaluate their efficiency on two sets of instances.

This chapter is organized as follows. Section 8.2 introduces two exact algorithms to solve the extended reformulation of the linear NORBiP, while the algorithm presented in Section 8.3 is a heuristic method guaranteed to compute a near-optimal robust bilevel-feasible solution. In Section 8.4, computational experiments and benchmarks are conducted, assessing the effectiveness and efficiency of the various presented methods. Section 8.5 concludes the chapter and highlights future work.

8.2 Exact lazy subproblem expansion

In this section, we present two novel methods for solving linear NORBiP instances. The linear NORBiP requires each dual-adversarial problem to be feasible with an objective value of less than $q_k - G_k^T x$. For $\mathcal{S} \subseteq [\![m_u]\!]$, we denote by the following problem as $\overline{\text{NORBiP}}(\mathcal{S})$:

$$\begin{aligned}
\min_{x,v}\ & c_x^T x + c_y^T v \\
\text{s.t. } & (8.1\text{b}) - (8.1\text{e}) \\
& G_k^T x + H_k^T v \leq q_k && \forall z \in \mathcal{Z}(x;\delta)\ \forall k \in \mathcal{S} \qquad (8.2) \\
& x \in \mathcal{X}.
\end{aligned}$$

It can be noted that $\overline{\text{NORBiP}}(\emptyset)$ does not have any near-optimality robustness constraint and corresponds to the optimistic bilevel problem. $\overline{\text{NORBiP}}([\![m_u]\!])$ is equivalent to NORBiP and integrates all near-optimality robustness constraints. $\overline{\text{NORBiP}}(\mathcal{S})$ is trivially a relaxation of NORBiP since only a subset of the robust constraints is applied.

Furthermore, for a given primal solution (x, v), verifying its near-optimality robustness with respect to the $k-th$ upper-level constraint can be done with an auxiliary linear optimization problem:

$$\begin{aligned}
\min_{\alpha,\beta} \; & (b - Ax)^T\alpha + (d^Tv + \delta)\beta \\
\text{s.t. } & B^T\alpha + \beta d \geq H_k \\
& \alpha \in \mathbb{R}_+^{m_l} \beta \in \mathbb{R}_+.
\end{aligned} \tag{8.3}$$

We consider a solution to the auxiliary problem as "robust" if the optimal value is below $q_k - G_k^T x$, in which case the constraint is robust to near-optimal deviations of the lower level. Algorithm 8.1 starts from $\mathcal{S} = \emptyset$ and iteratively adds violated constraints from (8.2) that ensure the near-optimality robustness of some upper-level constraints. We thus qualify it as lazy in contrats to the complete formulation from Chapter 6 introducing all disjunctive constraints upfront.

At any given iteration, the set $\overline{\mathcal{S}}$ is the complement of $\mathcal{S}$,

$$\overline{\mathcal{S}} \equiv [\![m_u]\!] \setminus \mathcal{S}$$

and $\mathcal{S}_{opt}$ contains the set of upper-level constraint indices that are robust for a current iterate. If $\overline{\mathcal{S}}$ is empty, there is no upper-level constraint that is not either added (already in $\mathcal{S}$) or already optimal (in $\mathcal{S}_{opt}$). Given that $\overline{NORBiP}$ is a relaxation of NORBiP, $\overline{\mathcal{S}} = \emptyset$ implies that the optimum is reached.

At any given iterate of Algorithm 8.1, $\overline{NORBiP}(\mathcal{S})$, a MILP, is solved from scratch. Resolving the problem at each iterate can create a high overhead in the computing time. Hence, we define a variant, Algorithm 8.2, that adds multiple violated near-optimality robustness constraints at each iteration, hence reducing the number of iterates compared to Algorithm 8.1. It takes a single parameter, $\mathcal{B}$, the maximum number of disjunctions added at each iteration.

The $\mathcal{S}_{opt}$ and $\overline{\mathcal{S}}$ sets are similar to Algorithm 8.1. $\mathcal{I}$ corresponds to the set of upper-level constraint indices that are added in the current batch.

Algorithm 8.1 Lazy Subproblem Expansion

function LazySubProblemExpansion($G, H, q, c_x, c_y, A, B, b, d, \delta$)
  $\overline{NORBiP} \leftarrow$ optimistic model with parameters $G, H, q, c_x, c_y, A, B, b, d$
  if Optimistic infeasible **then**
    Terminate with optimistic infeasible
  end if
  $\overline{\mathcal{S}} \leftarrow [\![m_u]\!]$
  $w_k \leftarrow \{\}\ \forall k \in [\![m_u]\!]$
  while $\overline{\mathcal{S}}$ not empty **do** ▷ subproblem exploration phase
    $\mathcal{S}_{opt} \leftarrow \emptyset$
    $(\hat{x}, \hat{v}) \leftarrow$ current solution
    Choose $k \in \overline{\mathcal{S}}$
    Solve $k-th$ dual subproblem (8.3) parameterized by $(\hat{x}, \hat{v})$
    if Solution is robust **then**
      $\mathcal{S}_{opt} \leftarrow \mathcal{S}_{opt} \cap \{k\}$
      $\overline{\mathcal{S}} \leftarrow \overline{\mathcal{S}} \setminus \{k\}$
    else ▷ subproblem expansion phase
      Add new variables $w_{kl} \in \{0,1\}\ \forall l \in 1 \dots |\mathcal{V}_k|$ to $\overline{NORBiP}$
      Add constraint $\sum_{l=1}^{|\mathcal{V}_k|} w_{kl} \geq 1$ to $\overline{NORBiP}$
      for $l \in 1 \dots |\mathcal{V}_k|$ **do**
        $(\alpha_k^l, \beta_k^l) \leftarrow \mathcal{V}_k^l$
        Add indicator constraint to $\overline{NORBiP}$:
         $w_{kl} = 1 \Rightarrow (b - Ax)^T \alpha_k^l + (d^T v + \delta)\beta_k^l \leq q_k - G_k^T x$
      end for
    end if
    Solve current iterate $\overline{NORBiP}$
    if Infeasible **then**
      Terminate with infeasible $k-$th subproblem
    end if
    $\overline{\mathcal{S}} \leftarrow \overline{\mathcal{S}} \cap \mathcal{S}_{opt}$
  end while
  return $(\hat{x}, \hat{v}, w)$
end function

8.3 Single-vertex heuristic

In this section, we present Algorithm 8.3, a heuristic algorithm to compute high-quality bilevel-feasible and near-optimal robust solutions. It only requires optimizing a finite sequence of MILPs with the same variables as the canonical bilevel problem and few additional linear constraints, instead of the disjunctive constraints with a number of terms equal to the number of vertices of the dual adversarial polyhedron.

At any given iterate, $\mathcal{C}$ is the set of upper-level constraint indices that were already added to

Algorithm 8.2 Lazy Batched Subproblem Expansion

function LazyBatched$(G, H, q, c_x, c_y, A, B, b, d, \delta, \mathcal{B})$
  $\overline{NORBiP} \leftarrow$ optimistic model with parameters $(G, H, q, c_x, c_y, A, B, b, d)$
  if Optimisic infeasible **then**
    Terminate with optimistic infeasible
  end if
  $\overline{\mathcal{S}} \leftarrow [\![m_u]\!]$
  $w_k \leftarrow \{\} \forall k \in [\![m_u]\!]$
  while $\overline{\mathcal{S}}$ not empty **do**
    $\mathcal{I} \leftarrow \emptyset$
    while $|\mathcal{I}| < \mathcal{B} \,\&\, \overline{\mathcal{S}}$ not empty **do**
      $\mathcal{S}_{opt} \leftarrow \emptyset$
      $(\hat{x}, \hat{v}) \leftarrow$ optimal solution of $\overline{NORBiP}$
      Choose $k \in \overline{\mathcal{S}}$
      Solve k-th dual subproblem parameterized by $(\hat{x}, \hat{v})$
      $\overline{\mathcal{S}} \leftarrow \overline{\mathcal{S}} \setminus \{k\}$
      if Solution is robust **then**
        $\mathcal{S}_{opt} \leftarrow \mathcal{S}_{opt} \cap \{k\}$
      else
        $\mathcal{I} \leftarrow \mathcal{I} \cup \{k\}$
        Add new variables $w_{kl} \in \{0, 1\} \; \forall l \in 1 \ldots |\mathcal{V}_k|$ to $\overline{NORBiP}$
        Add constraint $\sum_{l=1}^{|\mathcal{V}_k|} w_{kl} \geq 1$ to $\overline{NORBiP}$
        for $l \in 1 \ldots |\mathcal{V}_k|$ **do**
          $(\alpha_k^l, \beta_k^l) \leftarrow \mathcal{V}_k^l$
          Add indicator constraint to $\overline{NORBiP}$:
           $w_{kl} = 1 \Rightarrow (b - Ax)^T \alpha_k^l + (d^T v + \delta)\beta_k^l \leq q_k - G_k^T x$
        end for
      end if
    end while
    Solve current iterate $\overline{NORBiP}$
    if Infeasible **then**
      Terminate with infeasible subproblems $\mathcal{I}$
    end if
    $\overline{\mathcal{S}} \leftarrow \overline{\mathcal{S}} \cap \mathcal{S}_{opt}$
  end while
  return $(\hat{x}, \hat{v}, w)$
end function

the model. New constraints are added in a batched fashion, with the batch size controlled by the η parameter. Each k-th subproblem is added only once, by selecting a single vertex (α, β) and using it to enforce the constraint

$$(b - Ax)^T \alpha + (d^T v + \delta)\beta \leq q_k - G_k^T x.$$

Algorithm 8.3 Single-Vertex Heuristic

1: **function** SINGLEVERTEXHEURISTIC($G, H, q, c_x, c_y, A, B, b, d, \delta, \eta$)
2: $P \leftarrow$ optimistic model with parameters $(G, H, q, c_x, c_y, A, B, b, d)$
3: **if** Optimistic infeasible **then**
4: **Terminate** with optimistic infeasible
5: **end if**
6: **if** $\exists k$, k-th dual adversarial is infeasible **then**
7: **Terminate** with k-th dual adversarial infeasible
8: **end if**
9: count $\leftarrow 1$
10: $\mathcal{C} \leftarrow \emptyset$
11: **while** count > 0 **do** ▷ Outer loop
12: $k \leftarrow 1$
13: count $\leftarrow 0$
14: $(\hat{x}, \hat{v}) \leftarrow$ current solution of P
15: **while** count $\leq \eta$ and $k \leq m_u$ **do**
16: $(\alpha, \beta) \leftarrow$ solution to k-th dual adversarial (8.3) parameterized by $(\hat{x}, \hat{v})$
17: **if** $(b - A\hat{x})^T\alpha + (d^T\hat{v} + \delta)\beta > q_k - G_k^T\hat{x}$ **then**
18: $\mathcal{C} \leftarrow \mathcal{C} \cup \{k\}$
19: Add constraint to P: $(b - Ax)^T\alpha + (d^Tv + \delta)\beta \leq q_k - G_k^T x$
20: count $\leftarrow$ count $+ 1$
21: **while** $k \in \mathcal{C}$ **do**
22: $k \leftarrow k + 1$
23: **end while**
24: **end if**
25: **end while**
26: Re-optimize P
27: **if** Current iterate infeasible **then**
28: **Terminate** with no found solution
29: **end if**
30: **end while**
31: **return** $(\hat{x}, \hat{v})$
32: **end function**

Unlike the exact algorithms, Algorithm 8.3 initializes a MILP model and iteratively adds linear constraints to it. This procedure therefore lends itself to warm starts.

Proposition 8.3.1. *Algorithm 8.3 terminates in at most m_u iterations of the outer loop beginning Line 11 and solves optimization problems with the same variables as the canonical optimization problem and at most m_u linear constraints.*

Proof. Algorithm 8.3 maintains a cache $\mathcal{C}$ of the subproblems that have been explored. For each subproblem, exactly one vertex is chosen, which minimizes

$$(b - A\hat{x})^T\alpha + (d^T\hat{v} + \delta)\beta$$

with $(\hat{x}, \hat{v})$ the current iterate. For a chosen $\hat{x}$, the lower-level problem is feasible since $\hat{v}$ is computed, so the dual problem cannot be unbounded. It cannot be infeasible since its feasibility domain depends only on the problem data and is verified Line 6. Therefore, a vertex $(\alpha, \beta)_k$ is computed. If the objective is greater than $q_k - G_k^T\hat{x}$, then the current iterate is not near-optimal robust with respect to the k-th constraint, and the linear constraint:

$$(b - Ax)^T\alpha + (d^T v + \delta)\beta \leq q_k - G_k^T x$$

is added to the model. If no addition is made, the current iterate is near-optimal robust, the count variable remains at 0, and the outer loop exits, with the function returning the current iterate. Otherwise, the iterate was not near-optimal robust with respect to at least one upper-level constraint, which is turned into a constraint and added to the cache. The outer loop adds at least one constraint in a non-terminating iteration, therefore, m_u iterations suffice to add all constraints. Moreover, each subproblem adds exactly one linear constraint, so at most m_u linear constraints are added to the formulation of the canonical problem. □

Algorithm 8.3 verifies near-optimality of the upper-level constraints in a lexicographic order, with k going from 1 to m_u. This order can lead to over-verifying the same subproblems an excessive number of times. The procedure can be improved by randomizing the order of verification of near-optimal constraints with a permutation of $\{1..m_u\}\backslash\mathcal{C}$.

The η parameter controls the maximum number of linear constraints added for each outer iteration. $\eta = 1$ implies that the algorithm reoptimizes the problem after adding a single constraint, while $\eta \geq m_u$ will add all the linear constraints that correspond to upper-level constraints that are not near-robust at each iterate.

Finally, it can be noted that the single-vertex algorithm can be applied even when the number of vertices in the dual-adversarial is infinite, i.e. when the lower-level problem is a convex optimization problem. The only modification is the optimization of the dual adversarial problem for fixed values of (x, v) at Line 16, where a convex optimization problem is solved instead of a linear one.

8.4 Computational experiments

In this section, we present the implementation of the algorithms from Section 8.2 and Section 8.3 and computational experiments. The various algorithms are implemented in Julia v1.5 [104] using JuMP as a modelling framework [107, 108]. The implementation of the described algorithms is available and archived at [151]. SCIP v6.0 [138] is used as the MILP solver with SoPlex as the underlying linear solver. All experiments are run on the same hardware as Chapter 6. The time-out limit is set at two hours for all algorithms.

We generate instances from the 50 MIPS/RANDOM instances of the Bilevel Optimization Problem Library [135]. Since the original problems do not have upper-level constraints, we create new instances by using a third of the constraints at the upper level. In another group of instances, we use two thirds of the constraints at the upper level. We will refer to the first group of instances as MIPS instances and to the second group as alternative MIPS or ALTMIPS. The problem dimensions for the two groups are detailed below:

$$(m_u, m_l, n_l, n_u) = (6, 14, 10, 5) \qquad \text{(MIPS)}$$
$$(m_u, m_l, n_l, n_u) = (14, 6, 10, 5) \qquad \text{(ALTMIPS)}$$

The instances are archived and available on [136] in the JLD format. A δ value of 0.1 is applied. A comparison of runtimes between the exact method from Chapter 6 and the lazy expansion algorithm is presented in Fig. 8.1.

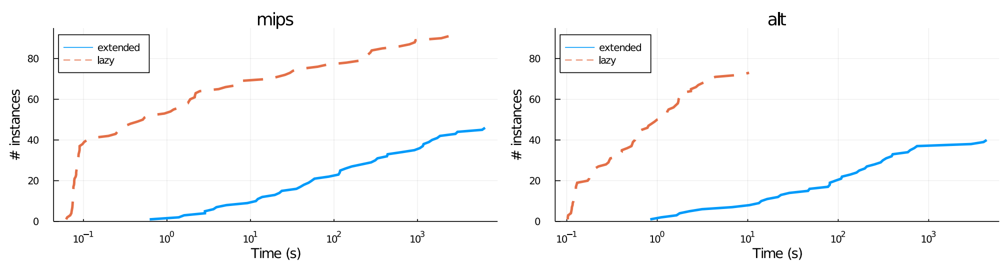

Figure 8.1 Comparison of exact methods

The lazy algorithm successfully terminates for all instances on both sets while the extended method results in a time-out for 38 instances of the ALTMIPS set and 45 instances of the MIPS set. The profiles highlight the efficiency of the lazy method for all instances. In Fig. 8.2, we compare the lazy simple and batched algorithm with various batch sizes.

The standard lazy algorithm slightly outperforms the batched versions for all batch sizes for

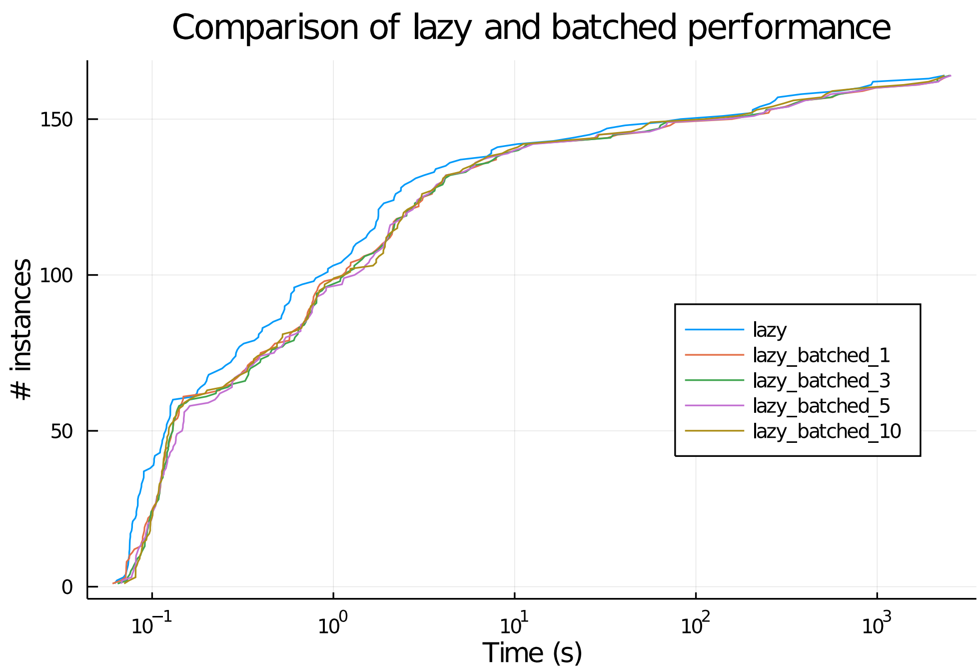

Figure 8.2 Comparison of simple batched lazy algorithms

instances solved below 10 seconds. For instances that take longer to optimize, the difference in runtime between the two methods is not significant enough on the set.

Runtime comparisons of the lazy algorithm and single-vertex heuristic are presented in Fig. 8.3 on the merged ALT and MIPS instances. All heuristic methods terminate much faster than the lazy exact method. Moreover, the heuristic always find a feasible solution for all ALT and MIPS instances. The solution is found to be optimal for all MIPS instances and all but two ALT instances, confirming that the solutions found by the heuristic are of high quality.

The runtimes of the standard and randomized single-vertex heuristics have similar distributions and cannot lead to any conclusion on the superiority of one method over the other.

One conclusion from the experiments in Chapter 6 was that solving the extended formulation for a near-optimal robust bilevel problem was much more computationally demanding than solving the optimistic version. In Fig. 8.4, we compare the deterministic and randomized single-vertex heuristics with the optimistic formulation.

The heuristic solves the MILP problem corresponding to the optimistic formulation and adds linear constraints based on the solution of linear subproblems. The additional time solving

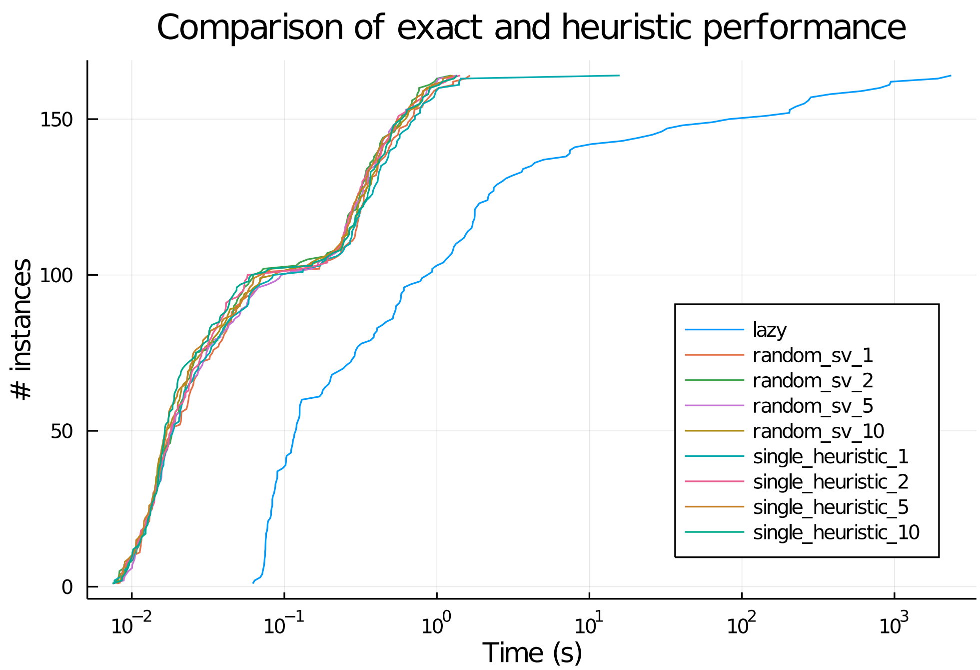

Figure 8.3 Comparison of exact and heuristic methods

the linear subproblems and the multiple iterations can explain the greater runtime.

8.5 Conclusion

In this chapter, new exact and heuristic methods were developed to accelerate the resolution of the extended formulation of the linear near-optimal robust bilevel optimization problem. The first group of algorithms is built on the relaxation of near-optimality robustness for a subset of the upper-level constraints, leading to an iterative procedure terminating on a near-optimal robust point. The single-vertex heuristic algorithm provides a feasible point when it terminates by solving a MILP with the same constraints as the canonical bilevel problem and few additional linear constraints. As a heuristic method, it can also provide good initial solutions to start branch-and-bound procedures.

Future research will consider the development of algorithms for other classes of near-optimal robust bilevel problems, such as problems where the lower level includes integrality constraints and where the lower level is convex and non-linear.

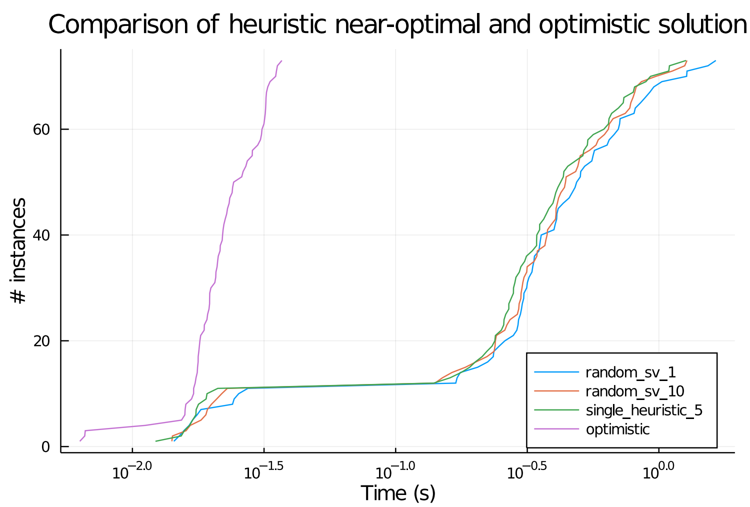

Figure 8.4 Comparison of heuristic and optimistic formulation

CHAPTER 9 CONCLUSION AND RECOMMENDATIONS

This thesis defines new problems in bilevel optimization and pricing for demand response, opening perspective for real-world applications in energy, improved modelling techniques and solution methods.

9.1 Thesis summary

The first part of this thesis explores bilevel models for price-setting problems in a power system where a supplier anticipates the decisions of users. In the context of power systems subject to increased stress due to distributed renewable energy production, these novel mechanisms offer additional levers for flexibility to the supplier, while letting customers reduce their cost through smart planning.

Using bilevel optimization, the logic of users is approached as a complete optimization problem and lets the supplier regard them as strategic players. Chapter 5 and Appendix A highlight that the power supplier can accommodate various constraints by adjusting the pricing components of TLOU, thus nudging flexible users into changing their consumption patterns. The pricing problem also presents the noticeable property that the baseline with a zero capacity can be perceived as safer and simpler, hence preferable, by users. This property challenges the optimistic assumption commonly used to ensure well-defined solutions in bilevel optimization. That challenge was overcome in our model by a concept of user conservativeness which is not a purely computational artifact used to regularize the problem but finds a natural economic interpretation of the user behaviour, or of its perception by the supplier.

The second part of the thesis focuses on bilevel problems with near-optimality robustness constraints (NORBiP). Chapter 6 introduces the model, generalizing the concept of conservativeness with respect to user behaviour beyond the TLOU pricing application. The fundamental concept of bilevel optimization is that a second optimization problem is solved to global optimality after the upper-level variables are fixed. NORBiP challenges this assumption and anticipates potential deviations from optimality of this second level, resulting in a more conservative model protecting feasibility in more cases through the lenses of robust optimization. The nature of the near-optimal set of the lower level integrates generalized semi-infinite constraints in the formulation of NORBiP, corresponding to a robust optimization problem with decision-dependent uncertainty. Focusing on the convex and linear cases, we exploit duality of the adversarial problem to derive a closed-form, single-level reformulation containing the same form of constraints as the single-level reformulation of the non-robust bilevel problem. In Chapter 7, complexity results are obtained for the near-optimal robust version of some multilevel problems. Under mild assumptions, near-optimality robustness does not change the complexity class of a given multilevel problem in the polynomial hierarchy. Chapter 8 builds upon the extended formulation defined in Chapter 6 for the linear-linear NORBiP with two groups of algorithms. The first group is composed of two exact algorithms expanding the dual adversarial subproblems on the fly only for constraints that are not near-optimal robust at a given iterate. The second group consists of two heuristics that compute a high-quality bilevel-feasible near-optimal robust solution. These two groups of algorithms significantly improve the solution time of NORBiP compared to the extended formulation.

9.2 Limitations

The TLOU framework provides power suppliers with an extension of price-based demand response programs. The work presented in this thesis focuses on the optimal setting of the pricing parameters which is only one of the aspects of the problem faced by the supplier. The use of TLOU has more impact when integrated into a complete model of the supplier decision with power purchase from a wholesale market and/or unit commitment constraints.

The proposed model relies on the consumption distribution in two ways:

- the distribution is assumed to be perfectly known by both the supplier and the user;
- the distribution is assumed to be fixed and independent from the decisions.

In a privacy-preserving environment, the supplier would not have access to the exact time-dependent distribution but to a data-based estimate. Methods from the stochastic and robust literature, such as distributionally robust optimization could provide a richer framework for modelling the additional uncertainty. Changes in the distribution can occur with user decisions, for example, if they operate a storage system.

For the multi-user TLOU formulation, numerical experiments highlighted that the pricing problems become computationally challenging for approximately 100 users for all variants with the current formulation. Even though such problem size is sufficient for a micro-grid or small neighbourhood, it implies that the formulation cannot be used for a supplier offering TLOU prices to even medium-size urban areas.

Near-optimality robustness brings more flexibility to bilevel optimization by relaxing the hypothesis of pure rationality of the lower level to bounded rationality. In the linear-linear case, even though we proposed a MILP-based reformulation, its size grows exponentially with the dimension and number of constraints of the lower-level problem. The exact algorithms developed in Chapter 8 reduce the size of the MILPs effectively solved. Those nonetheless remain of exponential size as a function of the lower-level problem size. The only method resulting in a polynomial problem size is the single-vertex heuristic, for which no guarantee on the quality of the solution could be provided despite its effectiveness in practice.

9.3 Future research

The bilevel formulation of TLOU offers a method for optimally choosing the pricing parameters to influence users. The next phase of the development of the pricing system would be its integration within the broader problem faced by a power supplier or aggregator. Another aspect to develop would be a more complete model of the user decision in the presence of flexible loads and storage capacity. In their presence, the user decision does not consist only in choosing a capacity to book but also scheduling flexible loads. The resulting two-stage optimization problem of the user creates new modelling and computational challenges, since the necessary optimality conditions leveraged in Chapter 4 and Chapter 5 do not apply with more lower-level variables.

The multi-user TLOU problem from Chapter 5 becomes computationally challenging when the number of users grows. Given the structure of the problem, decomposition techniques could improve the tightness of the linear relaxation and accelerate the branch-and-bound procedure. Another potential for acceleration could come from designing heuristics for rapidly computing initial solutions and prune nodes of the branch-and-bound tree.

In Chapter 6, two groups of valid inequalities were studied for NORBiP. One group did not improve the solution time for instances solved to optimality but helped to detect infeasible instances at the root node without solving the bilevel problem. This group of inequalities was initially defined for canonical bilevel problems and not specifically for their near-optimal robust counterpart. Valid inequalities specifically exploiting the primal-dual nature of the variables added by near-optimality robustness could strengthen the formulation and possibly lead to a faster resolution.

The near-optimality robustness of each upper-level constraint is independent of the feasibility or robustness of the other upper-level constraints. This fact is leveraged in Chapter 8 to design faster exact algorithms. This independence could also be exploited to derive less conservative forms of robustness, such as Γ-robustness, in the context of lower-level near-optimality.

As a last direction for future research, one can note that the single-level reformulation of NORBiP does not modify the structure of the KKT-based reformulation of the canonical bilevel problem, but adds new variables and constraints to it. The development of solution methods for NORBiP could therefore generalize to problems including two groups of variables, with each set of variables constrained in their own domain and linked only by a bilinear inequality. The extended method proposed in Chapter 6 was shown to solve NORBiP instances more effectively than the bilinear formulation relying on spatial branching despite the exponential size and indicator constraints; it could therefore lead to novel approaches for solving more generic linear-bilinear optimization problems.

REFERENCES

[1] Y. Parag and B. K. Sovacool, "Electricity market design for the prosumer era," *Nature Energy*, vol. 1, no. 4, pp. 1–6, 2016.

[2] E. Kuznetsova and M. F. Anjos, "Challenges in energy policies for the economic integration of prosumers in electric energy systems: A critical survey with a focus on Ontario (Canada)," *Energy Policy*, vol. 142, 2020, https://doi.org/10.1016/j.enpol.2020.111429.

[3] H. F. Stackelberg, *Marktform und Gleichgewicht*, 1934.

[4] H. P. Moreton and C. H. Séquin, "Functional optimization for fair surface design," *ACM SIGGRAPH Computer Graphics*, vol. 26, no. 2, pp. 167–176, 1992.

[5] D. Butnariu and A. N. Iusem, *Totally convex functions for fixed points computation and infinite dimensional optimization.* Springer Science & Business Media, 2000, vol. 40.

[6] B. Legat, "Set programming: Theory and computation," 2020, https://dial.uclouvain.be/pr/boreal/object/boreal:237650.

[7] J. Sokolowski and J.-P. Zolésio, "Introduction to shape optimization," in *Introduction to shape optimization.* Springer, 1992, pp. 5–12.

[8] H. Moss, D. Leslie, D. Beck, J. Gonzalez, and P. Rayson, "Boss: Bayesian optimization over string spaces," *Advances in neural information processing systems*, vol. 33, 2020.

[9] T. Achterberg, "Constraint integer programming," 2007. [Online]. Available: https://depositonce.tu-berlin.de/handle/11303/1931

[10] J. Nocedal and S. Wright, *Numerical optimization.* Springer Science & Business Media, 2006.

[11] P. Palensky and D. Dietrich, "Demand side management: Demand response, intelligent energy systems, and smart loads," *IEEE Transactions on Industrial Informatics*, vol. 7, no. 3, pp. 381–388, 2011.

[12] M. H. Albadi and E. F. El-Saadany, "A summary of demand response in electricity markets," *Electric Power Systems Research*, vol. 78, no. 11, pp. 1989–1996, 2008.

[13] N. G. Paterakis, O. Erdinç, and J. P. Catalão, “An overview of Demand Response: Key-elements and international experience,” *Renewable and Sustainable Energy Reviews*, vol. 69, no. July 2016, pp. 871–891, 2017. [Online]. Available: http://dx.doi.org/10.1016/j.rser.2016.11.167

[14] N. Good, K. A. Ellis, and P. Mancarella, “Review and classification of barriers and enablers of demand response in the smart grid,” *Renewable and Sustainable Energy Reviews*, vol. 72, pp. 57–72, 2017.

[15] L. Gelazanskas and K. A. Gamage, “Demand side management in smart grid: A review and proposals for future direction,” *Sustainable Cities and Society*, vol. 11, pp. 22–30, 2014.

[16] L. Gkatzikis, I. Koutsopoulos, and T. Salonidis, “The role of aggregators in smart grid demand response markets,” *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 7, pp. 1247–1257, 2013.

[17] F. Babonneau, R. Foguen, A. Haurie, and R. Malhamé, “Coupling a power dispatch model with a wardrop or mean-field-game equilibrium model,” *Dynamic Games and Applications*, 2020.

[18] R. F. Tchuendom, R. Malhamé, and P. Caines, “A quantilized mean field game approach to energy pricing with application to fleets of plug-in electric vehicles,” in *2019 IEEE 58th Conference on Decision and Control (CDC)*. IEEE, 2019, pp. 299–304.

[19] J. Vuelvas, F. Ruiz, and G. Gruosso, “Limiting gaming opportunities on incentive-based demand response programs,” *Applied Energy*, vol. 225, pp. 668–681, 2018.

[20] J. Vuelvas and F. Ruiz, “A novel incentive-based demand response model for cournot competition in electricity markets,” *Energy Systems*, Jan 2018. [Online]. Available: https://doi.org/10.1007/s12667-018-0271-2

[21] A. Asadinejad, A. Rahimpour, K. Tomsovic, H. Qi, and C.-f. Chen, “Evaluation of residential customer elasticity for incentive based demand response programs,” *Electric Power Systems Research*, vol. 158, pp. 26–36, 2018. [Online]. Available: http://dx.doi.org/10.1016/j.epsr.2017.12.017

[22] G. Aghajani, H. Shayanfar, and H. Shayeghi, “Demand side management in a smart micro-grid in the presence of renewable generation and demand response,” *Energy*, vol. 126, pp. 622–637, 2017.

[23] G. Aghajani and N. Ghadimi, “Multi-objective energy management in a micro-grid,” *Energy Reports*, vol. 4, pp. 218–225, 2018.

[24] G. Yuan, Y. Gao, B. Ye, and R. Huang, “Real-time pricing for smart grid with multi-energy microgrids and uncertain loads: a bilevel programming method,” *International Journal of Electrical Power & Energy Systems*, vol. 123, p. 106206, 2020.

[25] Y. Mou, A. Papavasiliou, and P. Chevalier, “A bi-level optimization formulation of priority service pricing,” *IEEE Transactions on Power Systems*, pp. 1–1, 2019.

[26] Y. Mou, “Nonlinear pricing schemes for mobilizing residential flexibility in power systems,” Ph.D. dissertation, Université Catholique de Louvain, 2020.

[27] J. A. Gomez-Herrera and M. F. Anjos, “Optimization-based estimation of power capacity profiles for activity-based residential loads,” *International Journal of Electrical Power & Energy Systems*, vol. 104, pp. 664–672, 2019. [Online]. Available: https://doi.org/10.1016/j.ijepes.2018.07.023

[28] M. F. Anjos, L. Brotcorne, and J. A. Gomez-Herrera, “Optimal time-and-level-of-use price setting for an energy retailer,” 2019, http://www.optimization-online.org/DB_FILE/2019/11/7474.pdf.

[29] J. A. Gomez Herrera, “User-oriented demand response for smart buildings,” Ph.D. dissertation, École Polytechnique de Montréal, 2017. [Online]. Available: https://publications.polymtl.ca/2663/

[30] J. L. R. Duarte, N. Fan, and T. Jin, “Multi-process production scheduling with variable renewable integration and demand response,” *European Journal of Operational Research*, vol. 281, no. 1, pp. 186–200, 2020.

[31] M. D. de Souza Dutra and N. Alguacil, “Optimal residential users coordination via demand response: An exact distributed framework,” *Applied Energy*, vol. 279, p. 115851, 2020.

[32] I. Soares, M. J. Alves, and C. H. Antunes, “A deterministic bounding procedure for the global optimization of a bi-level mixed-integer problem,” *European Journal of Operational Research*, 2020. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0377221720308018

[33] L. von Niederhäusen, “Design and pricing of new energy services in a competitive environment,” Ph.D. dissertation, Centrale Lille, 2019.

[34] S. Afşar, L. Brotcorne, P. Marcotte, and G. Savard, "Achieving an optimal trade-off between revenue and energy peak within a smart grid environment," *Renewable Energy*, vol. 91, pp. 293–301, 2016.

[35] S. Afsar, "Revenue optimization and demand response models using bilevel programming in smart grid systems," Ph.D. dissertation, Lille 1, 2016.

[36] J. Bracken and J. T. McGill, "Mathematical programs with optimization problems in the constraints," *Operations Research*, vol. 21, no. 1, pp. 37–44, 1973.

[37] S. Dempe, *Foundations of bilevel programming*. Springer Science & Business Media, 2002.

[38] M. H. Glass, "Bilevel optimization for parameter estimation in thermodynamics," Ph.D. dissertation, Universitätsbibliothek der RWTH Aachen, 2018.

[39] M. Glass and A. Mitsos, "Parameter estimation in reactive systems subject to sufficient criteria for thermodynamic stability," *Chemical Engineering Science*, vol. 197, pp. 420–431, 2019.

[40] M. G. Nicholls, "Optimizing the operations of an aluminium smelter using non-linear bi-level programming," in *Multilevel Optimization: Algorithms and Applications*. Springer, 1998, pp. 117–148.

[41] G. E. Stavroulakis and H. Günzel, "Optimal structural design in nonsmooth mechanics," in *Multilevel Optimization: Algorithms and Applications*. Springer, 1998, pp. 91–115.

[42] S. Albrecht, M. Leibold, and M. Ulbrich, "A bilevel optimization approach to obtain optimal cost functions for human arm movements," *Numerical Algebra, Control & Optimization*, vol. 2, no. 1, p. 105, 2012.

[43] J. Herskovits, A. Leontiev, G. Dias, and G. Santos, "Contact shape optimization: a bilevel programming approach," *Structural and Multidisciplinary Optimization*, vol. 20, no. 3, pp. 214–221, 2000.

[44] C. Audet, P. Hansen, B. Jaumard, and G. Savard, "Links between linear bilevel and mixed 0–1 programming problems," *Journal of Optimization Theory and Applications*, vol. 93, no. 2, pp. 273–300, 1997.

[45] L. J. Stockmeyer, "The polynomial-time hierarchy," *Theoretical Computer Science*, vol. 3, no. 1, pp. 1–22, 1976.

[46] R. G. Jeroslow, "The polynomial hierarchy and a simple model for competitive analysis," *Mathematical Programming*, vol. 32, no. 2, pp. 146–164, 1985.

[47] S. Dempe, V. Kalashnikov, G. A. Prez-Valds, and N. Kalashnykova, *Bilevel Programming Problems: Theory, Algorithms and Applications to Energy Networks*. Springer Publishing Company, Incorporated, 2015.

[48] W. Wiesemann, A. Tsoukalas, P.-M. Kleniati, and B. Rustem, "Pessimistic Bilevel Optimization," *SIAM Journal on Optimization*, vol. 23, no. 1, pp. 353–380, 2013.

[49] M. J. Alves and C. H. Antunes, "A semivectorial bilevel programming approach to optimize electricity dynamic time-of-use retail pricing," *Computers and Operations Research*, vol. 92, pp. 130–144, 2018. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0305054817303143

[50] D. Aussel, L. Brotcorne, S. Lepaul, and L. von Niederhäusern, "A trilevel model for best response in energy demand-side management," *European Journal of Operational Research*, vol. 281, no. 2, pp. 299–315, 2020.

[51] I. Tseveendorj, *Mathematical Programs with Equilibrium Constraints: A Brief Survey of Methods and Optimality Conditions*. New York, NY: Springer New York, 2013, pp. 49–61. [Online]. Available: https://doi.org/10.1007/978-1-4614-5131-0_4

[52] Y. Kim, S. Leyffer, and T. Munson, "MPEC methods for bilevel optimization problems," pp. 335–360, 2020.

[53] S. Dempe and J. Dutta, "Is bilevel programming a special case of a mathematical program with complementarity constraints?" *Mathematical programming*, vol. 131, no. 1-2, pp. 37–48, 2012.

[54] M. Benko and P. Mehlitz, "On implicit variables in optimization theory," *arXiv preprint arXiv:2008.08677*, 2020.

[55] L. Lozano and J. C. Smith, "A value-function-based exact approach for the bilevel mixed-integer programming problem," *Operations Research*, vol. 65, no. 3, pp. 768–786, 2017.

[56] S. Dempe and A. B. Zemkoho, "The bilevel programming problem: reformulations, constraint qualifications and optimality conditions," *Mathematical Programming*, vol. 138, no. 1-2, pp. 447–473, 2013.

[57] A. Zemkoho and S. Zhou, "Theoretical and numerical comparison of the Karush-Kuhn-Tucker and value function reformulations in bilevel optimization," *arXiv preprint arXiv:2004.10830*, 2020.

[58] F. Furini, I. Ljubić, E. Malaguti, and P. Paronuzzi, "On integer and bilevel formulations for the k-vertex cut problem," *Mathematical Programming Computation*, vol. 12, no. 2, pp. 133–164, 2020.

[59] A. G. Mersha and S. Dempe, "Linear bilevel programming with upper level constraints depending on the lower level solution," *Applied Mathematics and Computation*, vol. 180, no. 1, pp. 247–254, 2006.

[60] S. Pineda and J. M. Morales, "Solving Linear Bilevel Problems Using Big-Ms: Not All That Glitters Is Gold," *IEEE Transactions on Power Systems*, vol. 34, no. 3, pp. 2469–2471, May 2019, https://doi.org/10.1109/TPWRS.2019.2892607.

[61] T. Kleinert, M. Labbé, F. Plein, and M. Schmidt, "There's No Free Lunch: On the Hardness of Choosing a Correct Big-M in Bilevel Optimization," 2019, http://www.optimization-online.org/DB_FILE/2019/04/7172.pdf.

[62] S. Scholtes, "Convergence properties of a regularization scheme for mathematical programs with complementarity constraints," *SIAM Journal on Optimization*, vol. 11, no. 4, pp. 918–936, 2001.

[63] T. Hoheisel, C. Kanzow, and A. Schwartz, "Theoretical and numerical comparison of relaxation methods for mathematical programs with complementarity constraints," *Mathematical Programming*, vol. 137, no. 1-2, pp. 257–288, 2013.

[64] L. Schewe and M. Schmidt, "Computing feasible points for binary MINLPs with MPECs," *Mathematical Programming Computation*, vol. 11, no. 1, pp. 95–118, Mar 2019.

[65] T. Kleinert and M. Schmidt, "Computing feasible points of bilevel problems with a penalty alternating direction method," *INFORMS Journal on Computing*, 2019.

[66] E. M. L. Beale and J. A. Tomlin, "Special facilities in a general mathematical programming system for non-convex problems using ordered sets of variables," *OR*, vol. 69, no. 447-454, p. 99, 1970.

[67] T. Kleinert and M. Schmidt, "Why there is no need to use a big-M in linear bilevel optimization: a computational study of two ready-to-use approaches," 2020, http://www.optimization-online.org/DB_FILE/2020/10/8065.pdf.

[68] M. Fischetti, I. Ljubić, M. Monaci, and M. Sinnl, "Intersection cuts for bilevel optimization," in *International Conference on Integer Programming and Combinatorial Optimization*. Springer, 2016, pp. 77–88.

[69] M. Fischetti, I. Ljubić, M. Monaci, and M. Sinnl, "A New General-Purpose Algorithm for Mixed-Integer Bilevel Linear Programs," *Operations Research*, vol. 65, no. 6, pp. 1615–1637, 2017. [Online]. Available: https://doi.org/10.1287/opre.2017.1650

[70] L. Wang and P. Xu, "The watermelon algorithm for the bilevel integer linear programming problem," *SIAM Journal on Optimization*, vol. 27, no. 3, pp. 1403–1430, 2017.

[71] S. DeNegre, *Interdiction and discrete bilevel linear programming*. Lehigh University PhD, 2011.

[72] S. Tahernejad, T. K. Ralphs, and S. T. DeNegre, "A branch-and-cut algorithm for mixed integer bilevel linear optimization problems and its implementation," *Mathematical Programming Computation*, pp. 1–40, 2020.

[73] S. Liu, M. Wang, N. Kong, and X. Hu, "An enhanced branch-and-bound algorithm for bilevel integer linear programming," *European Journal of Operational Research*, 2020.

[74] X. Shi, O. Prokopyev, and T. Ralphs, "Mixed integer bilevel optimization with k-optimal follower: A hierarchy of bounds," 2020, http://www.optimization-online.org/DB_HTML/2020/06/7874.html.

[75] D. Bertsimas and M. Sim, "The price of robustness," *Operations Research*, vol. 52, no. 1, pp. 35–53, 2004.

[76] D. Bertsimas, D. B. Brown, and C. Caramanis, "Theory and Applications of Robust Optimization," *SIAM Review*, vol. 53, no. 3, pp. 464–501, 2011.

[77] V. Goel and I. E. Grossmann, "A class of stochastic programs with decision dependent uncertainty," *Mathematical Programming*, vol. 108, no. 2, pp. 355–394, Sep 2006.

[78] G. Still, "Generalized semi-infinite programming: Theory and methods," *European Journal of Operational Research*, vol. 119, no. 2, pp. 301–313, 1999.

[79] M. H. Zare, O. A. Prokopyev, and D. Sauré, "On bilevel optimization with inexact follower," *Decision Analysis*, vol. 17, no. 1, pp. 74–95, 2020.

[80] M. H. Zare, O. Y. Özaltın, and O. A. Prokopyev, "On a class of bilevel linear mixed-integer programs in adversarial settings," *Journal of Global Optimization*, vol. 71, no. 1, pp. 91–113, 2018.

[81] A. Aboussoror and P. Loridan, "Strong-weak Stackelberg problems in finite dimensional spaces," *Serdica Mathematical Journal*, vol. 21, no. 2, pp. 151–170, 1995.

[82] Y. Xie and U. V. Shanbhag, "On robust solutions to uncertain linear complementarity problems and their variants," *SIAM Journal on Optimization*, vol. 26, no. 4, pp. 2120–2159, 2016.

[83] S. Dempe, S. Ivanov, and A. Naumov, "Reduction of the bilevel stochastic optimization problem with quantile objective function to a mixed-integer problem," *Applied Stochastic Models in Business and Industry*, vol. 33, no. 5, pp. 544–554, 2017.

[84] S. V. Ivanov, "A bilevel stochastic programming problem with random parameters in the follower's objective function," *Journal of Applied and Industrial Mathematics*, vol. 12, no. 4, pp. 658–667, 2018.

[85] J. Burtscheidt, M. Claus, and S. Dempe, "Risk-averse models in bilevel stochastic linear programming," *SIAM Journal on Optimization*, vol. 30, no. 1, pp. 377–406, 2020.

[86] J. Burtscheidt and M. Claus, "Bilevel Linear Optimization Under Uncertainty," in *Bilevel Optimization*. Springer, 2020, pp. 485–511.

[87] S. Bolusani, S. Coniglio, T. K. Ralphs, and S. Tahernejad, *A Unified Framework for Multistage Mixed Integer Linear Optimization*. Cham: Springer International Publishing, 2020, pp. 513–560. [Online]. Available: https://doi.org/10.1007/978-3-030-52119-6_18

[88] I. Yanıkoğlu and D. Kuhn, "Decision rule bounds for two-stage stochastic bilevel programs," *SIAM Journal on Optimization*, vol. 28, no. 1, pp. 198–222, 2018.

[89] D. Salas and A. Svensson, "Well-posedness of deterministic bilevel games through a general stochastic approach," 2020, https://arxiv.org/abs/2010.05368.

[90] M. Besançon, M. F. Anjos, L. Brotcorne, and J. A. Gomez-Herrera, "A bilevel approach for optimal price-setting of time-and-level-of-use tariffs," *IEEE Transactions on Smart Grid*, vol. 11, no. 6, pp. 5462–5465, 2020, ©2020 IEEE. Reprinted, with permission, from the authors.

[91] H. A. Simon, "Theories of bounded rationality," *Decision and Organization*, vol. 1, no. 1, pp. 161–176, 1972.

[92] R. J. Aumann, "Rationality and bounded rationality," in *Cooperation: Game-Theoretic Approaches*. Springer, 1997, pp. 219–231.

[93] V. Grimm, G. Orlinskaya, L. Schewe, M. Schmidt, and G. Zöttl, “Optimal design of retailer-prosumer electricity tariffs using bilevel optimization,” *Omega*, p. 102327, 2020.

[94] M. Besançon, M. F. Anjos, L. Brotcorne, and J. A. Gomez-Herrera, “A bilevel approach to optimal price-setting of time-and-level-of-use tariffs,” https://arxiv.org/abs/1809.00512, 2018.

[95] US Federal Energy Regulatory Commission (FERC), “Assessment of Demand Response and Advanced Metering, https://www.ferc.gov/legal/staff-reports/12-20-12-demand-response.pdf,” 2012.

[96] L. S. Vargas, G. Bustos-Turu, and F. Larraín, “Wind power curtailment and energy storage in transmission congestion management considering power plants ramp rates,” *IEEE Transactions on Power Systems*, vol. 30, no. 5, pp. 2498–2506, 2014.

[97] A. Sinha, P. Malo, and K. Deb, “A review on bilevel optimization: From classical to evolutionary approaches and applications,” *IEEE Transactions on Evolutionary Computation*, vol. 22, no. 2, pp. 276–295, 4 2018.

[98] S. Dempe, *Bilevel Optimization: Theory, Algorithms, Applications and a Bibliography*. Cham: Springer International Publishing, 2020, pp. 581–672. [Online]. Available: https://doi.org/10.1007/978-3-030-52119-6_20

[99] J. A. Gomez-Herrera and M. F. Anjos, “Optimal collaborative demand-response planner for smart residential buildings,” *Energy*, vol. 161, pp. 370 – 380, 2018. [Online]. Available: https://doi.org/10.1016/j.energy.2018.07.132

[100] M. Zugno, J. M. Morales, P. Pinson, and H. Madsen, “A bilevel model for electricity retailers’ participation in a demand response market environment,” *Energy Economics*, vol. 36, pp. 182–197, 2013.

[101] J. Yang, J. Zhao, F. Wen, and Z. Y. Dong, “A framework of customizing electricity retail prices,” *IEEE Transactions on Power Systems*, vol. 33, no. 3, pp. 2415–2428, May 2018.

[102] J. Vuelvas and F. Ruiz, “Rational consumer decisions in a peak time rebate program,” *Electric Power Systems Research*, vol. 143, pp. 533–543, 2017.

[103] G. Hébrail and A. Bérard, “Individual household electric power consumption data set,” https://archive.ics.uci.edu/ml/datasets/individual+household+electric+power+consumption, 08 2012.

[104] J. Bezanson, A. Edelman, S. Karpinski, and V. Shah, "Julia: A fresh approach to numerical computing," *SIAM Review*, vol. 59, no. 1, pp. 65–98, 2017. [Online]. Available: https://doi.org/10.1137/141000671

[105] M. Besançon, D. Anthoff, A. Arslan, S. Byrne, D. Lin, T. Papamarkou, and J. Pearson, "Distributions.jl: Definition and modeling of probability distributions in the JuliaStats ecosystem," *arXiv preprint arXiv:1907.08611*, 2019.

[106] S. G. Johnson, "QuadGK.jl: Gauss–Kronrod integration in Julia," https://github.com/JuliaMath/QuadGK.jl.

[107] I. Dunning, J. Huchette, and M. Lubin, "JuMP: A Modeling Language for Mathematical Optimization," *SIAM Review*, vol. 59, no. 2, pp. 295–320, 2017.

[108] B. Legat, O. Dowson, J. D. Garcia, and M. Lubin, "MathOptInterface: a data structure for mathematical optimization problems," *arXiv preprint arXiv:2002.03447*, 2020.

[109] T. Walsh, "Symmetry breaking constraints: Recent results," *arXiv preprint arXiv:1204.3348*, 2012.

[110] F. Margot, "Symmetry in integer linear programming," in *50 Years of Integer Programming 1958-2008.* Springer, 2010, pp. 647–686.

[111] G. Chu, "Improving combinatorial optimization," in *Twenty-Third International Joint Conference on Artificial Intelligence*, 2013.

[112] A. Ghoniem and H. D. Sherali, "Defeating symmetry in combinatorial optimization via objective perturbations and hierarchical constraints," *IIE Transactions*, vol. 43, no. 8, pp. 575–588, 2011.

[113] C. Hojny and M. E. Pfetsch, "Polytopes associated with symmetry handling," *Mathematical Programming*, vol. 175, no. 1-2, pp. 197–240, 2019.

[114] D. Bates, S. Kornblith, A. Noack, M. Bouchet-Valat, M. K. Borregaard, A. Arslan, J. M. White, D. Kleinschmidt, G. Lynch, I. Dunning, P. K. Mogensen, S. Lendle, D. Aluthge, J. B. S. Calderón, B. Born, B. Setzler, C. DuBois, J. Quinn, P. Bastide, pdeffebach, P. Anthony Blaom, C. Caine, D. Lin, H. Q. Ngo, J. Adenbaum, J. TagBot, L. Hein, and M. Grechkin, "Juliastats/glm.jl: v1.3.10," Aug. 2020. [Online]. Available: https://doi.org/10.5281/zenodo.3974173

[115] M. Aghassi and D. Bertsimas, "Robust game theory," *Mathematical Programming*, vol. 107, no. 1, pp. 231–273, Jun 2006.

[116] F. Ordóñez and N. E. Stier-Moses, "Robust Wardrop Equilibrium," in *Network Control and Optimization*, T. Chahed and B. Tuffin, Eds. Springer, 2007, pp. 247–256.

[117] M. Jain, F. Ordónez, J. Pita, C. Portway, M. Tambe, C. Western, P. Paruchuri, and S. Kraus, "Robust solutions in stackelberg games: Addressing boundedly rational human preference models," 2008.

[118] J. Pita, M. Jain, M. Tambe, F. Ordóñez, and S. Kraus, "Robust solutions to stackelberg games: Addressing bounded rationality and limited observations in human cognition," *Artificial Intelligence*, vol. 174, no. 15, pp. 1142–1171, 2010.

[119] V. Gabrel, C. Murat, and A. Thiele, "Recent advances in robust optimization: An overview," *European Journal of Operational Research*, vol. 235, no. 3, pp. 471–483, 2014.

[120] P. Loridan and J. Morgan, "Weak via strong stackelberg problem: New results," *Journal of Global Optimization*, vol. 8, no. 3, pp. 263–287, 1996.

[121] F. Liers, L. Schewe, and J. Thürauf, "Radius of robust feasibility for mixed-integer problems," 2019.

[122] M.A. Goberna and V. Jeyakumar and G. Li and N. Linh, "Radius of robust feasibility formulas for classes of convex programs with uncertain polynomial constraints," *Operations Research Letters*, vol. 44, no. 1, pp. 67–73, 2016.

[123] P. Pardalos and J. Rosen, "Methods for global concave minimization: A bibliographic survey," *SIAM Review*, vol. 28, no. 3, pp. 367–379, 1986.

[124] H. Benson and R. Horst, "A branch and bound-outer approximation algorithm for concave minimization over a convex set," *Computers & Mathematics with Applications*, vol. 21, no. 6, pp. 67 – 76, 1991.

[125] S. Boyd and L. Vandenberghe, *Convex optimization.* Cambridge University Press, 2004.

[126] S. Dempe, "Computing Locally Optimal Solutions of the Bilevel Optimization Problem Using the KKT Approach," in *International Conference on Mathematical Optimization Theory and Operations Research.* Springer, 2019, pp. 147–157.

[127] Y. Nesterov and A. Nemirovskii, *Interior-point polynomial algorithms in convex programming.* SIAM, 1994, vol. 13.

[128] H. Scheel and S. Scholtes, "Mathematical programs with complementarity constraints: Stationarity, optimality, and sensitivity," *Mathematics of Operations Research*, vol. 25, no. 1, pp. 1–22, 2000.

[129] C. Audet, J. Haddad, and G. Savard, "Disjunctive cuts for continuous linear bilevel programming," *Optimization Letters*, vol. 1, no. 3, pp. 259–267, 2007.

[130] S. Siddiqui and S. A. Gabriel, "An sos1-based approach for solving mpecs with a natural gas market application," *Networks and Spatial Economics*, vol. 13, no. 2, pp. 205–227, 2013.

[131] S. Pineda, H. Bylling, and J. Morales, "Efficiently solving linear bilevel programming problems using off-the-shelf optimization software," *Optimization and Engineering*, vol. 19, no. 1, pp. 187–211, 2018.

[132] P. Bonami, A. Lodi, A. Tramontani, and S. Wiese, "On mathematical programming with indicator constraints," *Mathematical Programming*, vol. 151, no. 1, pp. 191–223, Jun 2015, 10.1007/s10107-015-0891-4.

[133] R. Rahmaniani, T. G. Crainic, M. Gendreau, and W. Rei, "The benders decomposition algorithm: A literature review," *European Journal of Operational Research*, vol. 259, no. 3, pp. 801–817, 2017.

[134] T. Kleinert, M. Labbé, F. Plein, and M. Schmidt, "Closing the gap in linear bilevel optimization: a new valid primal-dual inequality," *Optimization Letters*, pp. 1–14, 2020.

[135] T. Ralphs and S. Tahernejad, "Bilevel Optimization Problem Library v0.1," 2020, https://coral.ise.lehigh.edu/data-sets/bilevel-instances/.

[136] M. Besançon, M. F. Anjos, and L. Brotcorne, "Near-optimal robust bilevel linear instances," Aug. 2020. [Online]. Available: https://doi.org/10.5281/zenodo.4110025

[137] J. Bezanson, A. Edelman, S. Karpinski, and V. Shah, "Julia: A Fresh Approach to Numerical Computing," *SIAM Review*, vol. 59, no. 1, pp. 65–98, 2017.

[138] A. Gleixner, M. Bastubbe, L. Eifler, T. Gally, G. Gamrath, R. L. Gottwald, G. Hendel, C. Hojny, T. Koch, M. E. Lübbecke, S. J. Maher, M. Miltenberger, B. Müller, M. E. Pfetsch, C. Puchert, D. Rehfeldt, F. Schlösser, C. Schubert, F. Serrano, Y. Shinano, J. M. Viernickel, M. Walter, F. Wegscheider, J. T. Witt, and J. Witzig, "The SCIP Optimization Suite 6.0," Optimization Online, Technical Report, July 2018. [Online]. Available: http://www.optimization-online.org/DB_HTML/2018/07/6692.html

[139] B. Legat, R. Deits, O. Evans, G. Goretkin, T. Koolen, J. Huchette, D. Oyama, M. Forets, guberger, R. Schwarz, E. Saba, and C. Coleman, "JuliaPolyhedra/Polyhedra.jl: v0.5.1," May 2019.

[140] B. Legat, R. Deits, M. Forets, D. Oyama, F. Pacaud, and E. Saba, "JuliaPolyhedra/CDDLib.jl: v0.5.2," May 2019.

[141] M. Besançon, M. F. Anjos, and L. Brotcorne, "Near-optimal Robust Bilevel Optimization," *arXiv preprint arXiv:1908.04040*, 2019.

[142] S. Dempe, V. Kalashnikov, G. A. Pérez-Valdés, and N. Kalashnykova, "Bilevel programming problems," *Energy Systems. Springer, Berlin*, 2015.

[143] G. B. Allende and G. Still, "Solving bilevel programs with the KKT-approach," *Mathematical Programming*, vol. 138, no. 1-2, pp. 309–332, 2013.

[144] S. Dempe and S. Franke, "Solution of bilevel optimization problems using the KKT approach," *Optimization*, vol. 68, no. 8, pp. 1471–1489, 2019.

[145] S.-J. Chung, "NP-completeness of the linear complementarity problem," *Journal of Optimization Theory and Applications*, vol. 60, no. 3, pp. 393–399, 1989.

[146] S. Dempe, O. Khamisov, and Y. Kochetov, "A special three-level optimization problem," *Journal of Global Optimization*, vol. 76, no. 3, pp. 519–531, 2020.

[147] S. Tahernejad, T. K. Ralphs, and S. T. DeNegre, "A branch-and-cut algorithm for mixed integer bilevel linear optimization problems and its implementation," *COR@L Technical Report 16T-015-R3, Lehigh University*, 2016.

[148] A. Caprara, M. Carvalho, A. Lodi, and G. J. Woeginger, "A study on the computational complexity of the bilevel knapsack problem," *SIAM Journal on Optimization*, vol. 24, no. 2, pp. 823–838, 2014.

[149] T. Ralphs, A. Bulut, S. DeNegre, A. Lodi, F. Rossi, S. Smriglio, and G. Woeginger, "Complexity and multi-level optimization," https://coral.ise.lehigh.edu/~ted/files/talks/MIP12.pdf.

[150] J. T. Moore and J. F. Bard, "The mixed integer linear bilevel programming problem," *Operations Research*, vol. 38, no. 5, pp. 911–921, 1990.

[151] M. Besançon, "NORBiP.jl," 2020, dOI: 10.5281/zenodo.4160164. [Online]. Available: https://github.com/matbesancon/NORBiP.jl

[152] B. Colson, P. Marcotte, and G. Savard, "An overview of bilevel optimization," *Annals of Operations Research*, vol. 153, no. 1, pp. 235–256, 2007.

[153] X. Shi and H. S. Xia, "Model and interactive algorithm of bi-level multi-objective decision-making with multiple interconnected decision makers," *Journal of Multi-Criteria Decision Analysis*, vol. 10, no. 1, pp. 27–34, 2001. [Online]. Available: http://dx.doi.org/10.1002/mcda.285

[154] R. Fernández-Blanco, J. M. Arroyo, and N. Alguacil, "On the Solution of Revenue- and Network-Constrained Day-Ahead Market Clearing Under Marginal Pricing Part I: An Exact Bilevel Programming Approach," *IEEE Transactions on Power Systems*, vol. 32, no. 1, pp. 208–219, 2017.

[155] ——, "On the Solution of Revenue- and Network-Constrained Day-Ahead Market Clearing Under Marginal Pricing Part II: Case Studies," *IEEE Transactions on Power Systems*, vol. 32, no. 1, pp. 220–227, 2017.

[156] A. Street, A. Moreira, and J. M. Arroyo, "Energy and reserve scheduling under a joint generation and transmission security criterion: An adjustable robust optimization approach," *IEEE Transactions on Power Systems*, vol. 29, no. 1, pp. 3–14, Jan 2014.

[157] S. Bahrami, M. H. Amini, M. Shafie-khah, and J. P. S. Catalao, "A Decentralized Renewable Generation Management and Demand Response in Power Distribution Networks," *IEEE Transactions on Sustainable Energy*, vol. 9, no. 4, pp. 1783–1797, 2018. [Online]. Available: http://dx.doi.org/10.1109/TSTE.2018.2815502

[158] "Demand response in liberalized electricity markets: Analysis of aggregated load participation in the German balancing mechanism," *Energy*, vol. 71, pp. 245–254.

[159] D. Dheeru and E. Karra Taniskidou, "UCI machine learning repository," http://archive.ics.uci.edu/ml, 2017.

[160] J. D. Hunter, "Matplotlib: A 2d graphics environment," *Computing In Science & Engineering*, vol. 9, no. 3, pp. 90–95, 2007.

[161] JuliaStats, "KernelDensity.jl v0.4.1," https://github.com/JuliaStats/KernelDensity.jl, 02 2018.

[162] R. Lougee-Heimer, "The Common Optimization INterface for Operations Research: Promoting Open-source Software in the Operations Research Community," *IBM*

Journal of Research and Development, vol. 47, no. 1, pp. 57–66, 2003. [Online]. Available: http://dx.doi.org/10.1147/rd.471.0057

[163] A. N. Riseth, M. Besançon, and A. Naiem, "MultiJuMP.jl," https://doi.org/10.5281/zenodo.1042586, 02 2019.

[164] JuliaCI, "BenchmarkTools.jl v0.3," https://github.com/JuliaCI/BenchmarkTools.jl, 05 2018.

RÉSUMÉ ÉTENDU

La thèse porte sur les modèles d'optimisation mathématiques biniveaux et leurs applications à la réponse de la demande dans les réseaux électriques intelligents.

L'augmentation de la production d'énergie renouvelable fluctuante et l'apparition de nouveaux acteurs ont complexifié les opérations et décisions dans les réseaux électriques. Ces phénomènes, ainsi que l'amélioration des techniques de communication et la connexion massive d'appareils électriques, ont mené à l'émergence du paradigme des réseaux électriques intelligents, dans lequel les flux d'information et d'énergie circulent de manière multidirectionnelle entre plusieurs entités.

La réponse de la demande est une des solutions étudiées pour pallier à la part croissante de production électrique fluctuante et distribuée de sources renouvelables. Au lieu d'adapter la génération électrique à la demande par du stockage ou un ajustement des sources d'énergie contrôlables, son principe est de modifier une part flexible de la consommation à travers des mécanismes techniques ou économiques, lissant ainsi la courbe de demande au cours du temps.

Nous étudions un système de prix dynamique (TLOU) dans lequel un client réserve une capacité pour une période donnée, et paie un prix différent en fonction du dépassement de la capacité par sa consommation. Ce système offre un compromis entre les systèmes de réponse de la demande utilisant les prix et ceux utilisant une incitation ou entente. Il combine une simplicité pour l'usager et un gain d'information sur les intentions de consommation pour le fournisseur.

La consommation électrique prédite pour une tranche horaire donnée est modélisée comme une variable aléatoire dont la distribution est connue des deux partis. Nous étudions les propriétés de TLOU, en particulier du point de vue du fournisseur qui décide des paramètres du système de prix, et représentons l'interaction entre le fournisseur et les usagers par une approche d'optimisation mathématique biniveau. Le fournisseur, agissant comme premier niveau, définit les prix et anticipe une réaction de l'usager au deuxième niveau qui optimise l'espérance de son coût en décidant d'une capacité.

Les problèmes d'optimisation biniveaux sont caractérisés par un problème d'optimisation imbriqué dans les contraintes d'un autre problème d'optimisation. Toute solution, pour être réalisable pour le problème biniveau, doit être optimale pour le problème imbriqué au deuxième niveau. Vérifier la faisabilité nécessite donc dans le cas général de résoudre à l'optimalité un problème d'optimisation. Les méthodes exactes de résolution des problèmes d'optimisation biniveaux exploitent dans de nombreux cas sur une reformulation en un problème mathématique à un seul niveau, en utilisant des conditions nécessaires et suffisantes d'optimalité du deuxième niveau. Ces conditions peuvent être dérivées de la convexité du problème du deuxième niveau ou de caractéristiques spécifiques au problème considéré. Ces deux approches sont utilisées dans cette thèse, des conditions particulières d'optimalité étant déterminées pour les problèmes de tarification et les conditions de Karush-Kuhn-Tucker étant utilisées pour la résolution de problèmes biniveaux génériques dans la deuxième partie. Ces conditions sont nécessaires et suffisantes dans le cas où le problème d'optimisation du deuxième niveau est convexe, c'est-à-dire qu'il présente une fonction objectif et un ensemble réalisable convexes.

En utilisant la formulation mathématique du problème d'optimisation et certaines propriétés de toute solution optimale, le problème est réduit à la résolution d'une suite de problèmes d'optimisation linéaire dans le cas où le fournisseur offre des prix différents à chaque usager. Une généralisation de la tarification TLOU optimale à plusieurs usagers est proposée avec un même prix reçu par de multiples utilisateurs. Dans de nombreux cas, le fournisseur n'offre pas un tarif différencié à chaque usager, mais définit des tarifs qui s'appliquent par groupe.

L'application à la détermination de prix a motivé le développement plus général de formulations biniveaux où le deuxième niveau n'est plus nécessairement résolu exactement, mais peut dévier de son optimum d'une quantité limitée. Dans ce cadre, le premier niveau anticipe non seulement la réaction optimale du deuxième niveau, mais aussi les déviations potentielles. Il cherche donc à protéger la faisabilité de ses contraintes, quelle que soit la déviation limitée du deuxième niveau autour de sa solution optimale.

Nous développons une formulation biniveau robuste à la quasi-optimalité (NORBiP) dans laquelle le premier niveau s'assure de trouver une solution dont la faisabilité est garantie pour l'ensemble des solutions quasi optimales du deuxième niveau. Les solutions quasi optimales sont définies comme les solutions réalisables pour le deuxième niveau et dont la valeur objectif n'est pas supérieure à la valeur optimale de plus d'une tolérance prédéterminée. Ce modèle généralise la formulation pessimiste de l'optimisation biniveau et introduit une notion de robustesse spécifique à l'optimisation multiniveau. Une reformulation à un seul niveau est développée dans le cas où le deuxième niveau est un problème d'optimisation convexe, basée sur la dualisation des contraintes de robustesse. Le problème résultant contient en particulier des contraintes de complémentarité et contraintes bilinéaires. Dans le cas où le deuxième niveau est un problème linéaire, nous développons une formulation étendue remplaçant des contraintes non linéaires non convexes de la reformulation par une contrainte disjonctive linéaire qui peut être traitée par les solveurs génériques pour problèmes d'optimisation linéaire en variables mixtes. Des expériences numériques montrent également que la résolution de cette formulation étendue avec disjonction domine la résolution par approche de branchement spatiale de la formulation compacte utilisant un solveur global.

L'optimisation biniveau appartient à une classe de problèmes d'optimisation considérés comme difficiles à résoudre, dans le sens où il n'existe à priori pas d'algorithme en temps polynomial permettant leur résolution. Bien que la robustesse à la quasi-optimalité développée dans cette thèse ajoute une difficulté supplémentaire au problème biniveau classique, des résultats de complexité sont établis, démontrant qu'un problème biniveau ou plus généralement multiniveau robuste à la quasi optimalité appartient sous certaines conditions relativement générales à la même classe de complexité que le même problème sans contraintes de robustesse.

Dans le dernier chapitre de contributions, des algorithmes sont proposés pour accélérer la résolution de problèmes biniveaux robustes à la quasi-optimalité dans le cas où le problème au premier et deuxième niveau sont linéaires. Deux algorithmes exacts exploitent l'indépendance de chaque contrainte du premier niveau, permettant d'éviter la résolution de la formulation étendue avec une disjonction sur chacune de ces contraintes. Pour toute solution du premier et deuxième niveau, un sous-problème dit "adverse" définit pour chaque contrainte du premier niveau le pire cas quasi optimal pour cette contrainte. Ce problème adverse est un problème d'optimisation linéaire similaire au problème du deuxième niveau initial et dont la résolution vérifie qu'une solution est robuste. Si une solution n'est pas robuste, la contrainte disjonctive de la résolution étendue correspondant à cette contrainte du premier niveau est ajoutée à la formulation à la volée et le problème est résolu à nouveau.

Une méthode heuristique dite à sommet unique est également proposée pour la recherche rapide d'une solution réalisable, optimale pour le deuxième niveau et robuste à la quasi-optimalité. Au lieu d'ajouter des contraintes disjonctives comme dans les méthodes exactes proposées ci-dessus, cette méthode résout la version optimiste non robuste du problème biniveau, puis pour chaque sous-problème, vérifie la robustesse. Pour chaque sous-problème indiquant une non-robustesse de la contrainte, une seule contrainte linéaire est ajoutée, correspondant à un sommet du polyèdre dual au lieu d'une disjonction sur l'ensemble des sommets dans le cas de la méthode exacte. Cette méthode garantit la faisabilité et robustesse de la solution, mais pas son optimalité.

Ces différents algorithmes sont comparés à la simple formulation étendue en termes de temps de résolution et de qualité de la solution. Le temps de résolution est significativement réduit par les méthodes exactes accélérées. L'heuristique proposée permet de réduire encore le temps de calcul tout en obtenant des solutions de bonne qualité en termes de valeur objectif.

Les contributions de cette thèse sont donc d'une part le développement d'une approche d'optimisation mathématique biniveau appliquée à un nouveau système de tarification dans les réseaux électriques, et d'autre part le développement plus général d'une approche robuste pour les problèmes d'optimisation biniveaux génériques. Pour ces deux parties, plusieurs méthodes algorithmiques sont comparées durant la résolution, fournissant des indications sur la structure des problèmes mixtes en nombres entiers résolus par algorithmes exacts de branchement disponibles dans de multiples suites logicielles.

Les deux parties du travail présenté se concentrent principalement autour de la définition de nouveaux problèmes d'optimisation biniveau. De futurs travaux peuvent s'inscrire en continuité en proposant des variantes des problèmes définis, des applications et développements de nouvelles méthodes de résolution.

Plusieurs pistes de recherche permettent d'étendre les travaux sur le système de tarification TLOU. Une première sera l'intégration de la tarification dans un modèle global de décision d'un fournisseur ou agrégateur dans un réseau électrique, intégrant d'une part une représentation complète d'usagers potentiellement équipés d'une capacité de stockage et de production d'énergie distribuée, et d'autre part les contraintes de génération d'électricité dans un problème de planification à moyen terme type unit commitment.

Le modèle d'optimisation biniveau robuste à la quasi-optimalité tel que développé dans cette thèse se base sur le concept de robustesse sur les contraintes, une solution au problème d'optimisation est réalisable seulement si les contraintes sont valides pour toute valeur du paramètre incertain dans l'ensemble des réalisations possibles, l'ensemble des solutions quasi-optimales pour le deuxième niveau dans ce cas. De futurs travaux pourront étudier d'autres formes de robustesse développées dans la littérature (Γ-robustesse) et les appliquer à la robustesse à la quasi-optimalité qui est principalement caractérisée par son ensemble d'incertitude.

Un autre axe de recherche présenté par la définition de ce modèle concerne le développement de méthodes de résolution pour d'autres classes de problèmes biniveaux. Les algorithmes et formulations proposés dans ces travaux utilisent les propriétés spécifiques de linéarité des contraintes et de l'objectif du problème du deuxième niveau, et des contraintes du premier niveau par rapport aux variables du deuxième niveau. Deux sous-classes de problèmes d'optimisation biniveaux pourront en particulier bénéficier de méthodes de résolution pour la formulation robuste à la quasi-optimalité, les problèmes où le deuxième niveau est un problème linéaire à variables mixtes et les problèmes où le deuxième niveau est un problème convexe, mais non linéaire. Dans le cas du problème de deuxième niveau à variables mixtes, la reformulation utilisant la dualité du deuxième niveau ne s'applique plus, et un algorithme de branchement spécifique devra être défini pour éliminer les solutions non robustes. Dans le cas de problème de deuxième niveau convexe, mais non linéaire, la formulation étendue résulterait en des contraintes basées sur des disjonctions sur une infinité de contraintes.

APPENDIX A ARTICLE 4: A BILEVEL APPROACH TO OPTIMAL PRICE-SETTING OF TIME-AND-LEVEL-OF-USE TARIFFS

This appendix was presented as extended abstract at, and included in the program of, the 13th EUROGEN conference on Evolutionary and Deterministic Computing for Industrial Applications. It complements Chapter 4.

Authors: Mathieu Besançon, Miguel F. Anjos, Luce Brotcorne, Juan A. Gomez-Herrera

Summary

Time-and-Level-of-Use (TLOU) is a recently proposed pricing policy for energy, extending Time-of-Use with the addition of a capacity that users can book for a given time frame, reducing their expected energy cost if they respect this self-determined capacity limit. We introduce a variant of the TLOU defined in the literature, aligned with the supplier interest to prevent unplanned over-consumption. The optimal price-setting problem of TLOU is defined as a bilevel, bi-objective problem anticipating user choices in the supplier decision. An efficient resolution scheme is developed, based on the specific discrete structure of the lower-level user problem. Computational experiments using consumption distributions estimated from historical data illustrate the effectiveness of the proposed framework.

A.1 Introduction

The increasing penetration of wind and solar energy has marked the last decades, bringing a decentralization and higher stochasticity of power generation, yielding new challenges for the operation of electrical grids. Advances in communication technologies have enabled both scheduling of smart appliances and seamless data collection and exchange between different entities operating on power networks, from generators to end-consumers. Using such capabilities, agents can make decisions based on probability distributions of different variables, possibly conditioned on known external factors, such as the weather.

As a promising lead to tackle these challenges, Demand Response (DR) has attracted an increasing interest in the past decades. Instead of compensating ever-increasing fluctuations of renewable production with conventional power generation, this approach consists in leveraging the flexibility of some consuming units by reducing or shifting their load [12]. The Time-and-Level-of-Use system presented in this paper is primarily a price-based Demand Response program based on the classification found in the literature [12], although the self-determined consumption limit creates an incentive for respecting the commitment on the part of the user. Its design makes opting out of the program a natural extension and thus requires less coupling and interaction between users and suppliers to work. In a 2012 report, the US Federal Energy Regulatory Commission identified the lack of short-term estimation methods as one of the critical barriers to the effective implementation of Demand Response [95]. In a related approach [19], an incentive-based Demand Response program is developed with users signaling both a predicted consumption and a reduction capacity, with the supplier selecting reductions randomly. Both the program developed by the authors [19] and the TLOU policy require a signal sent from the user to the supplier.

Bilevel optimization has been used to model and tackle optimization problems in energy networks [47]. Applications on power systems and markets are also mentioned in several reviews on bilevel optimization [152,153]. It allows decision-makers to encapsulate utility-generators, user-utility interactions [154,155], or define robust formulations of unit commitment and optimal power flow [156]. Some recent work focuses on bilevel optimization for demand-response in the real-time market [157]. Multiple supplier settings are also considered, for instance in the determination of Time-of-Use pricing policies [101], where competing retailers target residential users.

The Time-and-Level-of-Use (TLOU) pricing scheme [27] was built as an extension of Time-of-Use (TOU), targeting specifically the issue with current large-scale Demand-Response programs identified in the FERC report [95]. The optimal planning and operation of a smart building under this scheme was developed, taking the TLOU settings as input of the decision. In this work, we consider the perspective of the supplier determining the optimal parameters of TLOU.

We integrate the optimal user reaction in the supplier decision problem, thus modelling the interaction as a Stackelberg game solved as a bilevel optimization problem. The specific structure of the lower-level decision is leveraged to reduce the set of possibly optimal solutions to a finite enumerable set. Through this reduction, the lower-level optimality conditions are translated into a set of linear constraints. The set of optimal pricing configurations is derived for different distributions corresponding to different time frames. These options can be computed in advance and then one is chosen by the supplier ahead of the consumption time to create an incentive to book and consume a given capacity. We show the effectiveness of the method with discrete probability distributions computed from historical consumption data.

A.2 Time-and-Level-of-Use pricing

The TLOU policy was initially developed from the perspective of smart consumption units [27] which can be a building, an apartment or a micro-grid monitoring its consumption and equipped with programmable consuming devices. It is a pricing model for energy built upon the Time-of-Use (TOU) implemented by several jurisdictions and for which the energy price is fixed by intervals throughout the day. TLOU extends TOU by allowing users to self-determine and book an energy capacity at each time frame depending on their planned requirements; and by doing so, they provide the supplier with information on the intended consumption. We will refer to the capacity as the amount of energy booked by the user for a given time frame, following the same terminology as the reference implementation [27]. Energy prices still depend on the time frame within the day, but also on the capacity booked by the user. This pricing scheme is applied in a three-phase process:

1. The supplier sends the pricing information to the user.
2. The user books a capacity from the supplier for the time frame before a given deadline. If no capacity has been booked, the Time-of-Use pricing is used.
3. After the time frame, the energy cost is computed depending on the energy consumed x and booked capacity c:
   - If $x \leq c$, then the applied price of energy is $\pi^L(c)$ and the energy cost is $\pi^L(c) \cdot x$.
   - If $x > c$, then the applied price of energy is $\pi^H(c)$ and the energy cost is $\pi^H(c) \cdot x$.

In the model considered in this paper, only the first two steps involve decisions from the agents, making the decision process a Stackelberg game given the sequentiality of these decisions. The price structure is composed of three elements: a booking fee K, a step-wise decreasing function $\pi^L(c)$ representing the lower energy price and a step-wise increasing function $\pi^H(c)$ representing the higher energy price. The steps of the lower and higher price functions are given at different breakpoints:

$$\begin{aligned} &\{c_0^L, c_1^L, c_2^L, ...\} = C^L && \& \; \{\pi_0^L, \pi_1^L, \pi_2^L, ...\} = \pi^L \\ &\{c_0^H, c_1^H, c_2^H, ...\} = C^H && \& \; \{\pi_0^H, \pi_1^H, \pi_2^H, ...\} = \pi^H \end{aligned}$$

$\pi^L(c)$ will refer to the function of the capacity and π_j^L to the value of the lower price at step j. In the initial version of the pricing [27], the energy consumed above the capacity is paid at the higher tariff while the rest is paid at the lower tariff.

Considering that most power systems try to prevent over-consumption and unplanned excess consumption, we introduce a variant where the whole energy consumed is paid at the lower tariff if it is less or equal to the capacity, and at the higher tariff otherwise, see Equation (A.1). In other words, if the consumption over the time frame remains below the booked capacity, the effective energy price is given by the lower tariff curve; if the consumption exceeds the booked capacity, the energy price is given by the higher tariff. The total cost for the user associated with a booked capacity c and a consumption X_t for a time frame t is:

$$\mathcal{C}(c_t; X_t) = \begin{cases} K \cdot c_t + \pi^L(c_t) \cdot X_t, & \text{if } X_t \ \leq \ c_t, \\ K \cdot c_t + \pi^H(c_t) \cdot X_t & \text{otherwise.} \end{cases} \tag{A.1}$$

The load distribution X_t is a random variable; both user and supplier make decisions on the expected cost over the set of possible consumption levels Ω, given as:

$$\mathcal{C}(c_t) = \mathbb{E}_\Omega \left[\mathcal{C}(c_t; x_\omega)\right] \tag{A.2}$$

where $\mathbb{E}_\Omega$ is the expected value over the support Ω of the probability distribution.

Furthermore, $\pi^L(c = 0) = \pi^H(c = 0) = \pi_0(t)$, with $\pi_0(t)$ the Time-of-Use price at the time frame of interest t. This property allows users to opt-out of the program for some time frames by simply not booking any capacity. TLOU is designed for the day-ahead market, where both the pricing components and the capacity are chosen ahead of the consumption time [27]. It can however be adjusted to other markets [158] or intra-day settings. The entity defining the TLOU pricing can also extend the possible settings by decoupling the time spans for one price setting choice from the time frames for capacity booking. For instance, a price setting can be chosen by the utility for the week, while the booking of capacity occurs the day before the consumption.

TLOU offers the user the possibility to reduce their cost of energy by load planning, and offers the supplier the prospect of improved load forecasting, because under-consumption is paid by the excess booking cost while over-consumption is paid by the difference between higher and lower tariffs.

We illustrate this phenomenon with a supplier decision (K, π^L, π^H) on Figures A.1 and A.2

using the relative cost:

$$x \mapsto \frac{\mathcal{C}(c;x)}{x} \tag{A.3}$$

for different values of c. For $c > 0$, the fixed cost $K \cdot c$ makes low consumption levels expensive per consumed unit, while the transition from lower to higher price makes over-consumption more expensive than the baseline. In the example, $c = 1.5$ cannot be an optimal booked capacity since it is always more expensive than the baseline for all possible consumption levels. The cases $c = 3.0$ and $c = 3.5$ both have ranges of consumption for which this choice is optimal for the user; these ranges always have the capacity as upper bound. The difference in cost occurring at the transition from lower to higher price increases with the booked capacity, illustrating the guarantee against over-consumption the supplier gains from the capacity booked by the user. Regardless of the ability to shift consumption or irrational decision-making, the commitment to a capacity creates a cost difference if the user is not making an optimal decision, which compensates the supplier for these unexpected deviations. This financial incentive against a consumption above the booked capacity creates a guarantee of interest to the supplier and is cast as a second objective expressed in Section A.3.

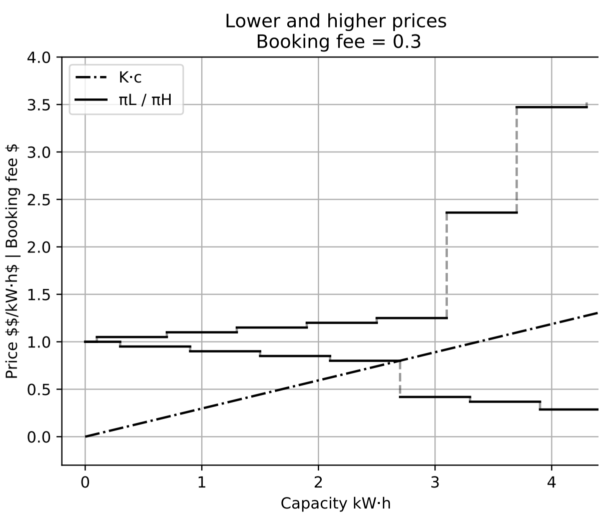

Figure A.1 Example of TLOU pricing

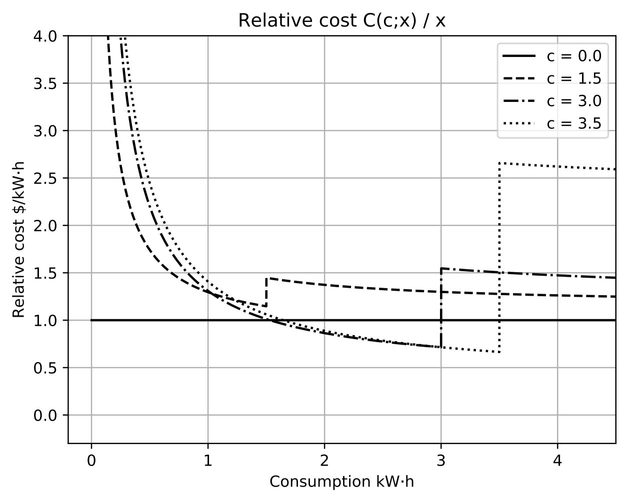

Figure A.2 Relative cost of energy vs consumption for different capacities

The supplier first builds their set of options based on prior consumption data. In the proposed method, the prior discrete distribution used can be conditioned on some independent variables if they are known and influence the consumption (e.g. forecast external temperature or day of the week). They can then pick a pricing setting for a given day based on generation-side considerations and constraints, including the option to stay at a flat Time-of-Use tariff for some or all time frames.

A.3 Bilevel model of the supplier problem

The supplier wishes to determine an optimal set of pricing options at any capacity level. In the model developed in this section, we consider a discrete probability distribution with a finite support, derive some properties of the cost structure which we then leverage to formulate the optimization problem in a tractable form.

At any capacity level, the decision process of the supplier involves two objectives, the expected revenue from the tariff and the guarantee of an upper bound on the consumption. The expected revenue is given by:

$$\mathcal{C}(c) = K \cdot c + \sum_{\omega \in \Omega^{-}(c)} x_\omega p_\omega \pi^L(c) + \sum_{\omega \in \Omega^{+}(c)} x_\omega p_\omega \pi^H(c) \tag{A.4}$$

with any capacity booked defining a partition of the set of scenarios:

$$\Omega^{-}(c) = \{\omega \in \Omega, x_\omega \leq c\} \tag{A.5}$$

$$\Omega^{+}(c) = \{\omega \in \Omega, x_\omega > c\} \tag{A.6}$$

The function $\mathcal{C}(c)$ is minimized by the user with respect to their decision c. It is non-linear, non-smooth and discontinuous because of the partition of the scenarios by c and the transition between steps of the pricing curves $\pi^L(c)$, $\pi^H(c)$. Both the user and supplier problems are thus intractable with this initial formulation. Proposition A.3.1 shows that only a discrete finite subset of capacity values are candidates to optimality for the user.

Proposition A.3.1. *The optimal booked capacity for a user at a time frame t belongs to a discrete and finite set of capacities S_t, defined as:*

$$S_t = \{0\} \cup C^L \cup \Omega_t \tag{A.7}$$

with Ω the set of consumption scenarios.

Proof. The user objective function is the sum of the booking cost and the expected electricity cost. The booking cost is linear in the booked capacity, with a positive slope equal to the booking fee. The expected electricity cost is piecewise constant in the booked capacity, with discontinuities at steps of both of the price curves because of the π^L and π^H prices and at possible load levels because of the transfer of a load from Ω^+ to the Ω^- set. This can be highlighted using the indicator functions associated with each of the two sets:

$$\mathbb{1}^-(\omega, c) = \begin{cases} 1, & \text{if } x_\omega \leq c, \\ 0 & \text{otherwise.} \end{cases} \tag{A.8}$$

$$\mathbb{1}^+(\omega, c) = 1 - \mathbb{1}^-(\omega, c) \tag{A.9}$$

The expression of the user expected cost becomes:

$$\mathcal{C}(c) = K \cdot c + \sum_{\omega \in \Omega} x_\omega \cdot p_\omega \cdot \left(\pi^L(c) \cdot \mathbb{1}^-(\omega, c) + \pi^H(c) \cdot \mathbb{1}^+(\omega, c)\right) \tag{A.10}$$

The sum of the two terms is therefore piecewise linear with a positive slope. On any interval between the discontinuity points, the optimal value lies on the lower bound, which can be any point of C^L, C^H, Ω or 0.

Furthermore, let $\mathcal{C}(c)$ be the user cost for a booked capacity and $\bar{c}$ such that $\bar{c} \in C^H$ and $\bar{c} \notin \{0\} \cup C^L \cup \Omega$. The higher tariff levels are monotonically increasing. Let $\varepsilon > 0$ be sufficiently small such that:

$$\begin{aligned}
&\pi^H(\bar{c} - \varepsilon) = \pi_n, \\
&\pi^H(\bar{c} + \varepsilon) = \pi_{n+1} > \pi_n, \\
&\pi^L(\bar{c} - \varepsilon) = \pi^L(\bar{c} + \varepsilon) = \pi^L_m, \\
&\nexists x_\omega,\ \omega \in \Omega \text{ s.t. } \bar{c} - \varepsilon \leq x_\omega \leq \bar{c} + \varepsilon.
\end{aligned}$$

The last condition guarantees there is no load value in the $[\bar{c} - \varepsilon, \bar{c} + \varepsilon]$ interval and can also be expressed in terms of the two load sets split by the capacity:

$$\Omega^+(\bar{c} - \varepsilon) = \Omega^+(\bar{c} + \varepsilon) \text{ and } \Omega^-(\bar{c} - \varepsilon) = \Omega^-(\bar{c} + \varepsilon).$$

Then if such ε exists, we find that:

$$\begin{aligned}
\mathcal{C}(\bar{c}-\varepsilon) &= K\cdot(\bar{c}-\varepsilon) + \sum_{\omega\in\Omega^-(\bar{c})} \pi^L_m \cdot x_\omega + \sum_{x\in\Omega^+(\bar{c})} \pi^H_n \cdot x_\omega,\\
\mathcal{C}(\bar{c}+\varepsilon) &= K\cdot(\bar{c}+\varepsilon) + \sum_{\omega\in\Omega^-(\bar{c})} \pi^L_m \cdot x_\omega + \sum_{\omega\in\Omega^+(\bar{c})} \pi^H_{n+1} \cdot x_\omega,\\
\mathcal{C}(\bar{c}+\varepsilon) - \mathcal{C}(\bar{c}-\varepsilon) &= 2\varepsilon K + \sum_{\omega\in\Omega^+(\bar{c})} (\pi^H_{n+1} - \pi^H_n)\cdot x_\omega,\\
\mathcal{C}(\bar{c}+\varepsilon) - \mathcal{C}(\bar{c}-\varepsilon) &> 0.
\end{aligned}$$

The discontinuity on any $x \in C^H$ is therefore always positive and cannot be a candidate for optimality. It follows that optimality candidates are restricted to the set $S = \{0\}\cup C^L \cup \Omega$. □

Proposition A.3.1 means we can replace the continuous decision set of capacities with a discrete set that can be enumerated.

The guarantee of an upper bound on the consumption $\mathcal{G}$ corresponds to the incentive given to the user against consuming above the considered capacity. It is the second objective, given by the difference in cost at the capacity, which is the immediate difference in total cost at the transition from lower to higher tariff:

$$\mathcal{G}(c, \pi^L, \pi^H) = c_t \cdot \left(\pi^H(c) - \pi^L(c)\right). \tag{A.11}$$

The supplier needs to include the user behavior and optimal reaction in their decision-making process, which can be done by a bilevel constraint:

$$c_t \in \arg\min_c \mathcal{C}(c). \tag{A.12}$$

The user thus books the least-cost option at each time frame, given the corresponding probability distribution. Given the finite set of optimal candidates S_t defined in A.3.1, this constraint can be re-written as:

$$c_t \leq c \;\forall c \in S_t. \tag{A.13}$$

If multiple choices of c yield the minimum cost, the choice of the user is not well-defined. The supplier would want to ensure the uniqueness of the preferred solution by making it lower than the expected cost of any other capacity candidate by a fixed quantity $\delta > 0$. This quantity can be interpreted as the conservativeness of the user (unwillingness to move to an optimal solution up to a difference of δ). It is a parameter of the decision-making process, estimated by the supplier. The lower-level optimality constraint is then for a preferred candidate k:

$$\mathcal{C}(c_{tk}, K, \pi^L, \pi^H) \leq \mathcal{C}(c_{tl}, K, \pi^L, \pi^H) + \delta \; \forall l \in S_t \backslash k. \tag{A.14}$$

A.3.1 Additional contractual constraints

In order to obtain regular price steps, the contract between supplier and consumer can include further constraints on the space of pricing parameters. We include three types of constraints: lower and upper bounds on the booking fee K, minimum and maximum increase at each step of the higher price and minimum and maximum decrease at each step of the higher price. All these can be expressed as linear constraints, and we gather them under the constraint set:

$$(K, \pi^L, \pi^H) \in \Phi. \tag{A.15}$$

A.3.2 Complete optimization model

The model is defined for each of the capacity candidates and is thus noted $\mathcal{P}_{tk}$ for candidate k and time frame t:

$$\max_{K,\pi^L,\pi^H} \left(\mathcal{C}(c_{tk}, K, \pi^L, \pi^H), \mathcal{G}(c_{tk}, \pi^L, \pi^H)\right) \tag{A.16}$$

$$(K, \pi^L, \pi^H) \in \Phi \tag{A.17}$$

$$\mathcal{C}(c_{tk}, K, \pi^L, \pi^H) \leq \mathcal{C}(c_{lt}, K, \pi^L, \pi^H) + \delta \; \forall l \in S_t \backslash k, \tag{A.18}$$

where

$$\mathcal{C}(c_{tk}, K, \pi^L, \pi^H) = K \cdot c_{tk} + \sum_{\omega \in \Omega_t^-(c_{tk})} x_\omega p \omega \pi_k^L + \sum_{\omega \in \Omega_t^+(c_{tk})} x_\omega p_\omega \pi_k^H, \tag{A.19}$$

$$\mathcal{G}(c_{tk}, \pi^L, \pi^H) = c_{tk} \cdot \left(\pi^H(c) - \pi^L(c)\right). \tag{A.20}$$

A.4 Solution method and computational experiments

The proposed model was implemented and tested using historical consumption data [159] measured on a pilot house. The instantaneous consumption was measured every two minutes during 47 months on a residential building by the energy supplier and grid operator EDF [103]. Since the focus is the energy consumed within a given time frame, the instantaneous power can be averaged for each hour, yielding the energy in $kW \cdot h$ and avoiding issues of missing measurements in the dataset. Data preprocessing, density estimation and discretization, and visualization were performed using Julia [104], matplotlib [160] and *KernelDensity.jl* [161]. The construction and optimization of the model were carried out using CLP from the COIN-OR project [162] as a linear solver and JuMP [107]. For all experiments where it is not specified, an inertia of $\delta = 0.05$ has been applied. For every time frame and for all capacity candidates, the bi-objective supplier problem with objectives $(\mathcal{C}, \mathcal{G})$ is solved with the ε-constraint method implemented in *MultiJuMP.jl* [163]. In all cases, the objectives are found to be non-conflicting, in the sense that the utopia point of the multi-objective problem is feasible and reached. This implies that a lexicographic multi-objective optimization solves the problem and reaches the utopia point, but does not guarantee that this holds for all problem configurations.

Figure A.3 shows the number of options computed at each hourly time frame and Figure A.4 the capacity level in $kW \cdot h$ of each option. The option of a capacity level of 0 is always possible. Two examples of TLOU settings obtained are presented in Figures A.5 and A.6.

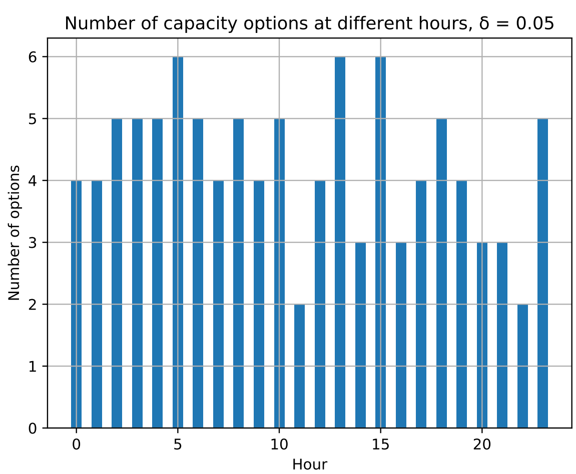

Figure A.3 Number of options at different hours

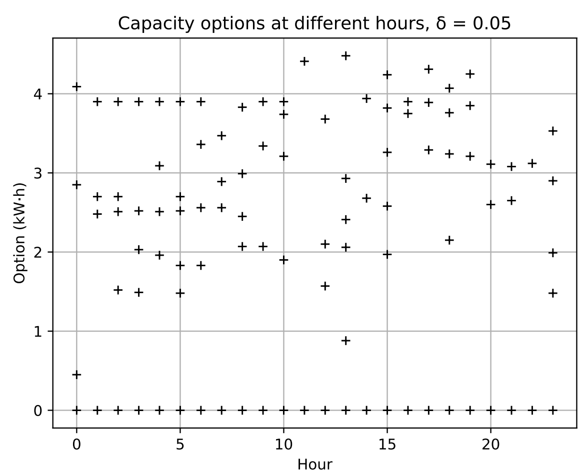

Figure A.4 Capacity levels of options at different hours

The expected cost for the user of booking any capacity, given the price setting provided in Figure A.5 is shown Figure A.7. The most notable result is that in all cases tested, the utopia point, defined as the optimal value of the two objectives optimized separately, is reachable.

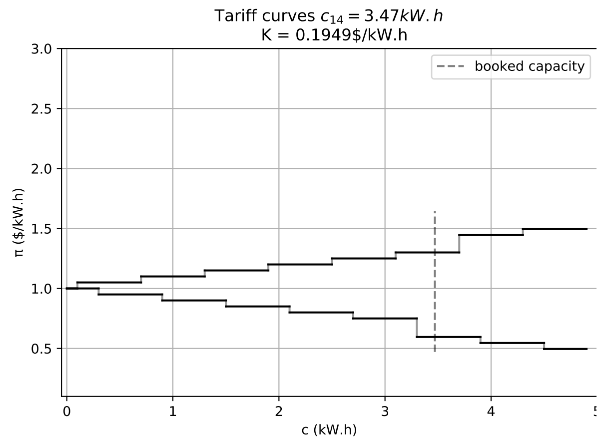

Figure A.5 TLOU price settings to incentive for capacity $c = 3.47kW \cdot h$

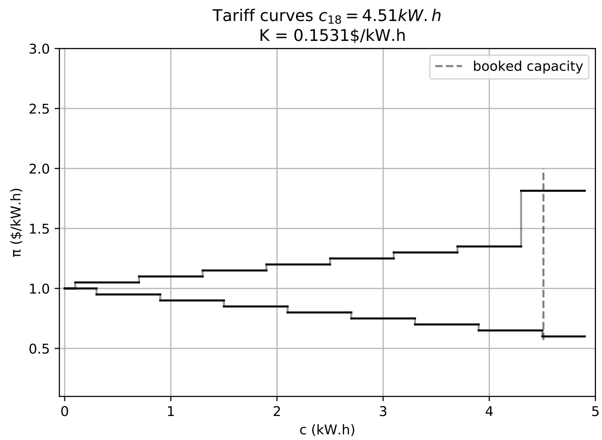

Figure A.6 TLOU price settings to incentive for capacity $c = 4.51kW \cdot h$

This result is conceivable given that the supplier decision is taken in a high-level space, allowing multiple solutions to be optimal with respect to the revenue. In order to ensure the guarantee-maximizing optimal solution, a two-step process lexicographic multi-optimization procedure is used:

1. Solve the revenue-maximizing problem to obtain the maximal reachable revenue v.

2. Solve the guarantee-maximizing problem, while constraining a revenue $\mathcal{C}(\cdot) \geq v$.

Figures A.8 and A.9 present a TLOU configuration optimized for costs only and for cost and guarantee.

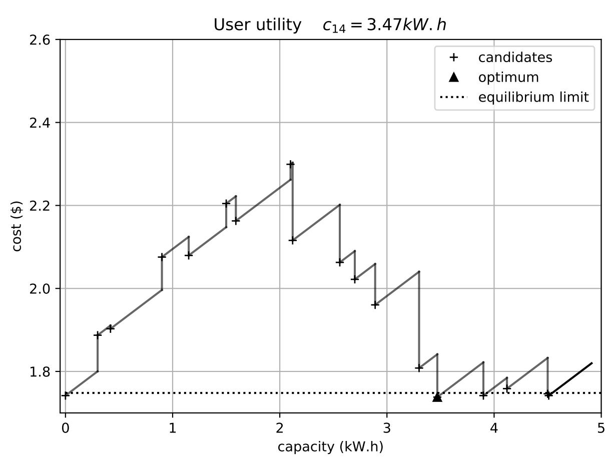

Figure A.7 Expected cost for any capacity booked under the price settings of Figure A.5

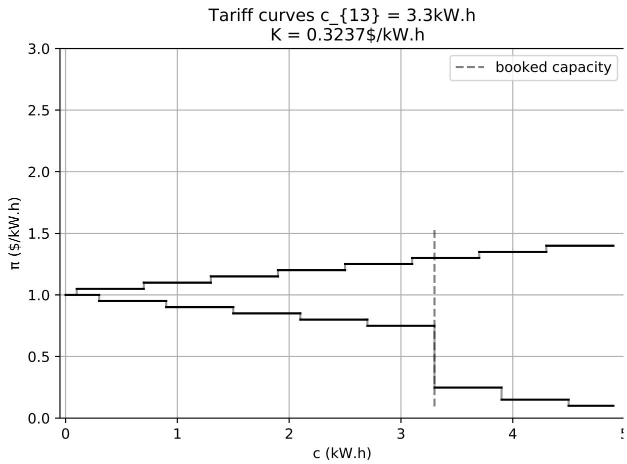

Figure A.8 TLOU price settings for revenue maximization only

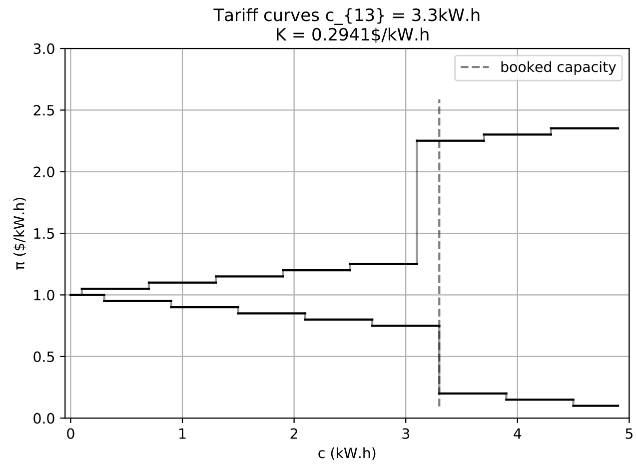

Figure A.9 TLOU price settings after lexicographic optimization

All the models solved are linear optimization problems with a fixed number of variables and a number of constraints growing linearly with the number of scenarios considered. However, all these constraints are of type:

$$\mathcal{C}(c_{tk}, K, \pi^L, \pi^H) \leq \mathcal{C}(c_{tl}, K, \pi^L, \pi^H) + \delta \ \forall l \in S_t \backslash k. \tag{A.21}$$

This is equivalent to:

$$\mathcal{C}(c_{tk}, K, \pi^L, \pi^H) \leq \min_{l \in S_t \backslash k} \mathcal{C}(c_{tl}, K, \pi^L, \pi^H) + \delta. \tag{A.22}$$

Therefore, at most one of the l will be active with a non-zero dual cost. The method can thus be scaled to a greater number of scenarios by adding these constraints on the fly.

With the current discretized distributions containing between 5 and 10 scenarios, the mean and median times to compute the whole solutions for all candidate capacities are below $\frac{1}{40}$ of a second. These metrics are obtained using the BenchmarkTools.jl package [164].

A study of the influence of the δ parameter is summarized Figure A.10. For any hour, there always exists a maximum value δ_{max} above which it becomes impossible to make a solution better for the lower-level than the baseline with a difference greater than δ_{max}.

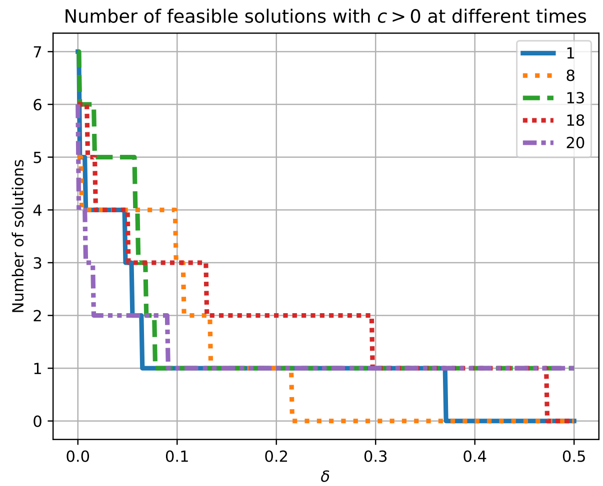

Figure A.10 Number of non-zero capacity solutions

A.5 Conclusion

TLOU is designed to price energy across time and to reflect varying costs and requirements from the generation side. Defining two objectives for the supplier, we built the set of cost-optimal price settings maximizing the guarantee in a lexicographic fashion. Computations on distributions built from real data show the effectiveness of the method, requiring a low runtime to compute the set of solutions. Future research will consider continuous probability distributions of the consumption and the price-setting problem with multiple users.

A.6 Acknowledgment

This work was supported by the NSERC Energy Storage Technology (NEST) Network.