université
PARIS-SACLAY

NNT : 2019SACLN064

Jeux Stochastiques sur des Graphes avec des Applications à l'Optimisation des Smart-Grids

Thèse de doctorat de l'Université Paris-Saclay
préparée à l'École Normale Supérieure Paris-Saclay (LSV)
& à CentraleSupélec (L2S)

École doctorale n°580 : Sciences et Technologies de l'Information et de la Communication (STIC)

Thèse préparée par

Mauricio GONZÁLEZ

Composition du jury :

David PARKER — Rapporteur
Professeur,
Université de Birmingham, School of Computer Science

Yezekael HAYEL — Rapporteur
Maitre de conférence,
Université d'Avignon, LIA/CERI

Clémence ALASSEUR — Examinatrice
Ingénieur de Recherche,
EDF, Laboratoire de Finance des Marchés de l'Énergie

Serge HADDAD — Examinateur
Professeur des Universités,
Ecole Normale Supérieure Paris-Saclay, LSV

Patrick PANCIATICI — Examinateur
Ingénieur,
RTE France

Patricia BOUYER — Directrice de thèse
Directeur de Recherche,
École Normale Supérieure Paris-Saclay, LSV

Samson LASAULCE — Co-directeur de thèse
Directeur de Recherche,
CentraleSupélec, L2S

Nicolas MARKEY — Co-directeur de thèse
Directeur de Recherche,
Université de Rennes 1, IRISA

Thèse de doctorat

STOCHASTIC GAMES ON GRAPHS WITH APPLICATIONS TO SMART-GRIDS OPTIMIZATION

MAURICIO GONZÁLEZ GÓMEZ

Ph.D Candidate
French National Center for Scientific Research (CNRS)
University of Paris-Saclay

Laboratory of
Specification and Verification (LSV)
École Normale Supérieure (ENS) Paris-Saclay
&
Laboratory of
Signal and Systems (LSS)
CentraleSupélec

November 2019– Version 2.3

SUPERVISORS:
Patricia Bouyer-Decitre
Samson Lasaulce
Nicolas Markey

RAPPORTEURS:
David Parker
Yezekael Hayel

LOCATION:
France

ABSTRACT

Applications such as energy networks become more and more important in our modern world. To design these networks, engineers resort more and more to advanced mathematical tools. Two key features for the design of a network are correctness and optimality. While correctness and optimality are in the core of formal methods, their effective application to energy networks remains largely unexplored. This constitutes one strong motivation for the work developed in this thesis, which is strongly based on formal methods of computer science and game theory. A special emphasis is made on the generic problem of power consumption scheduling. This is a scenario in which the consumers have a certain energy demand and want to have this demand to be fulfilled before a set deadline (e.g., an Electric Vehicle (EV) has to be recharged within a given time window set by the EV owner). Therefore, each consumer has to choose at each time the consumption power so that the final accumulated energy reaches a desired level. The way in which the consumption power profiles are chosen is according to a sequence of functions (namely, a "strategy") mapping at any time the relevant information of a consumer (e.g., the current accumulated energy for EV-charging) to a suitable power consumption level. The design of such strategies may be centralized (in which there is a single decision-maker for consumers) or decentralized (in which there are several decision-makers, each of them representing a consumer). We analyze both scenarios by exploiting game theory and formal methods of computer science. More specifically, the power consumption scheduling problem can be modeled using stochastic games and Markov Decision Processes. For instance, probabilities provide a way to model the environment of the electrical system, namely: the noncontrollable part of the total consumption (e.g., the non-EV consumption). The controllable consumption can be adapted to the constraints of the Distribution Network (DN) (e.g., to the maximum shutdown temperature of the DN-transformer), and to their objectives (e.g., all EVs are charged). At first glance, this can be seen as a stochastic system with multi-constraints objectives. Therefor, the contributions of this thesis also concern the area of multi-criteria objective models, which allows one to pursue several objectives at the time such as having design which is functionally correct and robust against changes of the environment.

PUBLICATIONS

The work presented in this thesis is based on the following publications:

[M1] M. González, O. Beaude, P. Bouyer, S. Lasaulce, and N. Markey. "Stratégies d'ordonnancement de consommation d'énergie en présence d'information imparfaite de prévision." In: *Gretsi*, Sep. 2017.

[M2] M. González, P. Bouyer, S. Lasaulce, and N. Markey. "Optimisation en présence de contraintes en probabilité et processus markoviens contrôllés." In: *Gretsi*, Aug. 2019.

[M3] P. Bouyer, M. González, N. Markey, and M. Randour. "Multi-weighted Markov decision processes with reachability objectives." In: *Gandalf 2018 - Ninth International Symposium on Games, Automata, Logics, and Formal Verification*, Sep 2018, Sarrebruck, Germany. pp.250 - 264.

[M4] S. Lasaulce, O. Beaude, and M. González. "Charging games and algorithms." In: Data Analytics for Power Systems. Cambridge University Press, 2019.

[M5] M. González, O. Beaude, P. Bouyer, S. Lasaulce, and N. Markey. "Distributed Power Load Scheduling Under Nonflexible Load Uncertainty.", In: *IEEE Transaction on Smart Grid*, 2019. (Forthcoming paper).

[M6] M. González, P. Bouyer, S. Lasaulce, and N. Markey. "Optimization in the presence of probability constraints and Markov decision processes." 2019. (Forthcoming paper).

We have seen that computer programming is an art,
because it applies accumulated knowledge to the world,
because it requires skill and ingenuity, and especially
because it produces objects of beauty.

— **knuth:1974** [**knuth:1974**]

ACKNOWLEDGEMENTS

I would like to thank the ERC project EQualIS and RTE-CentraleSupélec Chair.

CONTENTS

LIST OF FIGURES

LIST OF ALGORITHMS

Part I

BACKGROUND

This part presents a brief introduction of the basic concepts of the research field and the general context of this work.

1

INTRODUCTION

Abstract:

We start this work with a brief introduction of some concepts of the research field. Afterwards, we introduce our motivation and inspiration of this work, which is founded mainly on formal methods and game theory to provide algorithmic solutions and guarantees for systems control designs and applications, as in the are of smart grids. We focus mainly on a problem of power consumption scheduling and on its formalism. We discuss the main direction of our work and we close the chapter by mentioning the organization and structure of this thesis.

1.1 INSERTION OF THE METHODOLOGIES

Quantitative games and Markov Decision Processes (MDPs) are standard models in the fields of **reactive systems** [63, 64, 95] and **formal methods** [43, 48, 61]. These are also widely studied in the field of the (modern) **game theory** [79, 84, 111]. First of all, we briefly revisit some "standard" definitions to present the general context of this work.

(i) **Reactive systems**: they are architectural and design pattern of computer systems that continuously interact with their environment. Their correctness is often critical for building more responsive and capable strategies (system controllers) of sustaining a safe behavior despite the potentially adversarial effects of the environment. They are significantly more tolerant of failure and when failure does occur, they are not in a disaster.

(ii) **Formal methods**: they are particular mathematical tools for the specification, development and verification for reactive systems to prove that the system follows a given specification modeling desired behaviors. Building appropriate strategies is not easy and classical development techniques based on testing are largely inadequate. Formal methods are essential to assert the correctness of strategies controlling the system.

(iv) **Game theory**: this is a mathematical modeling of strategic interaction among cooperating and/or competing players (agents), who play to the best of their decisions in order to satisfy similar, opposed or mixed interests, and objectives over the outcomes that may result from these decisions. One player's payoff is contingent on the strategy implemented by the other players.

1.1.1 *DECISION-MAKING*

Games and MDPs are two known methods that can be represented on directed graphs to model the decision-making of a system against the behavior of its environment. For instance, the latter can behave as adversary (purely antagonistic player(s)) and/or completely stochastic (without strategic interests in the outcomes, playing randomly). A correct behavior of a decision-maker is often computed as a "winning" strategy. Aiming to be a winning strategy is to enforce a given specification, encoded through an objective. An objective is a predefined set of paths of states of the system. A path either satisfies the property (we therefore say that it is winning) or it does not (i.e., it is losing). Specifying objectives for a model can be formulated in different ways. We focus mainly on this work in qualitative objectives (e.g., in probability, in almost-surely or sure mode), and in quantitative objectives (e.g., computing expected values). Objectives can also be seen as constraints in the model of the system to be satisfied when the decision-making is made. Briefly, we identify two ways of decision-making: centralized (in which there is a single decision-maker) and decentralized (in which there are several decision-makers).

1.1.2 *UNCERTAINTY*

A main goal in this work is to provide appropriate and efficient strategies, ensuring that these behave correctly and in an optimal way (or at least with good performances) under the uncertainties of the environment in the system [75]. In this work, the uncertainties are considered in two ways: it is assumed that the environment is either deterministic or stochastic. For example, think about the Electric Vehicles (EVs) charging (which is a controllable consumption) made at home or in a business park. We may have a forecast of the noncontrollable part of the total electricity consumption to schedule a suitable consumption/charging strategy for the EVs. Such a forecast could be available for a day and to be completely deterministic (a vector of parameters) or stochastic (a vector of random variables).

1.1.3 *ROBUSTNESS*

A robust strategy is one that its performance does not change "much" if it is applied to a system that works slightly different from the system model used for its synthesis. In this work, when it is referred to good robustness properties in the numerical applications of the proposed algorithms to schedule strategies, this is always done a posteriori, i.e., forecast errors on data are simulated, and strategies are scheduled over a noisy (deterministic or stochastic) scenario of the environment to observe after the robustness of the strategies over a perfect forecast, i.e., over the real environment that is a simple sequence of values. A deterministic forecast represents thus a noisy scenario (a simple vector) of the real environment, and the stochastic one represents several scenarios with associated probabilities of occurrence (a stochastic vector) of the real environment. If the strategy is robust against noise in the first case (which is less complex to build), such an approach is suitable. If not, the strategy must adapt to the possible random behaviors of the environment (which could be more complex to build).

1.1.4 *SYNTHESIS*

A main question is to decide whether a decision-maker has a (winning) strategy satisfying certain an objective, e.g., reaching a set of goal states. Synthesis is not only concerned with the existence of winning strategies, but also in their synthesis, with concerns related to the simplicity of the computed solution. We favor strategies that are considered simple (less complex) to implement and/or using a minimal amount of resources, notable the memory requirements over paths (also understood as scenarios or history of the system). Memory can be necessary to fulfill a (complex) objective. Hence, it is of interest to find strategies with perfect recall (or full memory), as they correspond to strategies that can be implemented in practice. Simpler strategies include for instance so-called memoryless (also called Markov) strategies, which are strategies that their decisions only based on the current state of the system, not on the full history. Those strategies are easier to implement as a real program; hence they will usually be preferred (provided they are powerful enough for the considered objectives). On the contrary, strategies with prohibitive (infinite) memory requirements are not implementable [29]. In this work, we provide algorithmic solutions for control design to assert the correctness of scheduling strategies when constraints/objectives are taken into account for the modeling. To design suitable and appropriate strategies, we also resort to tools from **control theory** [11], **measure theory** [52], **dynamic programming** [96], **optimization** [30] and **model checking** [41]. To prove that a system follows a given specification modeling desired behaviors, model checking is of particular interest to be detailed a little more.

1.1.5 *MODEL CHECKING*

From the introduction of model checking, see for example [42, 109], the verification of strategies controlling a system has been improved for their designs. Model checking (see also [16] for another book about this topic) has demonstrated to be relevant on the technology industry to verify desired behaviors on systems. This is a verification method that applies a posteriori to check that a preexisting formal model of a scheduling strategy satisfies a given specification. Of course, one would like to start from the specification of a model with all objectives and constraints, and synthesize a solution from it, in such a way that the desired properties of the system are satisfied. However, that is very ambitious and considerably harder to get when the model is stochastic and covers a priori many objectives, constraints, and so on. See for example [100]. In this work, we use model checking as verification of objectives/constraints of strategies in the modeling after these were built. This will be specified when it is necessary, e.g., when the modeling is relaxed to synthesize a strategy because of the complexity of the scheduling problem.

1.2 INSPIRATION AND INVENTIVENESS

Application domains like smart grids or energy networks [88] are becoming more and more widespread in our modern word. Their design requires a formal mathematical approach which will ensure that they behave correctly and in an optimal way, or at least with good performances (see for instance [65] for a logical analysis favoring the use of formal methods). While correctness and optimality are in the core of formal methods, those applications have only seldomly been attacked through formal models (see [32] for some approach). Among the possible reasons for this lack of communications between those scientific areas, may be the relative complexity of using formal methods. One of the principal motivations of this work is founded on formal methods to provide algorithmic solutions and guarantees for control designs and applications, as in smart grids.

1.2.1 *MOTIVATION*

The main inspiration of this work started from a smart grid problem, namely the **power consumption scheduling problem**. In plain terms, consider a scenario in which several consumption entities (households, electric appliances, EVs, etc.) generically called “consumers” have a certain energy demand and want to have this demand to be fulfilled before a set deadline (e.g., a simple instance of such a scenario is the case of a pool of EVs which have to recharge their battery to a given state of charge within a given time window set by the EV owner). Therefore, each consumer has to choose at each time the consumption power so that the accumulated energy reaches a desired level. This is the problem of power consumption scheduling (this is introduced in Chapter 2 in a basic setting). In this work, we focus mainly on building appropriate and efficient strategies for consumers (provided of existenceness) face to the uncertainties of the environment, namely the part of the total electricity consumption that is uncontrollable.

1.2.2 *ASPECTS AND FEATURES*

One important aspect to reduce costs of particular consumers and those related to the management of electrical systems, is how to make the relation between network operators and particular consumers to be more flexible. Reducing the impact of consumption operations on the distribution network is also an important issue that must be analyzed to guarantee the accuracy of the system. Notably, part of the total electricity consumption of particular consumers (e.g., in a district) can be adapted to the constraints of the electrical system (e.g, the maximum shutdown temperature of the distribution transformer), even adaptable to particular objectives and/or needs of the consumers (e.g., minimizing the total electrical consumption payment). In the whole of the potentially controllable electric consumption, EVs have a very favorable position according to its popularity and expected future development in the domain of smart grids, see for example [57]. They will be the representatives of the controllable electrical uses in the numerical applications in this work, wherein we examine the technical and economic interactions between EVs and the electricity distribution network. Some approximations have been made to reduce the impact of electricity consumption (see for example [17, 44, 72, 99]), but no formal methods have been provided to guarantee correctness and optimality, which must be in line with smart grid topics. This is the main objective of the Chapter 4 and Chapter 5.

1.2.3 *NUMERICAL ANALYSIS*

In the numerical applications, a forecast of the noncontrollable part of the total electricity consumption is available to compute strategies for consumers (these assumed to be EVs as we stated above). More precisely, the noncontrollable consumption is forecasted either by means of a deterministic function, or by random variables built from statistics. The main objective of the numerical applications in this work, is to show that the built strategies are adaptable to forecast errors (robustness accuracy). This is discussed mainly in the Chapter 3 and Chapter 6, in which resp. the centralized and decentralized decision-making are demanded. In such chapters it is also presented the general mathematical modeling for a deterministic and stochastic points of view.

1.3 ORGANIZATION AND STRUCTURE

This thesis is divided into five parts with several chapters. We have tried to draft each chapter so that each one can be read independently. The structure of this work is the following:

Part I : **BACKGROUND**

Following this introductory Chapter 1 is Chapter 2. In which the background of the chosen application framework and the research tools are presented. The main (power consumption scheduling) problem of this work is introduced in a basic setting and the background of Markov processes and formal methods take place.

Part II : **CENTRALIZED MODELING**

Here, we discuss the centralized approach for the power consumption scheduling problem, in which a single entity controls and builds the strategies for the consumers. This part is divided into three chapters. These begin with Chapter 3 that shows the modeling and the different approaches to solve the (centralized) problem, in which these depend on whether the forecast of the environment is deterministic or stochastic. Chapter 4 covers the research related to the existence of strategies for the general multi-constrained problem of this work under a stochastic setting of the environment. The algorithmic approach is based on two sequences of optimization problems, in which their solutions are assumed under randomized strategies. In Chapter 5, the existence and synthesis of a solution for the optimization problems is proved under mixed strategies. So that, the synthesis of randomized strategy solution of the multi-constrained problem comes naturally by using tools of game theory.

Part III : **DECENTRALIZED MODELING**

This discusses the decentralized modeling for the power consumption scheduling problem, in which each consumer builds his own strategy. This part contains one chapter. This is the Chapter 6, which describes the modeling and approaches to solve the (decentralized) problem, in which these depend on whether the forecast of the environment is deterministic or stochastic. The main idea is to use a technique so-called sequential best response dynamic between the consumers, in which convergence of the method can be shown for the deterministic case, and only a numerical approach is developed in the stochastic case due to its complexity.

Part IV : **CONCLUSION**

Within this part, the summary of this work is covered in the Chapter 7, which includes conclusions and suggestions for related future works (several ideas, computational challenges, approaches, etc.).

Part V : **APPENDIX**

This covers the Appendix B wherein the mathematical proofs of our research are drafted.

2

RESEARCH BACKGROUND

Abstract:

This chapter first places the work of this thesis on the subject of smart grids. It specifies the chosen application framework oriented towards the load flexibility of controllable consumption entities (consumers) to limit their impact on a distribution network. Second, after a basic problem of consumption scheduling is presented, the main tools used in this work are introduced. In particular, the consumers are faced to an antagonistic adversary (the noncontrollable part of the total load consumption) that can be deterministic or stochastic. The latter opens the possibility to use Markov decision processes, wherein fixing the controllable consumption of consumers according to well-chosen scheduling strategies, the model becomes purely stochastic (Markov chain). Strategies are analyzed with regard to several criteria such as their expected cost, their memory requirements and so on. We illustrate many usual objectives and notably discuss reachability and shortest path objectives.

LIST OF ABBREVIATIONS AND SYMBOLS

DN	Distribution Network	HS	Hot-Spot (temperature)
EVs	Electric Vehicle(s)	PaC	Plug-and-Charge
SNR	Signal-to-Noise Ratio	SSP	Stochastic Shortest Path
MC	Markov Chain	MDP	Markov Decision Process
h	hour	min	minutes
kW	Kilowatt	kWh	Kilowatt-hour
TWh	Terawatt hour	dB	decibel
$°C$	Degree Celsius		

$\mathcal{A}$	finite action space in $\mathcal{M}$
$\mathcal{B}(\Omega_{x_1})$	Borel sigma-algebra over Ω_{x_1}
$\mathcal{C}$	MC model
C_{x_1}	cost function under an initial condition x_1
Cyl	cylinder set generating $\mathcal{B}(\Omega_{x_1})$
δ	randomized scheduling strategy
Δ	set of randomized strategies
$\Delta[\Pi]$	set of mixed strategies over Π
$\underline{\Delta}$	set of randomized Markov strategies
$\mathcal{D}(\mathcal{S})$	set of probability distributions over a finite set $\mathcal{S}$
Δ_t	time-step duration of each time-slot
e_i	energy demand of i
$\mathbb{E}$	expectation operator
f	evolution law of the system state
$\mathcal{G}$	set of goal states of a system
i	controllable electric device or consumer
I	number of controllable electric devices or consumers
$\mathcal{I}$	set of controllable electric devices or consumers
$\ell_i^{\min}$	minimum power of i
$\ell_i^{\max}$	maximal power of i
$\ell_{i,t}$	controllable load of i at t
$L_{i,t}$	function representing $\ell_{i,t}$
$\boldsymbol{\ell}_i$	controllable load profile of i of length T
$\boldsymbol{L}_i$	function representing $\boldsymbol{\ell}_i$
$\boldsymbol{\ell}_t$	controllable load vector at t of length I
$\boldsymbol{L}_t$	function representing $\boldsymbol{\ell}_t$ at t
$\ell_{0,t}$	real noncontrollable load at t
$\boldsymbol{\ell}_0$	real noncontrollable load profile of length T

$\widetilde{\ell}_{0,t}$	deterministic forecast of $\ell_{0,t}$ at t
$\widetilde{\boldsymbol{\ell}}_0$	deterministic forecast profile of length T
$\widetilde{L}_{0,t}$	stochastic forecast of $\ell_{0,t}$ at t
$\widetilde{\boldsymbol{L}}_0$	stochastic forecast profile of length T
$\ell^{\max}$	maximal power of the DN-transformer
ℓ_t	total load consumption at t
$\widetilde{L}_t$	function representing the total load under $\widetilde{\boldsymbol{L}}_0$ at t
$\mathcal{L}$	total load consumption space
$\mathcal{M}$	MDP model
$\mathcal{M}_T$	unfolding of $\mathcal{M}$ with depth T
$\mathcal{N}(\mu, \sigma^2)$	Gaussian (normal) distribution with mean μ and variance σ^2
ω	history or path of a system
ω_t	history or path of length t of the system
Ω_{x_1}	set of histories or paths of a system from an initial state x_1
proj_j	projection function on the j-component of a sequence
π	pure scheduling strategy
Π	set of pure strategies
$\underline{\Pi}$	set of pure Markov strategies
P	transition probability between the states of a system
P_T	transition probability between states in $\mathcal{M}_T$
P_0	probability distribution of $\widetilde{L}_{0,t}$
$\mathbb{P}$	probability measure on $\mathcal{B}(\Omega_{x_1})$
σ	mixed scheduling strategy
s_t	state in $\mathcal{S}_T$ at t
$\mathcal{S}_T$	augmented space of states in $\mathcal{M}_T$
t	time-slot
T	finite horizon time-slot
$T_{\mathcal{G}}$	random reachability time to $\mathcal{G}$
$\mathcal{T}$	set of time-slots
TS	truncated sum function
$\mathbb{TS}$	expected truncated sum
$x^{\max}$	upper bound of the system state values
x_t	system state at t
$\widetilde{x}_t$	system state under $\widetilde{\ell}_{0,t}$ at t
$\widetilde{X}_t$	function representing the system state under $\widetilde{\boldsymbol{L}}_0$ at t
$\mathcal{X}$	finite set of system states
$\underline{\Omega}_{x_1}$	set of finite histories or paths of a system from an initial state x_1

2.1 SMART GRIDS

Being guided by [17, 73], let's start with the seven goals set by the U.S. Department of Energy [1] to go to the electrical grids more "intelligent":

(i) **Improve information to consumers**: this point will be discussed indirectly in this work. The coordination of the consumption, e.g., the one over a set of Electric Vehicles (EVs), will be studied from an algorithmic point of view. A signal is in fact sent by a central operator (which is the most common scenario) from the electrical Distribution Network (DN) to the consumers. This signal is partly intended to give information about the state of the electrical grid to the consumers so that they can schedule their controllable consumption decisions by integrating the potential impact on the electrical system, namely the DN.

(ii) **Integrate new production and storage opportunities**: this will not be addressed in this document. In this work, each consumer is seen as a load consumption on the DN. For instance, the storage to be able to transport energy (e.g., from workplace to home) will not be used here. This choice results in more constraints in the architecture of the model, making the DN more complex. See [92] for an example of "re-injection to buildings" made by EVs-to-Building.

(iii) **Integrate new products, services and markets**: this is largely evoked in this work. This will be reflected on the coordination of the consumers, which leads to necessarily define new services.

(iv) **Improve the quality of electric power**: although it could be a physical metric to inject into the proposed algorithms in this work, it will not be treated here. The main reason is that it often requires to use more complex functions to reflect the impacts generated on the DN. These are not easy to integrate into the models that often take, at least initially, approximate metrics to show their contribution. See for example [59] in which the aging metric of the DN-transformer is finely studied by integrating, e.g., electric current.

(v) **Efficient management and optimization of electrical equipment**: this is directly related to the algorithms proposed in this work to compute suitable consumption scheduling strategies. The technical and economical models used in the numerical applications in this document, allow to carry out the scheduling that takes into account not only the preferences of the consumers, but also the impacts on the DN. For instance, the recharge of EVs is used efficiently, both from the point of view of the consumers and the operator of the DN.

(vi) **Move towards an independent management of incidents**: this is not studied in this work. A priori, it involves placing at much shorter time scales than those studied in this document. Such a point of view must be designed to be fast. It may not open the door to an "iterative communication" between the network operator and the consumers (as is the case in this work).

(vii) **Be flexible face to the randomness**: this is the kernel of this work to evaluate the performance of the proposed algorithms in the context where forecast errors are made on the model. This is opposite to the works reported in [17, 19, 20, 56, 106] for example, in which all simulations have been performed over a noisy setting. More precisely, in those early works, forecast errors over the non-controllable part of the total consumption are ignored, and all their work is based on a deterministic forecast (even under the ambitious hypothesis of perfect forecast). The main algorithms proposed in our work are performed on a stochastic forecast, allowing us to build consumption scheduling strategies that are robust to noises of forecast. Moreover, formal methods developed here take as input any stochastic model of the environment, providing existence and synthesis of scheduling strategies.

EVs currently play the role of the controllable consumption whose impact on the distribution networks, markets and electrical systems in general must be analyzed, according to its popularity and expected future development in the domain of smart grids. They will be the representatives of the controllable electrical uses in the numerical applications in this work.

2.1.1 *ELECTRIC VEHICLES*

Recent numbers indicate a significant increase in registrations of EVs. For example, the worldwide number of EVs was more than 3 million in 2018, see Figure 2.1. Statistics also show a significant increase for 2030 in the world fleet of EVs: according to the "New Policies" scenario of the AIE [7] (based on current or announced policies), the world fleet of EVs could be around 125 million units in 2030 (and up to 220 million in the scenario "EV30@30" aiming to increase the market share of electric mobility worldwide to 30%). In addition, the worldwide electricity consumption from EVs will reach 404 TWh in the "New Policies' scenario and 928 TWh in the "EV30@30" scenario. These values represent, respectively, a 7-fold and 17-fold increase when compared with the electricity consumed by EVs in 2017. It is therefore essential to anticipate the issue of charging/consuming, well before the peak of EVs development.

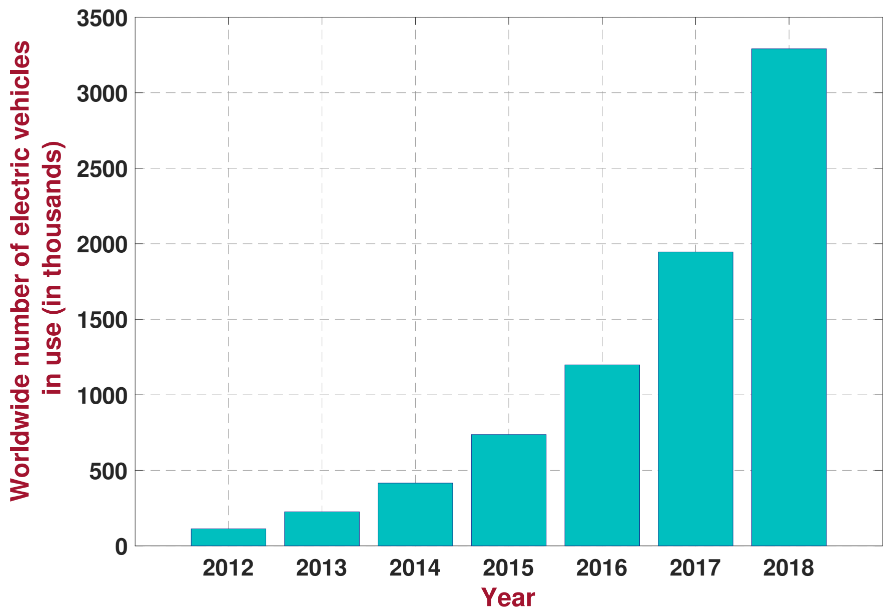


Figure 2.1: Worldwide number of EVs from 2012 to 2018. There were some 3 million EVs in use globally in 2018. Source [2, 7].

CHARGING POWER

An important point is the mode of charging/consuming of EVs. This can be summarized as normal, fast or accelerated mode. The latter two have typically high costs (even prohibitive for quick drop) for a very low use. The charging mode that is predominant, is the normal one (mostly made at home and business parks, with an acceptable cost), that is usually to be 3 kW of power and implemented in the night. This is the main charging mode that we use in the numerical applications in this work, but the results of this thesis remain invariant under any charging mode.

In summary, the significant development of EVs must be accompanied by reflecting on the ecosystem. The charging infrastructure is in particular a point of interest. In this work, we focus in particular on the interaction between EVs and the electrical system (the DN) when the vehicles are charging/consuming. The objective is to propose and explore methods to build suitable consumption scheduling strategies, that can be applied to the EVs domain to limit the impact of consuming on the DN. This is part of discussions on smart grids [69, 98], which is intended to increase the flexibility and the "intelligence" or electrical networks, in particular by strengthening the link between DN-operators and consumers (e.g., EVs). For all these reasons, EVs are good actors to use in the numerical applications in this work.

2.2 AIM OF MODELLING POWER CONSUMPTIONS

In almost all the contributions of the literature on Smart grids, an intermediate step to modeling problems of power consumption scheduling is the calculation of the total electricity consumption profile of a district. In this work, we consider a finite set $\mathcal{I}$ representing several controllable consumption entities (also called controllable electric devices or simply consumers here). The **total electricity consumption** at time t can therefore be written of the form:

$$\underbrace{\ell_t}_{\substack{\text{total electricity}\\ \text{consumption}}} = \underbrace{\ell_{0,t}}_{\substack{\text{noncontrollable}\\ \text{electricity consumption}}} + \underbrace{\sum_{i\in\mathcal{I}} \ell_{i,t}}_{\substack{\text{sum of controllable}\\ \text{electricity consumption}}} , \tag{2.1}$$

where the first term on the right side, $\ell_{0,t}$, represents the electricity consumption in the district that excludes the consumption of the controllable electric devices at time t. This value contains the aggregated consumption of all traditional electrical uses, e.g., heating, lighting, cooking, etc. This consumption is supposed to be **noncontrollable**, i.e., it does not adapt to the impacts measured on the electric DN (e.g., the degradation of the DN in terms of the DN-transformer lifetime, the electrical consumption payment due to the total electricity consumption, etc.). This amount at each time t can be forecasted deterministically or stochastically. The two forms presented remains valid and are taken into account in this work. We will specify through this work if we consider the modeling based on a **deterministic forecast**, i.e., when the noncontrollable consumption is forecasted by means of a deterministic function (and then $\ell_{0,t}$ boils down to a simple parameter denoted by $\widetilde{\ell}_{0,t}$), or when this is forecasted as a random variable (and then $\ell_{0,t}$ is statistically estimated using past databases, that we denote by $\widetilde{L}_{0,t}$).

The second term on the right side in the eq. (2.1), $\sum_{i\in\mathcal{I}} \ell_{i,t}$, is the accumulated consumption due to the consumers of the district, where each $i \in \mathcal{I}$ consumes $\ell_{i,t}$ at time t and can be scheduled within a time interval with a fixed time horizon. This component is said **controllable** since each one is, and is adaptable to the constraints of a scheduling problem, i.e., to the impacts that can be measured on the DN. Controllable devices (consumers) that can be considered in this work could include, e.g., EVs, dishwasher, water-heaters, etc., which can be scheduled to reach a given (cumulative) energy demand to complete a corresponding task, e.g., charging EVs, arranging a dishwasher program, etc. In this work, we refer in general to controllable electric devices or consumers, but we will refer to EVs when it is necessary to clarify the presentation and the numerical results.

Note that the total electricity consumption (2.1) can be taken into account as a current, a power or an energy; depending on the physical models used. The results presented in this work are unchanged, and only the functions expressing the physical impacts on the DN will have to be adapted. Subsequently, we will usually refer to power loads or simply as loads.

Independently of the knowledge of the noncontrollable part of the total load consumption (2.1), which can be deterministic or stochastic, the DN experiences an increased amount of variable loads depending on the consumption of the consumers. In particular, the way in which the controllable loads are scheduled, has an impact on the total load consumption, introducing new load variations and, possibly causing overloading on the DN [72]. For example, depending on the features of the load strategies for a large number of EVs, could result in energy losses [20], in increasing of the DN-transformer aging [19][M1]. Such load consumption variations depend on when the controllable devices are available to schedule, at which power loads, the energy demanded to complete the tasks, among others. For all these reasons, we aim to construct suitable scheduling load strategies for each controllable electric device (consumer), minimizing their resulting impact on the DN.

2.3 CONTROLLABLE LOAD CONSUMPTION

Standard electrical load consumption typically starts as soon as an electric device is plugged, e.g., when an EV is plugged into the grid to charge its battery, and typically the load consumption is carried out at the maximum admissible power by the charging system. The credit of this very simple scheduling method, so called **Plug-and-Charge** (PaC), is that it does not require any interaction between the users and the electric grid, and minimizes the time needed to reach a given cumulative energy to complete an associate task, e.g., recharging the battery of an EV. On the other hand, such a method has the disadvantage of potentially strongly impacting the grid, since it ignores the total load consumption demand associated with all the other electric devices. For instance, if a large number of EVs are most usually charged in a grid by the PaC scheduling method, the total load consumption will be amplified, generating demand peaks and potentially causing that the DN-transformer reaches its (maximum) shutdown temperature, see Figure 2.2 for instance. In addition, in scenarios where time-differentiated electrical consumption payment are implemented, the monetary cost paid by the users can also increase [86].

One of the goals of the work reported in this thesis is to provide new load scheduling strategies whose main purpose is to minimize the impact of the power loads of the controllable electric devices on a precise part of the electric grid namely: the DN. For example, minimizing the degradation of the DN in terms of the DN-transformer aging, the electrical payment due to the total load consumption, etc. The derived results can be re-exploited for other problems in smart grids such as problems of load scheduling with delay constraints [28, 107]; therein the system state is given by the available stored energy and the cost are market-based or generation ones. The scheduling problem studied in this work is based mainly on the knowledge of the noncontrollable part of the total load consumption (2.1), which can be given by a deterministic or stochastic forecast.

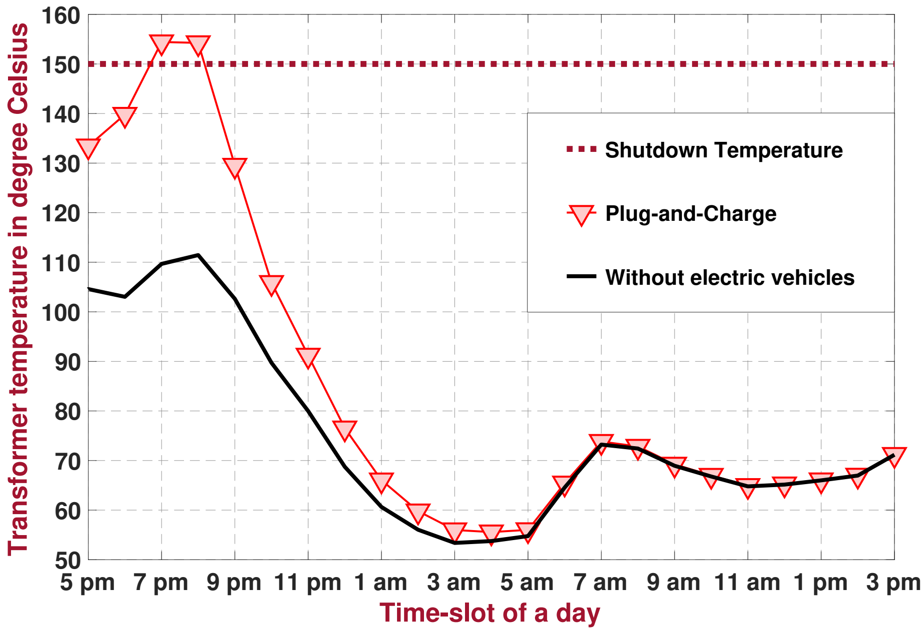


Figure 2.2: Impact of the charging scheme Plug-and-Charge (PaC) on the evolution of the DN-transformer Hot-Spot (HS) temperature, under the assumption of $I = 10$ Electric Vehicles (EVs). In this case, the shutdown temperature is exceeded for such a method of charging. (the HS model used was the one of [19] and the real data was taken from [3]).

2.4 NONCONTROLLABLE LOAD UNCERTAINTY

Besides the controllable load consumption, there is a noncontrollable part in the total load consumption (2.1) that is uncertain, i.e., it is not perfectly known when the controllable loads are being scheduled. In this work, the **real noncontrollable load profile**:

$$\boldsymbol{\ell}_0 \quad := \quad (\ell_{0,t})_{t\in\mathcal{T}} \ , \tag{2.2}$$

is represented by a given **forecast** (or prediction) on a finite time set $\mathcal{T}$, to schedule power consumption strategies. Here, we consider two approaches of how the real noncontrollable loads are forecasted: a deterministic approach and a stochastic one.

2.4.1 DETERMINISTIC FORECAST APPROACH

When the forecast of the noncontrollable load consumption is assumed to be deterministic, one can express the profile $\boldsymbol{\ell}_0$ as a simple vector of parameters on the time set $\mathcal{T}$, which is used to schedule controllable loads by means of strategies. To assess the impact of not being able to forecast the noncontrollable load consumption perfectly, we assume that the scheduling problem is fed in the numerical applications with a **noisy deterministic forecast**, expressed as:

$$\widetilde{\boldsymbol{\ell}}_0 \quad := \quad (\widetilde{\ell}_{0,t})_{t\in\mathcal{T}} \ , \tag{2.3}$$

where each parameter $\widetilde{\ell}_{0,t}$ at time t can be of the form:

$$\widetilde{\ell}_{0,t} \quad = \quad \ell_{0,t} \ + \ z_t \ , \tag{2.4}$$

where, for each t fixed we take a single value $z_t \sim \mathcal{N}(0, \sigma^2_{\mathcal{T}})$, which represents a noise of forecast. The variance $\sigma^2_{\mathcal{T}}$ can be computed on a day if $\mathcal{T}$ represents a day, e.g., by means of Signal-to-Noise Ratio[1] (SNR) expressed in decibel (dB) (see for instance [18]), which allows one to measure to what extent the noncontrollable load consumption can be forecasted (details about such a measure will be shown in the numerical analysis). Figure 2.3 shows several deterministic forecast of the real noncontrollable load consumption on time-slots t of 30 min for under different noises expressed in decibel (dB).

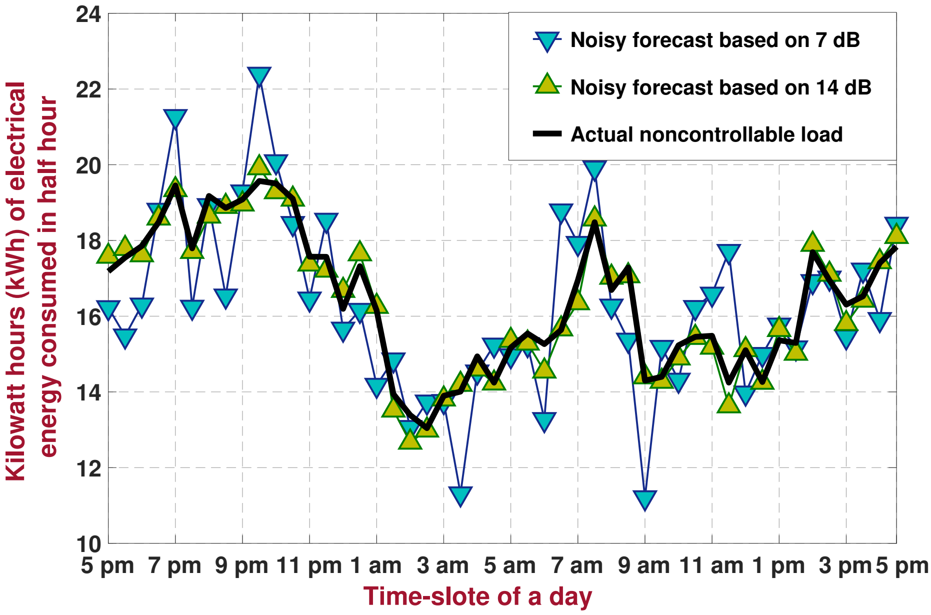


Figure 2.3: Real noncontrollable energy consumption $\Delta_t \boldsymbol{\ell}_0$ (black line), where $\Delta_t = 0.5$ represents an half hour. Two deterministic forecast (obtained with a sampling noise) are plotted over this figure, built by the model (2.4): the first one is based on a SNR = 7 dB and the second one is based on a SNR = 14 dB. Real data was extracted from [3].

1 This will be explained a little further in the numerical applications.

2.4.2 *STOCHASTIC FORECAST APPROACH*

Besides the previous Section 2.4.1, the forecast on the real noncontrollable load consumption can be assumed to be stochastic. This can be understood as many scenarios with respective probability of occurrence. In this work, a **stochastic forecast** of the noncontrollable load profile (2.2) is supposed to be represented by the following random vector:

$$\widetilde{\boldsymbol{L}}_0 \;:=\; (\widetilde{L}_{0,t})_{t\in\mathcal{T}}\ , \tag{2.5}$$

which is a finite collection of i.i.d. random variables[2]. Following the model (2.4), we can express for instance the following stochastic forecast of the real noncontrollable load $\ell_{0,t}$ at time t as following:

$$\widetilde{L}_{0,t} \;=\; L_{0,t} \;+\; Z_t\ , \tag{2.6}$$

where $(L_{0,t})_{t\in\mathcal{T}}$ is the sequence taking the real noncontrollable load profile (2.2), and Z_t be a random noise independent of $L_{0,t}$. For example, as in (2.4), for each t fixed we take a $Z_t \sim \mathcal{N}(0,\, \sigma^2_{\mathcal{T}})$, where the variance $\sigma^2_{\mathcal{T}}$ can be computed on a day if $\mathcal{T}$ represents a day, e.g., by means of Signal-to-Noise Ratio(SNR) expressed in decibel (dB). The Figure 2.4 shows a stochastic forecast over the real noncontrollable loads on time, which can be seen as several deterministic forecast with associated probabilities of occurrence.

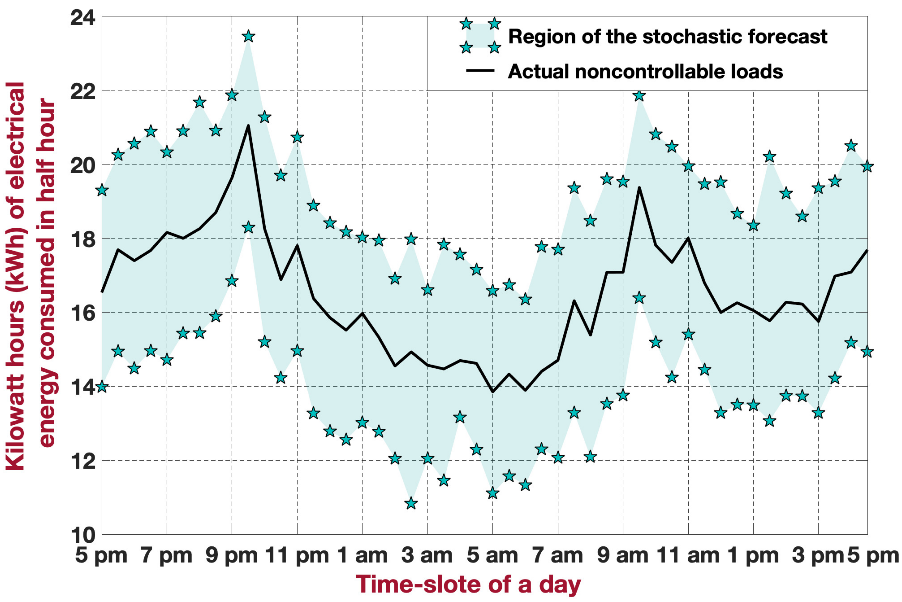


Figure 2.4: Real noncontrollable energy consumption $\Delta_t\,\boldsymbol{\ell}_0$ (black line), where $\Delta_t = 0.5$ represents an half hour. The complete (continuous) region of a stochastic forecast is plotted over this figure. Real data was extracted from [3] to build the forecast, which is based on the model (2.6) with a $\mathrm{SNR} = 7\,\mathrm{dB}$.

2 Which can be alternatively defined on a common probability space, let say $(\Omega, \mathcal{F}, P)$, or by means of a probability distribution P_0 without explicit reference of the underlying probability space. Either the original probability measure P or the induced distribution P_0 can be used to compute probabilities of the events involving $\widetilde{\boldsymbol{L}}_0$.

Based on the forecast of the noncontrollable load consumption, the state of the system can be obtained by means of a dynamic evolution in function of the (controllable and noncontrollable) load consumption. The basic scheduling problem of power consumption is therefore presented in next.

2.5 THE BASIC SCHEDULING PROBLEM

The basic problem that we aim to study through this document is introduced in this section. Firstly, we consider a scenario in which $I \in \mathbb{N}$ controllable electric devices (or consumers) have a certain energy demand and want to have this demand to be fulfilled before a set deadline $T \in \mathbb{N}$. A simple instance of such a scenario is the case of a pool of EVs which have to recharge their battery to a given state of charge within a given time window set by the EV owner. Thus, each consumer $i \in \mathcal{I} := \{1, ..., I\}$ has to choose at each time $t \in \mathcal{T} := \{1, ..., T\}$ the consumption power $\ell_{i,t} \in \mathbb{R}_0^+$ so that the accumulated energy reaches a desired level $e_i \in \mathbb{R}^+$. This is the problem of power consumption scheduling that we aim to study and exploit throughout this work. The basic form of such a problem is presented as follows:

Basic Power Consumption Scheduling Problem

$$\min_{\boldsymbol{\ell}} \quad \sum_{t \in \mathcal{T}} C_{x_1}(x_t, \boldsymbol{\ell}_t, \ell_{0,t}) \tag{2.7}$$

$$\text{s.t.} \quad x_{t+1} = f(x_t, \boldsymbol{\ell}_t, \ell_{0,t}) \tag{2.8}$$

$$\Delta_t \sum_{t \in \mathcal{T}} \ell_{i,t} \geq e_i \tag{2.9}$$

where $x_1 \in \mathbb{R}^+$ is an initial condition for the system state value (namely, a precise metric of the DN). More about such a variable and the constraint are discussed in the Section 2.5.2.

2.5.1 *CENTRALIZED AND DECENTRALIZED SCHEDULING*

The controllable variable $\boldsymbol{\ell}$ represents the consumption power of consumers $i = 1, ..., I$ over time $t = 1, ..., T$. This can be seen as the following matrix:

$$\boldsymbol{\ell} = \begin{bmatrix} \ell_{1,1} & \ell_{1,2} & \cdots & \ell_{1,T} \\ \ell_{2,1} & \ell_{2,2} & \cdots & \ell_{2,T} \\ \vdots & \vdots & \ddots & \vdots \\ \ell_{I,1} & \ell_{I,2} & \cdots & \ell_{I,T} \end{bmatrix},$$

which can be either controlled by a single entity (centralized form), or controlled separately by I decision-makers (decentralized form). In the first case, there is a single decision-maker that controls in each instant $t = 1, ..., T$ the **controllable load vector**:

$$\boldsymbol{\ell}_t \quad := \quad (\ell_{i,t})_{i \in \mathcal{I}} \ ,$$

i.e., all the load variable $\boldsymbol{\ell}$. In the second case, there are several decision-makers, where each one represents a consumer i and then, the decentralized form can be seen as each one controls its respective column i of the matrix $\boldsymbol{\ell}$, i.e., controlling its **controllable load profile**:

$$\boldsymbol{\ell}_i \quad := \quad (\ell_{i,t})_{t \in \mathcal{T}} \ .$$

Details about the two scheduling architectures (centralized and decentralized) is explained in the following.

CENTRALIZED AND DECENTRALIZED APPROACHES

In order to solve the scheduling problem (2.7), the controllable loads need to be coordinated. This coordination is done through each load consumption profile of each controllable electric device (consumer). We identify two principal approaches of communication and control that can be performed between them, see for instance [99]. These two categories refer to a decentralized and centralized scheduling architectures, which relate to the level on which the controllable loads are scheduled, given an objective function and constraints that need to be met given a certain scenario (that in this work it can be deterministic or stochastic due to the noncontrollable load consumption forecast).

First, we gain reliability in the controllable loads under a **centralized scheduling**, because all the information of the load consumption of the controllable electric devices is available to scheduling. However, since such approach controls jointly the vectors $\boldsymbol{\ell}_t$ for each $t = 1, ..., T$ (i.e., the complete variable $\boldsymbol{\ell}$), it requires a high degree of information in order to get precise scheduling when T or I is large, and also, it does not preserve the privacy of the controllable load consumption of the consumers. For this reason, a centralized architecture increases the complexity of scheduling [82]. In order to address such complexity, a decentralized architecture is often more appropriate.

Second, when a **decentralized scheduling** architecture is used, the set $\mathcal{I}$ is considered as a set of different decision-makers, each one of them representing a consumer $i \in \mathcal{I}$. This approach allows each i to choose its load profile $\boldsymbol{\ell}_i$ according to the objective function of the problem to arrange the load consumption process. Thus, the information between the decision-makers (also refereed as consumers) has to be coordinated and incorporated in the scheduling problem in order to find effective scheduling load profiles while the constraints are satisfied, in particular satisfying the required energy (2.9) of each i. Although a decentralized approach requires more exchange of information, this can be made iteratively between the consumers and the size of the scheduling problem is confined to one unit in each iteration, i.e., to find a load profile $\boldsymbol{\ell}_i$ for i while the consumption of the other consumers is fixed.

2.5.2 *VARIABLES AND CONSTRAINTS*

In the basic problem (2.7), the variable x_t represents a certain metrics of interest of the system at t (namely the DN), so called the **system state**, e.g., this can represent the total electricity bill corresponding to the total load consumption, the the hot-spot DN-transformer temperature, etc. This metric could be either deterministic or stochastic depending of the noncontrollable load knowledge (the deterministic forecast (2.3) or the stochastic one (2.5) resp.). In addition, the system state is assumed to evolve at each time with a function f representing a dynamic law, see eq. (2.8). An important assumption to make an effective calculation with the proposed algorithms in this work, is the one shown in the Assumption 2.5.1.

Assumption 2.5.1

The function f is a function that depends of the total load consumption, i.e., of the form: $f'(x_t, \ell_t)$, with ℓ_t of eq. (2.1).

Explicit expressions of f are developed in the numerical applications, mainly in the Section 3.5 and Section 6.5.

Example 2.5.2. *Let us now consider an example where the state of the system represents the hot-spot temperature of a DN-transformer. The system state is assumed to obey the following dynamic law [M1] [68, 70], which has a single period time-lag in the load consumption:*

$$x_{t+1} \;=\; \alpha\, x_t \;+\; \beta\, \ell_t^2 \;+\; \gamma\, \ell_{t-1}^2 \;+\; z_t\;,$$

where $\alpha, \beta, \gamma \in \mathbb{R}$ are constants of the model [70], and $z_t \in \mathbb{R}_0^+$ is a known deterministic function (it typically represents the ambient temperature in Celsius degrees). The initial conditions of the model are assumed to be given.

On the other hand, the constraint (2.9) expresses the **energy demand** $e_i \in \mathbb{R}^+$ to complete a corresponding task of a controllable electric devise (consumer) i, e.g., an energy required can be $e_i = 3.6\,\mathrm{kWh}$ for a full-load dishwasher cycle, $e_i = 24\,\mathrm{kWh}$ for an EV to recharge its battery, etc. In the eq. (2.9), Δ_t represents the time-step duration, e.g., to be $\Delta_t = 0.5$ if each t represents a time-slot of $30\,\mathrm{min}$.

VARIABLES UNDER A NONCONTROLLABLE LOAD FORECAST

Under a deterministic (2.3) or stochastic (2.6) forecast of the noncontrollable load consumption, the total load (2.1) and the system state (which evolves with the dynamic law (2.8)) are evidently affected. When a deterministic or stochastic case is assumed, we implement different notations. This is summarized in the following table, which shows how we write the main variables considered in this document.

Mode	Deterministic Forecast	Stochastic Forecast
Noncontrollable Load	$\widetilde{\ell}_{0,t}$	$\widetilde{L}_{0,t}$
Controllable Load	$\ell_{i,t}$	$L_{i,t}$
Total Load	$\widetilde{\ell}_t$	$\widetilde{L}_t$
System State	$\widetilde{x}_t$	$\widetilde{X}_t$

ADDITIONAL CONSTRAINTS

The constraints in this work, are also referred as "objectives" to satisfy. Some constraints concerning the DN that are not been presented in the basic problem (2.7), but that will be studied in this work, are the following:

(*i*) The system state x_t must be bounded by an **upper safety limit**, denoted by $x^{\max} \in \mathbb{R}^+$, so that the system operates in safe conditions. Mathematically, this is expressed for each $t \in \mathcal{T}$ by:

$$x_t \leq x^{\max} . \tag{2.10}$$

For example, if the state of the system represents the hot-spot DN-transformer temperature (see Example 2.5.2 for instance), exceeding a high temperature $x^{\max}$ (e.g., $x^{\max} = 150\,°\mathrm{C}$) might put the transformer in dangerous conditions of shut down.

(*ii*) Concerning the DN-transformer, there is a **maximal power** admissible. It is for this reason that we want to keep the total load consumption (2.1) upper bounded by a certain value $\ell^{\max} \in \mathbb{R}_+$. This can be expressed by:

$$\ell_t \leq \ell^{\max} . \tag{2.11}$$

For example, the maximal power of a typical DN-transformer in a urban district is $\ell^{\max} = 90\,\mathrm{kW}$.

(*iii*) An additional constraint is considered for each consumer i when he is charging to satisfy the constraint of energy demand (2.9). The power load $\ell_{i,t}$ at time t is assumed to be at least a minimum power $\ell_i^{\min} \in \mathbb{R}_0^+$ and also, it cannot exceed the maximum power at which the consumer i can be charging. Mathematically, this is expressed by:

The minimum power refers when, e.g., i is switched off but is designed to draw some load in standby mode.

$$\ell_i^{\min} \leq \ell_{i,t} \leq \ell_i^{\max} .$$

For example, dishwashers may use $\ell_i^{\max} = 1.8\,\mathrm{kW}$ of maximum power, an EV may be charged at $\ell_i^{\max} = 3\,\mathrm{kW}$ in a slow charging mode (typically at home).

Due to the available knowledge about the noncontrollable load consumption (2.2) to schedule the controllable loads of the consumers, i.e., the forecast that can be deterministic (2.3) or stochastic (2.6), all the constraints (and even more, the scheduling problem in question) that involve the stochastic case, become stochastic as well. When the forecast is deterministic, the noncontrollable load at each time boil down to a simple parameter and then, the (deterministic) constraints (without going into details of how to satisfy them) are easier to express, e.g., see the ones in (2.10) and (2.11). On the contrary, when the constraints become stochastic, these can be formulated in different ways [10, 71], e.g., in expectation, in almost surely mode, in probability, etc. We explain more about that in the Section 2.6.5.

2.5.3 *COST FUNCTION*

In the basic problem (2.7), the impact of the charging operation on the DN of the different consumers, is measured by an objective function, which is expressed as a sum of a cost function C. At each instant t, each controllable load $\ell_{1,t}, ..., \ell_{I,t}$ has an effect of increasing the load consumption, causing a cost that is measured by $C(x_t, \boldsymbol{\ell}_t, \ell_{0,t}) \in \mathbb{R}$. The latter amount takes into account the state of the system x_t, the load vector $\boldsymbol{\ell}_t = (\ell_{1,t}, ..., \ell_{I,t})$, and the noncontrollable load consumption $\ell_{0,t}$. Such a cost function can represent, e.g., the degradation of the DN in terms of the DN-transformer lifetime, the Joule losses, the electrical consumption payment due to the load operations, among other. Based on a (deterministic or stochastic) forecast of the noncontrollable load consumption, the cost function C is expressed in this work either a deterministic function, or an expectation, which also depends on the way in which the controllable loads are managed (in a centralized or decentralized) way. We summarize how we write the cost incurred at each time t in the following table:

Mode	**Centralized**	**Decentralized**
Deterministic	$C(\widetilde{x}_t, \boldsymbol{\ell}_t ; \widetilde{\ell}_{0,t})$	$C(\widetilde{x}_t, \ell_{1,t}, ..., \ell_{I,t} ; \widetilde{\ell}_{0,t})$
Stochastic	$\mathbb{E}\Big[C(\widetilde{X}_t, \boldsymbol{L}_t; \widetilde{L}_{0,t})\Big]$	$\mathbb{E}\Big[C(\widetilde{X}_t, L_{1,t}, ..., L_{I,t}; \widetilde{L}_{0,t})\Big]$

A suitable method to model the scheduling problems studied in this work when a stochastic forecast on the noncontrollable part of the total load consumption is assumed, is the so-called Markov decision process [96]. To go to such a model, we revisit some research background in the next section.

2.6 BACKGROUND OF MARKOV PROCESSES

All the models studied in this work are based on directed graphs. The vertices are called states. A pebble is placed in an initial state and moved from state to state creating an infinite path in the graph, so-called "play". How the pebble moves depends on who possesses the origin state. States can either belong to a player or be stochastic. In the first case, the player chooses where to move the pebble according to his scheduling strategy. In the second one, the pebble is moved according to a predefined probability distribution over successor states in the graph.

First, for a finite set $\mathcal{X}$, we let $\mathcal{D}(\mathcal{X})$ as the set of probability distributions over $\mathcal{X}$. This can be expressed mathematically as:

$$\mathcal{D}(\mathcal{X}) \quad := \quad \left\{ p : \mathcal{X} \to [0,1] \;\middle|\; \sum_{x \in \mathcal{X}} p(x) = 1 \right\} .$$

2.6.1 *MARKOV CHAINS*

Definition 2.6.1: Markov Chain

A Markov Chain (MC) is a structure:

$$\mathcal{C} \quad = \quad (\mathcal{X}, P) \,, \tag{2.12}$$

where $\mathcal{X}$ is a finite set of states, and $P : \mathcal{X} \to \mathcal{D}(\mathcal{X})$ is a probability transition function between states.

Since $\mathcal{X}$ is assumed to be a finite set, we can take any bijective function $\mathcal{X} \hookrightarrow\!\!\!\!\!\rightarrow \{0, 1, ..., |\mathcal{X}| - 1\}$, to label the states with natural numbered indices. Based on this, for a Markov chain $\mathcal{C} = (\mathcal{X}, P)$ and a given initial state[3] $x_0 \in \mathcal{X}$, we define the set Ω_{x_0} of **infinite paths** in $\mathcal{C}$ from x_0, also called **sample space**, by:

$$\Omega_{x_0} \quad := \quad \left\{ (x_0, x_1, ...) \in \mathcal{X}^{\mathbb{N}} \;\mid\; P(x_{t-1})(x_t) > 0, \;\; \forall\, t \in \mathbb{N} \right\} . \tag{2.13}$$

When we refer to **finite paths**, we denote by $\underline{\Omega}_{x_0}$ the set of these. We let $\mathcal{B}(\Omega_{x_0})$ as the **Borel sigma-algebra** over Ω_{x_0} generated by all possible cylinder sets:

$$\mathrm{Cyl}(x_0, x_1, ..., x_T) \quad := \quad \left\{ \omega \in \Omega_{x_0} \;\mid\; \mathrm{proj}_t^{\mathcal{X}}(\omega) = x_t, \;\; \forall\, t \le T \right\} , \tag{2.14}$$

for each $T \in \mathbb{N}$, where $\mathrm{proj}_t^{\mathcal{X}}$ is the t^{th}-projection map on paths. In words, a **cylinder set** specifies the first T states after x_0 for the paths. To reside in such a set, an ω must match these on the first T states,

3 Although all the modeling of power consumption scheduling problems studied in this work using Markov process, start from a given initial state of the system (namely, the DN) denoted by x_1, here we write x_0 since it is more natural in this mathematical context. When necessary for the consumption modeling, we turn to the pertinent notation of x_1. All the results shown here remain invariant.

but can have arbitrary assignment thereafter. From any initial state $x_0 \in \mathcal{X}$, there is induced a **probability space** denoted by:

$$\left(\Omega_{x_0}, \mathcal{B}(\Omega_{x_0}), \mathbb{P}_{x_0}\right) , \tag{2.15}$$

with probability distribution over set of paths[4] $\mathbb{P}_{x_0} : \mathcal{B}(\Omega_{x_0}) \to [0, 1]$, such that for each cylinder $\mathrm{Cyl}(x_0, x_1, ..., x_T) \in \mathcal{B}(\Omega_{x_0})$,

$$\mathbb{P}_{x_0}[\mathrm{Cyl}(x_0, x_1, ..., x_T)] \;=\; \prod_{t=1}^{T} P(x_{t-1})(x_t) \;. \tag{2.16}$$

The latter probability $\mathbb{P}_{x_0}$ is such that $\mathbb{P}_{x_0}[\varnothing] = 0$ and $\mathbb{P}_{x_0}[\Omega_{x_0}] = 1$. Note that the elements of $\mathcal{B}(\Omega_{x_0})$ are those subsets of Ω_{x_0} which have well-defined probabilities. We call each one an **event**.

Definition 2.6.2: State Process

A Markov chain $\mathcal{C} = (\mathcal{X}, P)$, with a given initial state $x_0 \in \mathcal{X}$, induces a stochastic state process $(X_t)_{t\in\mathbb{N}_0}$ over Ω_{x_0}, which is a sequence of $\mathcal{X}$-valued random variables (r.vs.), defined as $X_t(\omega) := \mathrm{proj}_t^{\mathcal{X}}(\omega)$ for $\omega \in \Omega_{x_0}$.

In other words, for a path $\omega = x_0 x_1 ...$, $X_t(\omega) = x_t$ to each $t \in \mathbb{N}_0$. Note that for a state $x_t \in \mathcal{X}$, the probability that the state process $(X_t)_{t\in\mathbb{N}}$ is x_t at time-step t from an initial state $x_0 \in \mathcal{X}$, is:

$$\begin{aligned}\mathbb{P}_{x_0}[X_t = x_t] \quad &= \sum_{x_1,\dots,x_{t-1}} \prod_{\tau=1}^{t} \mathbb{P}_{x_0}[X_\tau = x_\tau \mid X_{\tau-1} = x_{\tau-1}] \\ &= \sum_{x_1,\dots,x_{t-1}} \prod_{\tau=1}^{t} P(x_{\tau-1})(x_\tau) \;,\end{aligned} \tag{2.17}$$

where $\sum_{x_1,\dots,x_{t-1}}$ denotes the summation over all tuples of the form $(x_1, ..., x_{t-1}) \in \mathcal{X}^{t-1}$. In addition, $(X_t)_{t\in\mathbb{N}_0}$ satisfies the **Markov property** or "memorylessness", i.e., for all states $x_1, ..., x_t \in \mathcal{X}$, it holds:

$$\begin{aligned}\mathbb{P}_{x_0}[X_t = x_t \mid X_{t-1} = x_{t-1}, ..., X_1 = x_1] &= \mathbb{P}_{x_0}[X_t = x_t \mid X_{t-1} = x_{t-1}] \\ &= P(x_{t-1})(x_t)\end{aligned}$$

We denote the **expectation operator** over the probability space (2.15) as $\mathbb{E}_{x_0}$. Thus, for a measurable function, let say $F : \underline{\Omega}_{x_0} \to \mathbb{R}$, we can compute its expected value as:

$$\mathbb{E}_{x_0}[F(X_0, ..., X_T)] = \sum_{x_1,\dots,x_T} F(x_0, ..., x_T) \prod_{t=1}^{T} P(x_{t-1})(x_t) \;. \tag{2.18}$$

For example, this operator can help us compute the expected lifetime of the DN-transformer, to quantify the impact of consuming on the DN. The expected operator (2.18) can be naturally extended to functions over infinite paths by considering the limit behaviour. In general, such expected value does not need to converge as $T \to +\infty$, but it is

4 Carathéodory's extension theorem induces a unique probability measure on the Borel sigma-algebra over the simple space [52].

also possible to define two variants: the $\liminf$ and the $\limsup$ limits. Provided that

$$\mathbb{E}_{x_0}[\,|F(X_0, \ldots, X_T)|\,] \tag{2.19}$$

is convergent as $T \to +\infty$ from x_0, the expected value (2.18) is thus extended for functions $F : \Omega_{x_0} \to \mathbb{R}$, and can be written as:

$$\lim_{T \to +\infty} \mathbb{E}_{x_0}[F(X_0, \ldots, X_T)]\ .$$

We mainly focus on doubly-weighted processes in this work, but the technical developments mainly rely on simply-weighted process. We therefore define the setting with an arbitrary number of weights (cost functions).

Definition 2.6.3: Multi-Weighted Markov Chain

A multi-weighted Markov Chain is a structure:

$$\mathcal{C} \;=\; (\mathcal{X},\, P,\, (C_j)_{j=1}^{J})\ ,$$

where $J \in \mathbb{N}$, $(\mathcal{X},\, P)$ is a Markov chain (2.12), and $C_j : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ is a cost functions over transitions for each $j = 1, .., J$.

Note that above, the cost functions are defined on transitions. Similarly, a multi-weighted MC can be defined with costs over states, i.e., with costs $C_j : \mathcal{X} \to \mathbb{R}$, $j = 1, \ldots, J$. In any case, there is a correspondence between costs defined on transitions and on states. Indeed, let $j \in \{1, \ldots, J\}$ and let any state $x_t \in \mathcal{X}$ with cost $C_j(x_t)$. The latter can be transformed into a cost on transitions from such state x_t as:

$$C_j(x_t, x_{t+1}) \;\coloneqq\; C_j(x_t)$$

for each x_{t+1} such that $P(x_t)(x_{t+1}) > 0$. Conversely, if a transition takes place from any x_t to some x_{t+1}, i.e., $P(x_t)(x_{t+1}) > 0$, and the cost $C_j(x_t, x_{t+1})$ is incurred, we can define the cost on x_t as:

$$C_j(x_t) \;\coloneqq\; \sum_{x_{t+1} \in \mathcal{X}} C_j(x_t, x_{t+1}) P(x_t)(x_{t+1}) \tag{2.20}$$

Note that the latter is the expected value (2.18) of the cost in one-step, i.e., for each state $x_t \in \mathcal{X}$, $C_j(x_t) = \mathbb{E}_{x_t}[C_j(X_t, X_{t+1})]$.

An important example to compute the expected value involving the costs functions of a multi-weighted MC, is considering the sum of cost with a given finite-horizon time. It is known as the finite truncated sum, also called total-payoff.

Definition 2.6.4: Finite-Truncated Sum

Let $\mathcal{C} = (\mathcal{X}, P, (C_j)_{j=1}^{J})$ a multi-weighted MC, with $J \in \mathbb{N}$. For $j \in \{1, ..., J\}$ and $T \in \mathbb{N}$, we define the Truncated Sum (TS) up to the finite-horizon T, by:

$$\mathrm{TS}_{j,T} \;:=\; \sum_{t=1}^{T} C_j(X_{t-1}, X_t) \tag{2.21}$$

From the latter, computing an expected cost over the truncated sum comes naturally with the eq. (2.18), which formalizes the accumulation procedure of costs along paths by simple expected summations. In addition, it is also possible to extend the summations to the limit, provided there is convergence. Formally, we define the following.

Definition 2.6.5: Expected Truncated Sum

Let $\mathcal{C} = (\mathcal{X}, P, (C_j)_{j=1}^{J})$ a multi-weighted MC, with $J \in \mathbb{N}$, and $T \in \mathbb{N}$. For each $j = 1, ..., J$, we define the expected truncated sum up to T, as the function $\mathbb{TS}_{j,T} : \mathcal{X} \to \mathbb{R}$, which take any state $x_0 \in \mathcal{X}$ and defines:

$$\mathbb{TS}_{j,T}(x_0) \;:=\; \mathbb{E}_{x_0}[\,\mathrm{TS}_{j,T}\,]\,. \tag{2.22}$$

In addition, if the latter is convergent for each $x_0 \in \mathcal{X}$ as $T \to +\infty$, the expected truncated sum with infinite-horizon is defined as:

$$\mathbb{TS}_{j,\infty}(x_0) \;:=\; \lim_{T\to+\infty} \mathbb{TS}_{j,T}(x_0)\,.$$

In general, the function $\mathbb{TS}_{j,T}$ need not converge as $T \to +\infty$. For example, if costs have different signs, then their accumulate sum can oscillate. In order not to be concerned with such oscillations, we can impose the stronger condition on the absolute convergence for the infinite-horizon case, as in eq. (2.19). In the following, we provide a method which helps to evaluate the expected truncated sum [60].

Proposition 2.6.6

Let $\mathcal{C} = (\mathcal{X}, P, (C_j)_{j=1}^{J})$ be a multi-weighted MC, with $J \in \mathbb{N}$. For a fixed $j \in \{1, ..., J\}$ and any state $x_0 \in \mathcal{X}$, the following holds:

(i) for $T \in \mathbb{N}$ fixed, $\mathbb{TS}_{j,T}(x_0)$ can be computed iteratively through:

$$\mathbb{TS}_{j,T}(x_0) \;=\; C_j(x_0) \;+\; \sum_{x_1\in\mathcal{X}} P(x_0)(x_1)\mathbb{TS}_{j,T-1}(x_1)\,,$$

where $\mathbb{TS}_{j,0}(x_0) := 0$ for each $x_0 \in \mathcal{X}$.

(ii) if $\mathbb{TS}_{j,\infty}(x_0)$ exists, then it satisfies the system of linear eqs.:

$$\mathbb{TS}_{j,\infty}(x_0) = C_j(x_0) + \sum_{x_1 \in \mathcal{X}} P(x_0)(x_1)\mathbb{TS}_{j,\infty}(x_1) .$$

Proof. See Proof B.1.1 in Appendix B.1. ■

In this work, we are principally interested in computing the truncated sum until some **goal set**, also called **target set**, i.e., an a priori defined subset of states, is reached by the state process (2.6.2). Formally, a state $x \in \mathcal{X}$ is said a **reachable state** from a fixed initial state x_0, if there exists $t \in \mathbb{N}_0$ such that $\mathbb{P}_{x_0}[X_t = x] > 0$. Moreover, such a state is called **almost-surely reachable state** if this is reached with probability one. Based on this, we define the following random variable (r.v.), called **reachability time**, which represents when the state process eventually reaches x. It is defined by:

$$T_x := \inf \left\{ t \in \mathbb{N}_0 \cup \{+\infty\} \mid X_t = x \right\} . \tag{2.23}$$

Note that T_x is finite if x is an almost-surely reachable state, and it is $+\infty$ if along a path $\omega \in \Omega_{x_0}$ the process never reaches x. For example, if the state process represents the dynamic of charging the battery of an EV, the reachability time can represent when the battery is recharged. In the following, we define the truncated sum (2.6.4), but now considering the stochastic reachability time to reach a goal set.

Definition 2.6.7: Truncated Sum

Let $\mathcal{C} = (\mathcal{X}, P, (C_j)_{j=1}^J)$ a multi-weighted MC, with $J \in \mathbb{N}$. For $j \in \{1, ..., J\}$, and a given set of goal states $\mathcal{G} \subset \mathcal{X}$, we define the Truncated Sum (TS) by:

$$\mathrm{TS}_j^{\mathcal{G}} := \begin{cases} \sum_{t=1}^{T_{\mathcal{G}}} C_j(X_{t-1}, X_t) & \text{if } T_{\mathcal{G}} \in \mathbb{N} \\ +\infty & \text{otherwise} \end{cases} \tag{2.24}$$

where $T_{\mathcal{G}}$ is the reachability time defined in (2.23) for the set $\mathcal{G}$.

In the rest of this chapter, we assume that the goal set $\mathcal{G}$ is **absorbent**, i.e., that there is a single loop in each state of $\mathcal{G}$ whose costs are all equal to zero. This assumption is w.l.o.g., since we will study the truncated sums, which only consider the finite paths up to the first visit to the goal set $\mathcal{G}$. Mathematically, we assume the following.

Assumption 2.6.8: Absorbent Goal Set

For any defined goal set $\mathcal{G} \subset \mathcal{X}$, we assume that $\forall\, x \in \mathcal{G}$, $P(x)(x) = 1$ and $C_j(x, x) = 0$ for each $j = 1, ..., J$.

2.6.2 *MARKOV DECISION PROCESSES*

Roughly speaking, an MDP extends the pure stochastic behaviour of an MC by introducing actions, which can be used by a scheduling controller in order to control state transitions.

Definition 2.6.9: Multi-Weighted MDP

A multi-weighted Markov Decision Process (MDP) is a structure:

$$\mathcal{M} \quad := \quad \left(\mathcal{X},\, \mathcal{A},\, P,\, (C_j)_{j=1}^{J}\right) ,$$

where $J \in \mathbb{N}$, $\mathcal{X}$ is a finite set of states, $\mathcal{A}$ is an alphabet of actions, $P : \mathcal{X} \times \mathcal{A} \to \mathcal{D}(\mathcal{X})$ is a probability transition function between states, and $C_j : \mathcal{X} \times \mathcal{A} \times \mathcal{X} \to \mathbb{R}$ is a cost function over transitions for each $j = 1, .., J$.

In analogy with Markov chains, see eq. (2.20), we denote for $j = 1, ..., J$; the cost on a state x_t when an action a_t is taken and the transition to some x_{t+1} takes place, i.e., $P(x_t, a_t)(x_{t+1}) > 0$, as the expected cost:

$$C_j(x_t, a_t) \quad := \sum_{x_{t+1} \in \mathcal{X}} C_j(x_t, a_t, x_{t+1}) P(x_t, a_t)(x_{t+1}) \; . \tag{2.25}$$

We assume that there exists a single entity, called **scheduling controller** or system operator, that schedules the actions from the set $\mathcal{A}$, based on some information of the system. More precisely, such controller uses a function called scheduling strategy or decision rule to define an action to play in each state. The controller takes into account the finite paths (also called histories) of the system, which represents the visited system states and the actions chosen previously to when he should choose the action. First, analogously to the set of infinite paths (2.13) in a MC, we extend it for an MDP $\mathcal{M} = (\mathcal{X}, \mathcal{A}, P, (C_j)_{j=1}^{J})$ as:

$$\Omega_{x_0} = \left\{ (x_0, a_0, x_1, ...) \in \mathcal{X} \times (\mathcal{A} \times \mathcal{X})^{\mathbb{N}} \mid P(x_{t-1}, a_{t-1})(x_t) > 0,\ \forall\, t \in \mathbb{N} \right\} .$$

We write $\mathrm{proj}_t^{\mathcal{X}}$ and $\mathrm{proj}_t^{\mathcal{A}}$, to refer to the t^{th}-projection map on paths, resp. over $\mathcal{X}$ and $\mathcal{A}$ for $t \in \mathbb{N}_0$, e.g., for a path $\omega = x_0 a_0 x_1 ...$, $\mathrm{proj}_t^{\mathcal{X}}(\omega) = x_t$ and $\mathrm{proj}_t^{\mathcal{A}}(\omega) = a_t$.

The notion of Borel sigma-algebra $\mathcal{B}(\Omega_{x_0})$, cylinder sets (2.14), and thus the induced probability space (2.15) are naturally extended, wherein for each $\mathrm{Cyl}(x_0, a_0, x_1, ..., x_T) \in \mathcal{B}(\Omega_{x_0})$, the probability distribution (2.16) is now as:

$$\mathbb{P}_{x_0}[\mathrm{Cyl}(x_0, a_0, x_1, ..., x_T)] \quad = \quad \prod_{t=1}^{T} P(x_{t-1}, a_{t-1})(x_t) \; .$$

A useful technique that we use in this work is called "unfolding", which takes a multi-weighted MDP and defines another MDP without costs between transitions. More precisely, the unfolding is constructed by explicitly keeping track of the cost functions in the states.

Definition 2.6.10: Unfolding of an MDP

Let $\mathcal{M} = (\mathcal{X}, \mathcal{A}, P, (C_j)_{j=1}^J)$ a multi-weighted MDP, with $J \in \mathbb{N}$. For a given initial state $x_0 \in \mathcal{X}$ and $T \in \mathbb{N}$, we define the unfolding of $\mathcal{M}$ of depth T as the MDP:

$$\mathcal{M}_T \; := \; (\mathcal{S}_T,\, s_0,\, \mathcal{A},\, P_T) \; ,$$

where the space of states here is:

$$\mathcal{S}_T \; := \; \mathcal{X} \times \bigtimes_{j=1}^{J} [T\,C_j^{\min},\; T\,C_j^{\max}] \times \{0,\, 1,\, ...,\, T\} \; ,$$

with $C_j^{\min}, C_j^{\max} \in \mathbb{R}$ to be resp. the minimum and maximum cost appearing in the transitions of $\mathcal{M}$ due to C_j, $j = 1, ..., J$. The initial state is $s_0 := (x_0, 0, 0)$, $\mathcal{A}$ is the set of actions from $\mathcal{M}$, and $P_T : \mathcal{S}_T \times \mathcal{A} \to \mathcal{D}(\mathcal{S}_T)$ is the probability transition function between states defined as $P_T := P \circ (\mathrm{proj}_1, \mathrm{id}_{\mathcal{A}})$. In addition, each state $s_T \in \mathcal{S}_T$ such that $\mathrm{proj}_{J+2}(s_T) = T$, is considered as an initial state of $\mathcal{M}$, i.e., we keep a copy of $\mathcal{M}$ below each leaf of $\mathcal{M}_T$.

2.6.3 *STRATEGIES*

As we said before, the way in which the actions are chosen by the scheduling controller in each state is according to a function called strategy. This can be defined in several manners, we focus here on pure, randomized and mixed strategies [84]. In the following, we fix an initial state $x_0 \in \mathcal{X}$ in the MDP $\mathcal{M}$, and we consider $\underline{\Omega}_{x_0}$ be the set of finite paths.

First, a **pure strategy** is defined as a function:

$$\pi : \underline{\Omega}_{x_0} \longrightarrow \mathcal{A}$$

assigning an action for each finite path $\omega_t = (x_0, a_0, ..., x_t) \in \underline{\Omega}_{x_0}$, where $t \in \mathbb{N}_0$, i.e., $\pi(\omega_t) = a_t$. This strategy is also called **deterministic strategy**. We denote by Π the set of all pure strategies. On the other hand, a **randomized strategy** is defined as a function:

$$\delta : \underline{\Omega}_{x_0} \longrightarrow \mathcal{D}(\mathcal{A})$$

defining a probability distribution over actions for each finite path, i.e., for each $\omega_t = (x_0, a_0, ..., x_t) \in \underline{\Omega}_{x_0}$, with $t \in \mathbb{N}_0$, this is defined as $\delta(\omega_t)(a_t) \in [0, 1]$. From the traditional game theory [84], the previous definition corresponds to the one of **behavioural strategy**, i.e., putting a randomization on the choice of actions. We denote by Δ the set of all randomized strategies. Note that a pure strategy may be regarded as a special case of a randomized one in which the probability distribution on the set of actions is degenerate, i.e., $\delta(\omega_t)(a_t) = 1$ for some $a_t \in \mathcal{A}$. This type of strategies is in contrast to mixed strategies, which directly put a probability distribution over pure strategies. Mathematically, a **mixed strategy** is a function:

$$\sigma : \Pi \longrightarrow [0, 1]$$

We denote by $\Delta[\Pi]$ the set of mixed strategies. When the set of pure strategies is finite, i.e., when $|\Pi| \in \mathbb{N}$, the set $\Delta[\Pi]$ can be seen as a simplex in $\mathbb{R}^{|\Pi|}$, which is compact, convex and of dimension $|\Pi| - 1$. Thus, a mixed strategy can be regarded as a convexification of pure strategies.

Intuitively, the difference between randomized and mixed strategies, is that a randomized one defines whenever an action must be chosen, whereas a mixed strategy defines once before starting to play, how to choose randomly a pure strategy. An important result is that in a game with several players (also-called controllers or decision-makers here) in extensive form[5] under a setting of perfect information about the past of finite paths (which is the case in this work), the well-known Kuhn's theorem applies [12, 77], which states that randomized and mixed strategies have the same "power". In other words, if the scheduling controller can fulfill an objective with any of those strategy types, he can also fulfill the objective when restricted to the other one. We quote the theorem that interests us:

Theorem 2.6.11: [Kuhn, 1953]

In every game in extensive form, if a player has perfect recall, then for every mixed strategy of such a player, there exists an equivalent behavior strategy.

Proof. See, e.g., Proof of Theorem 6.15 in [84] ∎

In our context, an unfolded-MDP can be seen as a game in extensive form with two players: the scheduling controller and the "Nature". The latter is a player who has no strategic interests in the outcome (or paths) in the system and plays randomly. Thus, the probability transition between state can be seen as a fixed randomized strategy of the Nature. In this work, we are interested in defining a randomized strategy from a mixed one of the controller. This will be useful in the Chapter 5. From the proof of the Theorem 2.6.11, let σ a mixed strategy, and ω_t a fixed finite path. An equivalent randomized strategy for σ in an unfolded-MDP can be defined for each action a_t available at t, by:

$$\delta(\omega_t)(a_t) \quad = \quad \sum_{\pi \in \Pi} \sigma(\pi)\, \mathbb{1}_{\pi(\omega_t)}(a_t) \ , \tag{2.26}$$

where $\mathbb{1}_{\pi(\omega_t)}(a_t) = 1$ if $\pi(\omega_t) = a_t$, and $\mathbb{1}_{\pi(\omega_t)}(a_t) = 0$ otherwise. Of course, Kuhn's theorem is more general since a randomized strategy can be defined over sets of information. Here, such sets are reduced to be

5 That is, a game tree, which consists of a directed graph in which the set of vertices represents positions in the game. In each vertex a player is assigned to take an action, see [84] for more details

singleton sets. The definition in eq. (2.26) is sufficient for our purposes in Chapter 5, where we also see that the number of pure strategies to define the randomized one that interests us is finite.

MEMORY

We say that a pure strategy π (resp. a randomized one δ) is said **memoryless** or **Markov** if:

$$\pi(\omega_t) \;=\; \pi(\omega'_t) \qquad\qquad (\text{resp. } \delta(\omega_t) \;=\; \delta(\omega'_{t'})) ,$$

where $\omega_t, \omega'_{t'} \in \underline{\Omega}_{x_0}$ are any two finite paths, such that $\text{proj}_t^{\mathcal{X}}(\omega_t) = \text{proj}_{t'}^{\mathcal{X}}(\omega'_{t'})$. That is to say that the only relevant information that a Markov strategy needs is contained in the current state of the system. The set of pure (resp. randomized) Markov strategies is denoted by $\underline{\Pi}$ (resp. $\underline{\Delta}$). In such a case, we sometimes represent the action a_t taken in a state x_t by a pure (resp. randomized) Markov strategy as:

$$\pi(x_t) \;=\; a_t \qquad\qquad (\text{resp. } \delta(x_t)(a_t) \;\in\; [0,1]) .$$

Note that we do not define mixed strategies with memory, because if it is the case, this is due to the memory of pure strategies.

On the other hand, we say that a strategy is **non-memoryless** if it does not take the same action every time it visits the same state, and is thus dependent of time. In such case, we denote π_t (resp. δ_t) a non-memoryless pure (resp. randomized) strategy. The mainly goal here is to provide efficient implementable strategies. Hence, we favor strategies that are considered simple to implement and/or using a minimal amount of resources, notable the memory requirements over paths, where memoryless or Markov strategies are preferred. However, memory can be necessary to fulfill a (complex) objective. Hence, it is of interest to find strategies with finite-memory, as they correspond to strategies that can be implemented in practice. On the contrary, strategies with prohibitive (infinite) memory requirements are not implementable, see for example [29].

2.6.4 *INDUCED MARKOV CHAIN*

When a strategy is fixed, the resulting paths in $\mathcal{M}$ are fully stochastic, where no more decisions are made anymore. More precisely, a strategy resolves the non-deterministic choice between actions and thus, an MDP is reduced into a MC [96]. Indeed, let $\mathcal{M} = (\mathcal{X}, \mathcal{A}, P, (C_j)_{j=1}^J)$ a multi-weighted MDP and a fixed initial state $x_0 \in \mathcal{X}$. First, suppose that the scheduling controller uses some pure Markov strategy π. In such a case, we can define a multi-weighted MC by:

$$\mathcal{M}^\pi \;:=\; \left(\mathcal{X},\, x_0,\, P^\pi,\, (C_j^\pi)_{j=1}^J\right) ,$$

where, P^π is the probability transition function between states defined from P, i.e., if a transition from x_t to x_{t+1} takes place using π, then:

$$P^\pi(x_t)(x_{t+1}) \;:=\; P(x_t, \pi(x_t))(x_{t+1}) , \tag{2.27}$$

Analogously, for each $j = 1, ..., J$ the cost function C_j^π is defined from C_j resp., i.e.,

$$C_j^\pi(x_t, x_{t+1}) \quad := \quad C_j(x_t, \pi(x_t), x_{t+1}) \; .$$

On the other hand, suppose that the scheduling controller uses some randomized Markov strategy δ. In such case, we can define a multi-weighted MC by:

$$\mathcal{M}^\delta \quad := \quad \left(\mathcal{X},\, x_0,\, P^\delta,\, (C_j^\delta)_{j=1}^J\right) ,$$

where as before, if a transition from x_t to x_{t+1} takes place using δ, then:

$$P^\delta(x_t)(x_{t+1}) \quad := \quad \sum_{a_t \in \mathcal{A}} \delta(x_t)(a_t) P(x_t, a_t)(x_{t+1}) \; , \tag{2.28}$$

and for each $j = 1, ..., J$ the cost function C_j^π is defined from C_j resp. by:

$$C_j^\delta(x_t, x_{t+1}) \quad := \quad \sum_{a_t \in \mathcal{A}} \delta(x_t)(a_t) C_j(x_t, a_t, x_{t+1}) \; .$$

From the definition of P^π in (2.27) (resp. P^δ in (2.28)), the set of infinite paths in the Markov chain $\mathcal{M}^\pi$ (resp. $\mathcal{M}^\delta$) is well-defined, which we denoted by $\Omega_{x_0}^\pi$ (resp. $\Omega_{x_0}^\delta$). Analogously for the set of finite paths, we write $\underline{\Omega}_{x_0}^\pi$ (resp. $\underline{\Omega}_{x_0}^\delta$). In this way, the notion of the induced probability space $(\Omega_{x_0}^\pi, \mathcal{B}(\Omega_{x_0}^\pi), \mathbb{P}_{x_0}^\pi)$ and $(\Omega_{x_0}^\delta, \mathcal{B}(\Omega_{x_0}^\delta), \mathbb{P}_{x_0}^\delta)$ are naturally understood. Also, expectation operator (2.18) over the probability space induced by a pure strategy π (resp. a randomized strategy δ), is denoted by $\mathbb{E}_{x_0}^\pi$ (resp. $\mathbb{E}_{x_0}^\delta$).

Remark 2.6.12: MC by a non-Markov Strategy

Analogous constructions apply to finite history-dependent strategies, i.e., strategies that are not Markov and take the information of the set of finite paths $\underline{\Omega}_{x_0}$ to select actions. Its probability transition function is defined based on the distributions prescribed by the strategy and in order to accurately account for the memory updates, i.e., the finite paths taken into account for decisions. For instance, if the set of states in the MDP is $\mathcal{X}$ and a non-Markov strategy is used, let say a pure one π, the state space for the induced MC will be $\underline{\Omega}_{x_0}^\pi \times \mathcal{X}$. Still, there exists a bijection between paths of the MC and their traces in the MDP, thanks to the projection operator [34].

2.6.5 OBJECTIVES AND SYNTHESIS PROBLEM

The goal of the scheduling controller is to enforce a given specification, encoded through an objective. From the tradition formal methods theory, see for example [61], an objective is a predefined set of paths in an MDP. A path either satisfies the property (we therefore say that it is winning) or it does not (i.e., it is losing). Specifying objectives for a model can be formulated in different ways. We focus in this work in *(i)* qualitative objectives, e.g., in probability, in almost-surely or sure mode; and in *(ii)* quantitative objectives, e.g., the expected value. Because we are interested in measuring predefined sets of paths in an MDP, objectives are events that have well-defined probabilities. Also, objectives can be seen as constraints in modelling a system. In the following, we present the principal objectives considered in this work. Since pure strategies can be seen as a particular case of randomized strategies, we define the principal objectives for randomized strategies.

QUALITATIVE OBJECTIVES

This type of objectives express Boolean properties [61] on paths of the system, e.g., the reachability objective to some predefined set of goal states. Qualitative specifications are sufficient to model YES/NO properties. Thus, a path is either correct or incorrect, with no interpretation of how well it behaves, and then cost functions are not relevant here. We consider two types of objectives: the almost-surely and the probability objectives. We fix an ordinary MDP, i.e., without cost functions,

$$\mathcal{M} \;=\; (\mathcal{X}, \mathcal{A}, P) \tag{2.29}$$

an initial state $x_0 \in \mathcal{X}$, and we let $(\Omega_{x_0}, \mathcal{B}(\Omega_{x_0}), \mathbb{P}_{x_0})$ the underlying probability space.

Definition 2.6.13: Almost-Sure Objective

A strategy $\delta \in \Delta$ is said that it satisfies almost-surely an objective $\Theta \in \mathcal{B}(\Omega_{x_0})$, if $\mathbb{P}^{\delta}_{x_0}[\Theta] \;=\; 1$.

The term "almost-surely" refers to the possible exception of a set of paths whose probability measure is zero. On the other hand, if the sample space Ω_{x_0} is finite, there is no difference between the terms **almost-sure** and **sure**. The latter is defined as follows.

Definition 2.6.14: Sure Objective

A strategy $\delta \in \Delta$ is said that it satisfies surely an objective $\Theta \in \mathcal{B}(\Omega_{x_0})$, if for all $\omega \in \Omega^{\delta}_{x_0}$, $\omega \in \Theta$.

In other words, the Definition 2.6.14 refers to that δ ensures an objective Θ against any stochastic behaviour of the paths in $\mathcal{M}$. More precisely, $\Theta = \Omega^{\delta}_{x_0}$ in the induced MC $\mathcal{M}^{\delta}$.

Example 2.6.15. *Consider that the states of the MDP (2.29) represents the non-controllable part of the total load consumption generated on the DN-transformer, where the controllable part is assumed to be the actions. If it is required that the total load consumption at each time should not exceed the maximal load $\ell^{\max}$ of the DN-transformer, then the set of all possible values of the total load consumption almost-never exceed such an upper bound. Mathematically, for a strategy $\delta \in \Delta$, it is expressed by:*

$$\mathbb{P}^{\delta}_{x_0}[\{\omega \in \Omega_{x_0} \mid \exists t \in \mathbb{N}_0 \ : \ \mathrm{proj}^{\mathcal{X}}_t(\omega) + \mathrm{proj}^{\mathcal{A}}_t(\omega) > \ell^{\max}\}] \ = \ 0 \, .$$

If the sample space is finite, we can write it as "surely":

$$\forall \omega \in \Omega^{\delta}_{x_0} \, , \qquad \mathrm{proj}^{\mathcal{X}}_t(\omega) \ + \ \mathrm{proj}^{\mathcal{A}}_t(\omega) \ \leq \ \ell^{\max}$$ □

A principal objective that we focus on in this work is the stochastic reachability objective that must be satisfied with probability one. It is presented in the following.

Definition 2.6.16: Stochastic Reachability Objective

Given a goal set $\mathcal{G} \subset \mathcal{X}$, the (stochastic) reachability objective $\Theta_{\mathcal{G}} \in \mathcal{B}(\Omega_{x_0})$ is defined as:

$$\Theta_{\mathcal{G}} \ := \ \{\omega \in \Omega_{x_0} \ \mid \ \exists t \in \mathbb{N}_0 \ : \ \mathrm{proj}^{\mathcal{X}}_t(\omega) \in \mathcal{G}\}$$

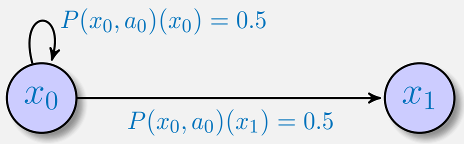
and, it is required that $\mathbb{P}^{\delta}_{x_0}[\Theta_{\mathcal{G}}] \ = \ 1$.

To satisfy this objective, the scheduling controller must be able to force visiting a state of $\mathcal{G}$ at least once. Note that $\Theta_{\mathcal{G}}$ can also be defined with the r.v. reachability time (2.23) by:

$$\Theta_{\mathcal{G}} \ = \ \{\omega \in \Omega_{x_0} \ \mid \ T_{\mathcal{G}}(\omega) \in \mathbb{N}_0\} \, .$$

Pure memoryless strategies suffice for the scheduling controller for this type of objective, and it can be constructed in POLYNOMIAL-TIME (if it exists), see [16] for instance.

Example 2.6.17. *Consider the following MDP:*



with the goal set defined as $\mathcal{G} = \{x_1\}$. *The only available action in the initial state* x_0 *is* a_0, *and there is a transition to* $\mathcal{G}$ *with probability* 0.5. *In that case, for the pure strategy* $\pi(x_0) = a_0$, *it satisfies the reachability objective almost-surely because* π *allows to reach (eventually) the goal set with probability one. Note that* $\mathcal{G}$ *is not reached surely, since there is a path*

$$\omega \quad = \quad x_0 a_0 x_0 a_0 ... \in (\{x_0\} \times \{a_0\})^{\mathbb{N}}$$

that never reaches the goal set, although this one has a zero probability. □

Sometimes, when the size of the set of paths in an MDP is large, observing all resulting paths may be a very strict assumption and generally inappropriate from realistic problems. If some risk may be accepted, it is useful to reason about the risk in probability of an objective over paths.

Definition 2.6.18: Probability Objective

Given a threshold $\varepsilon \in [0, 1]$ and an objective $\Theta \in \mathcal{B}(\Omega_{x_0})$, a strategy $\delta \in \Delta$ is said that it satisfies the objective with ε-risk in probability, if $\mathbb{P}^{\delta}_{x_0}[\,\Theta\,] \;\geq\; 1 - \varepsilon$.

In other words, this means that one accepts that an event remains under a certain risk of probability parametrized by a given thresholds. Note that considering $\varepsilon = 0$, it is the same as the almost-surely objective.

Example 2.6.19. *Consider that the states of the MDP* (2.29) *represents the DN-transformer temperature. It is natural to think that exceeding a high temperature might put the DN-transformer in dangerous conditions of shut down. Here, one assumes a risk in probability* $\varepsilon \in [0, 1]$ *of exceeding the maximum temperature* $x^{\max}$, *a strategy* $\delta \in \Delta$ *satisfies it requirement if:*

$$\mathbb{P}^{\delta}_{x_0}[\,\{\,\omega \in \Omega_{x_0} \;\mid\; \mathrm{proj}^{\mathcal{X}}_t(\omega) \;\leq\; x^{\max} \;\;,\;\; \forall t \in \mathbb{N}_0\}\,] \;\;\geq\;\; 1 - \varepsilon$$ □

A principal objective that we focus on in this work is the stochastic shortest path objective to some set of goal states. Here we can accept some probability risk when satisfying the objective. This is known as the stochastic shortest path probability objective.

Definition 2.6.20: Probability Shortest Path Objective

Given a goal set $\mathcal{G} \subset \mathcal{X}$, a threshold $\nu \in \mathcal{R}$, and a probability threshold $\varepsilon \in [0,1]$; let $\mathrm{TS}_j^{\mathcal{G}}$ the truncated sum (2.24) for $j \in \{1, ..., J\}$. The probability (stochastic) shortest path objective is defined as:

$$\mathbb{P}_{x_0}^{\delta}[\,\mathrm{TS}_j^{\mathcal{G}} \;\geq\; \nu] \;\;\geq\;\; 1 - \varepsilon$$

This objective is PSPACE-HARD as shown in [62]. Pure strategies with exponential memory suffice [101]. A strategy for this objective can be obtained by a PSEUDO-POLYNOMIAL-TIME algorithm based on the unfolding method, see Definition 2.6.10.

QUANTITATIVE OBJECTIVES

While qualitative specifications are sufficient to model correct/incorrect properties, quantitative extensions to model resource constraints such as power consumption, can be considered. Thus, MDPs with costs come into play. Under such a context, the expectation operator (2.18) is an attractive approach for this type of objectives.

Definition 2.6.21: Expected Value Objective

Let $\mathcal{M} = (\mathcal{X}, \mathcal{A}, P, (C)_{j=1}^{J})$ a multi-weighted MDP with $J \in \mathbb{N}$, and $x_0 \in \mathcal{X}$ an initial state. Given a function $F : \Omega_{x_0} \to \mathbb{R} \cup \{\infty\}$ and a threshold $\nu \in \mathbb{R}$, a strategy $\delta \in \Delta$ is said that it satisfies the expected value objective for F and μ, if $\mathbb{E}_{x_0}^{\delta}[\,F\,] \;<\; \nu$.

Note that it is possible that F can be a composed function of the costs C_j, $j = 1, ..., J$; e.g., it can represent the sum of costs from the initial state. Also, the expected value objective can be written as a constraint in expectation over transitions for each time-step. For instance, we consider the following example.

Example 2.6.22. *Let C a cost function of an MDP, which gives the temperature of the DN-transformer at each time-step. Suppose that we expect that such temperature stays below a maximum prescribed temperature $x^{\max}$, for which the DN-transformer works in safety conditions. A constraint in expectation can be written as:*

$$\forall\, t \in \mathbb{N}\,, \qquad \mathbb{E}_{x_0}^{\delta}[\,C(X_{t-1}, A_{t-1}, X_t)\,] \;\;<\;\; x^{\max}\,.$$

For instance, exceeding a critical thresholds may cause DN-transformer accelerated ageing, see, e.g., [M1][19]. □

Although this type of objectives is mathematical attractive, it is difficult to understand the involved risk in choosing a prescribed threshold as we have seen in the precedent example. Here, we will focus in minimizing sum of costs until some goal set in an MDP is reached from some initial state. Such a quantitative setting is a generalization of the classical shortest path over graphs (see for example [38] for the graph setting). It is defined as follows.

Definition 2.6.23: Expected Shortest Path Objective

Given a goal set $\mathcal{G} \subset \mathcal{X}$ and a threshold $\nu \in \mathbb{R}$, let $\mathrm{TS}_j^{\mathcal{G}}$ the truncated sum (2.24) for $j \in \{1, \ldots, J\}$. The expected (stochastic) shortest path objective is defined as:

$$\mathbb{E}_{x_0}^{\delta}[\,\mathrm{TS}_j^{\mathcal{G}}\,] \;<\; \nu$$

Pure memoryless strategies suffice, and satisfying the requirement of the threshold can be solved in POLYNOMIAL-TIME via, e.g., linear programming [14, 25] or value iteration [47]. Alternatively to the expected shortest path objective, we can consider probability instead of expectation. In such a case, the requirement is on strategies that maximize the probability of short paths to a goal set, see Definition 2.6.20.

SYNTHESIS PROBLEM

The principal question is to decide whether the scheduling controller has a strategy satisfying an objective, e.g., reaching a set of goal states. For all the objectives considered in this work, see Section 2.6.5, we are also interested in constructing such strategies (provided of the existence). This is known as the **synthesis problem**. Another objective is to provide efficient implementable strategies. Hence, we favor strategies that are considered simple to implement and/or using a minimal amount of resources, notable the memory requirements over paths, where memoryless or Markov strategies are preferred, see Section 2.6.3 for details.

All numerical analysis in this thesis is done through MATLAB and PRISM [4, 78] tools. The latter is a tool for formal modeling and analysis of systems that exhibit random or probabilistic behavior, as in MDPs; including multiple efficient engines to describe and verify the models. It has been used to analyze systems from many different application domains, including randomized distributed algorithms, biological systems, communication and multimedia protocols, and many others. Here, we use it for the domain of smart grids. This tool is used in this work when the forecast on the noncontrollable part of the total load consumption is stochastic, to synthesize an optimal scheduling strategy. When the decision-making is made in a decentralized way, we developed an algorithm to execute iteratively and automatically between MATLAB and PRISM.

Part II

CENTRALIZED MODELING

This part presents the centralized approach of the power consumption scheduling problem. A single entity controls and builds the scheduling strategies of consumers.

3

CENTRALIZED POWER CONSUMPTION SCHEDULING PROBLEM

Abstract:

We start this chapter with a general mathematical modeling of the centralized problem of power consumption scheduling. To solve it, we describe four sorts of strategies for consumers. First, the strategies are imposed to be rectangular profiles. As a consequence, these merely consist of choosing the consumption start time. Second, the consumption are arbitrary and we take into account explicitly the evolution law of the system state. Under some conditions, the problem boils down to a standard optimization problem. Third, we replace the general problem by a valley-filling algorithm. Therefore, this relies only on the minimization of the total consumption and not on any other measure of impact over the network. Fourth, the stochastic modeling comes into play through Markov Decision processes with multi-constrained objectives, to adapt the strategies to several consumption scenarios.

LIST OF ABBREVIATIONS AND SYMBOLS

DN	Distribution Network	HS	Hot-Spot (temperature)
EVs	Electric Vehicle(s)	PaC	Plug-and-Charge
SNR	Signal-to-Noise Ratio	AAF	Accelerated Aging Factor
MDP	Markov Decision Process	ECP	Electrical Consumption Payment
SSP	Stochastic Shortest Path	h	hour
min	minutes	kV	Kilovolt
V	Volt	kWh	Kilowatt-hour
kW	Kilowatt	km	Kilometres
kVA	Kilovolt-Ampere	¢	cents of Australian dollar
dB	decibel		
$°C$	Degree Celsius		

$\mathcal{A}$	finite action space in $\mathcal{M}$
C	cost function of the power consumption problem
$C^{\boldsymbol{\pi}}_{x_1}$	cost function under a fixed $\boldsymbol{\pi}$ and x_1
C_i	cost function, instantaneous energy of consumer i
$C_{\mathcal{I}}$	cost function, instantaneous aggregated energy of consumers
$\boldsymbol{C}$	cost function between transitions in $\mathcal{M}_T$
Δ_t	time-step duration of each time-slot t
$\mathcal{D}(\mathcal{S})$	set of probability distributions over a finite set $\mathcal{S}$
ε_ℓ	risk in probability of exceeding $\ell^{\max}$
ε_x	risk in probability of exceeding $x^{\max}$
e_i	energy demand of consumer i
$e_{\mathcal{I}}$	aggregate energy demand of consumers
e_t	aggregate energy of consumers at t
$\mathcal{E}$	aggregate energy space of e_t
$\mathbb{E}^{\boldsymbol{\pi}}_{x_1}$	expectation operator under a fixed $\boldsymbol{\pi}$ and x_1
f	evolution law of the system state
$\mathcal{G}$	set of goal states in an MDP model
i	consumer
I	number of consumers
$\mathcal{I}$	set of consumers
$\mathrm{id}_{\mathcal{A}}$	identity function on a set $\mathcal{A}$
$\ell_i^{\min}$	minimum power of consumer i
$\ell_i^{\max}$	maximal power of consumer i
$\ell_{i,t}$	controllable load of consumer i at t
$L_{i,t}$	function representing $\ell_{i,t}$

$\boldsymbol{\ell}_t$	controllable load vector at t of length I
$\boldsymbol{L}_t$	function representing $\boldsymbol{\ell}_t$ at t
$\ell_{0,t}$	real noncontrollable load at t
$\boldsymbol{\ell}_0$	real noncontrollable load profile of length T
$\widetilde{\ell}_{0,t}$	deterministic forecast of $\ell_{0,t}$ at t
$\widetilde{\boldsymbol{\ell}}_0$	deterministic forecast profile of length T
$\widetilde{L}_{0,t}$	stochastic forecast of $\ell_{0,t}$ at t
$\widetilde{\boldsymbol{L}}_0$	stochastic forecast profile of length T
$\ell^{\max}$	maximal power of the DN-transformer
ℓ_t	total load consumption at t
$\widetilde{L}_t$	function representing the total load under $\widetilde{\boldsymbol{L}}_0$ at t
$\mathcal{L}$	total load consumption space
$\mathcal{M}$	MDP model
$\mathcal{M}_T$	unfolding of $\mathcal{M}$ with depth T
$\mathcal{N}(\mu, \sigma^2)$	Gaussian (normal) distribution with mean μ and variance σ^2
ω_t	history or path of length t of the system
P	transition probability between states in $\mathcal{M}$
P_T	transition probability between states in $\mathcal{M}_T$
P_0	probability distribution of $\widetilde{L}_{0,t}$
proj_j	projection function on the j-component of a sequence
$\boldsymbol{\pi}$	centralized strategy profile of length I
$\boldsymbol{\pi}_t$	centralized strategy of consumers at t
$\boldsymbol{\pi}^{\mathrm{RP}}$	a $\boldsymbol{\pi}$ built by rectangular profile method
$\boldsymbol{\pi}^{\mathrm{DC}}$	a $\boldsymbol{\pi}$ built by dynamic charging method
$\boldsymbol{\pi}^{\mathrm{VF}}$	a $\boldsymbol{\pi}$ built by valley-filling method
s_t	state in $\mathcal{S}_T$ at t
$\mathcal{S}_T$	augmented space of states in $\mathcal{M}_T$
t	time-slot
t_i^{start}	time at which the consumption starts for i
t_i^{stop}	time at which the consumption ends for i
T	finite horizon time
$\mathcal{T}$	set of time-slots
$x^{\max}$	upper bound of the system state values
x_t	system state at t
$\widetilde{x}_t$	system state under $\widetilde{\ell}_{0,t}$ at t
$\widetilde{X}_t$	function representing the system state under $\widetilde{\boldsymbol{L}}_0$ at t
$\mathcal{X}$	finite set of system states

3.1 MOTIVATION AND CONTRIBUTIONS

Motivated by our practical results reported in [M1] concerning an Electric Vehicles (EVs) charging problem, and also by the basic problem of power consumption mentioned in the Chapter 2, we aim a generalized mathematical modeling of such a problem, taking into account constrains of the Distribution Network (DN).

In this chapter, we consider a scenario in which several controllable consumption entities (also called controllable electric devices or simply consumers here) have a certain energy demand and want to have this demand to be fulfilled before a set deadline. A simple instance of such a scenario is the case of a pool of EVs which have to recharge their battery to a given state of charge within a given time window set by the EV owner. Thus, each consumer has to choose at each time the consumption power so that the accumulated energy reaches a desired level. This is the problem of power consumption scheduling that we want to study and generalize in this chapter, taking into account other constraints, e.g., the maximal admissible power of the DN-transformer.

The consumption operations for this problem are assumed to be centralized here, i.e., it has to be centralized in the sense that a single entity (a centralized system operator) controls and builds the power consumption scheduling strategies of consumers satisfying the requirements of the individual energy demands. Additionally, the centralized system operator needs a certain knowledge about the noncontrollable part of the total load consumption to schedule the strategies. A typical scenario considered in this chapter is that a day-head decision has to be made and some knowledge (imperfect deterministic/stochastic forecast on the noncontrollable consumption) is available.

3.1.1 *STRUCTURE*

The main contributions and structure of this chapter can be summarized as follows.

(i) In Section 3.2, we formulate the problem of centralized power consumption scheduling to be solved in a general mathematical form, where the objective is to minimize the impact of the total load consumption on the DN, e.g., maximizing the DN-transformer lifetime or minimizing the electrical consumption payment. Such an impact is taken into account by an objective function of interest, namely: a linear combination of cost functions that can be deterministic or stochastic depending of the noncontrollable load consumption forecast.

(ii) In Section 3.3, a deterministic forecast on the noncontrollable part of the total load consumption is considered to schedule the controllable part. In this case, the strategies are reduced to be parameter vectors in which each component is a consumption power. Three sorts of strategies are provided.

First, the consumption strategies are imposed to be rectangular, i.e., it must be either the minimum or the maximum power at which a consumer can be consuming. Also, it has to be uninterrupted until the energy demand is met. Each strategy thus boils down to a simple decision, namely: the consumption start time. This approach was inspired from [20][M4], but the centralized case was not analyzed there.

Second, the consumption are considered arbitrary here, and the dynamic of the system state is taken into account explicitly. This approach was motivated from [19], but here we provide a generalized method on the dynamic law of the system. Under some convexity conditions, the problem boils down to a simple (convex) optimization problem.

Third, the general power consumption scheduling problem is replaced by a valley-filling algorithm, which is based mainly on [106]. This methods is presented for purposes of numerical comparison.

(iii) On the other hand, the stochastic forecast case is presented in Section 3.4. Here, the strategies are no longer parameter vectors and they become dynamic. The dynamic structure of the problem is exploited here in a Markov decision process (see [96] for example). This allows us to build strategies more robust to forecast noises, as we have shown in [M1].

(iv) Numerical results are shown in Section 3.5, where the objective function is exploited as the degradation of the DN. First, in terms of the DN-transformer lifetime in Section 3.5.1, and second, in terms of the total electricity consumption payment in Section 3.5.2.

3.2 PROBLEM FORMULATION

We give here a brief summary of the formulation of the basic problem presented in Section 2.5, but turning to the general form. The formulation of the problem can be explained as follows.

LOAD CONSUMPTION

We consider a Distribution Network (DN) comprising one transformer (referred to here as DN-transformer) to which two groups of electric devices are connected: a set $\mathcal{I} = \{1, ..., I\}$, $I \in \mathbb{N}$, of controllable electric devices or consumers, e.g., Electric Vehicles (EVs), dishwashers, water-heaters, etc.; and a set of other electric devices. The latter group of electrical appliances is assumed to induce a load consumption which is independent of the loads of the controllable electric devices and therefore referred to as the noncontrollable load consumption (e.g., heating, lighting, cooking, etc.). Assuming that time is slotted and indexed by $t \in \mathcal{T} = \{1, ..., T\}$, $T \in \mathbb{N}$, the corresponding **real noncontrollable load** consumption profile is expressed as:

Each time-slot t may be of typical duration of 30 min.

$$\boldsymbol{\ell}_0 \;\; := \;\; (\ell_{0,t})_{t\in\mathcal{T}} \; . \tag{3.1}$$

On a given time-slot, a controllable electric device $i \in \mathcal{I}$ may be active or not. The extent to which each one of these is active on time $t \in \mathcal{T}$, we measure it by the **controllable load** consumption that i generates, which is denoted by $\ell_{i,t} \in \mathbb{R}_0^+$. The **controllable load vector** of the devices in $\mathcal{I}$ at time t is denoted by:

$$\boldsymbol{\ell}_t \;\; := \;\; (\ell_{i,t})_{i\in\mathcal{I}} \; . \tag{3.2}$$

The **total load** consumption on the DN-transformer at time t can be then expressed by:

$$\ell_t \;\; := \;\; \ell_{0,t} + \sum_{i\in\mathcal{I}} \ell_{i,t} \; . \tag{3.3}$$

Figure 3.1 illustrates a typical scenario that is encompassed by the considered model. This figure represents a set of consumers, where each of them is represented by an household and an EV. Both are connected to a DN-transformer. A single entity (centralized decision-maker) chooses the controllable load vectors at each time, see eq. (3.2), to reach the individual state of charges (energy demand) for the EVs, faces to the aggregated noncontrollable consumption of the households.

A natural constraint from the DN-transformer is due to its maximal admissible power (e.g., the maximal power of a typical DN-transformer in a urban district is 90kW). The consumers are aware of that and must keep the total load consumption lower then the maximal power $\ell^{\max} \in \mathbb{R}^+$. Thus, it is required that:

$$\ell_t \;\; \leq \;\; \ell^{\max} \; . \tag{3.4}$$

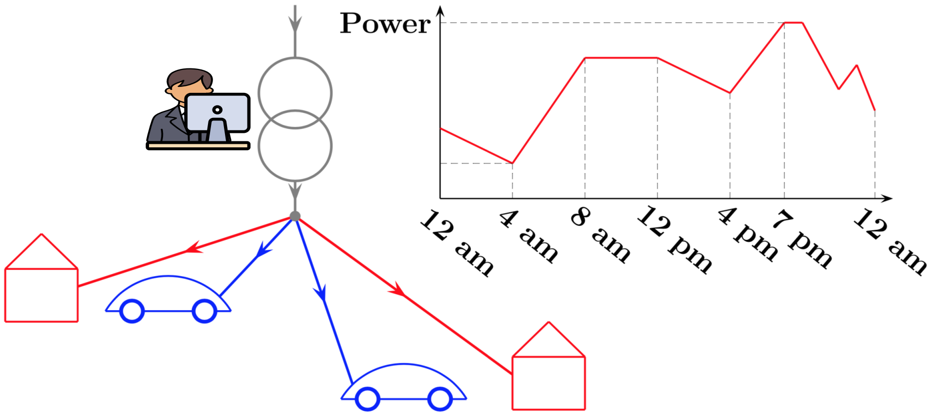


Figure 3.1: A centralized decision-maker controls the charging power of EVs to reach the individual energy demands. The centralized scheduler have some knowledge about the day-ahead (aggregated) noncontrollable part of the total consumption.

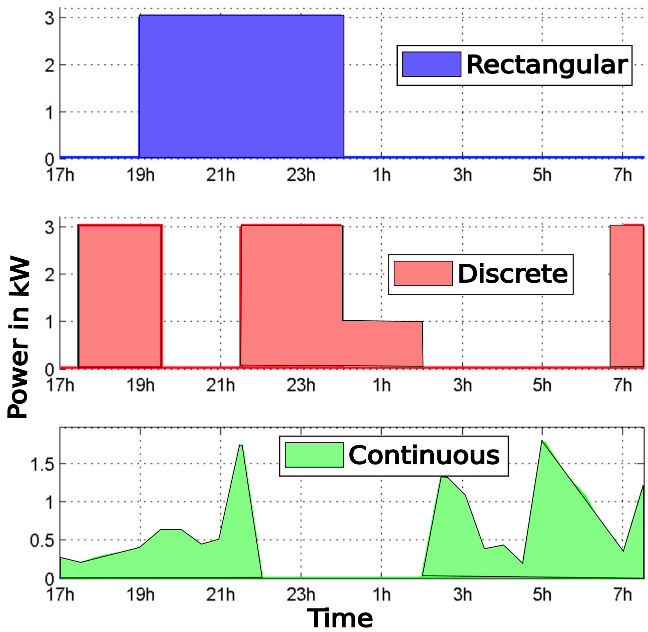


Figure 3.2: Four charging models of an EV: rectangular, continuous and discrete charging. Each model corresponds to a class of electrical uses far broader than the case of EVs.

It is assumed that Consumer i requires some energy to complete a corresponding task. That is why i wants to reach the **energy demand** $e_i \in \mathbb{R}^+$ before time $t = T$. The corresponding constraint for the scheduling problem can be written as:

$$\Delta_t \sum_{t \in \mathcal{T}} \ell_{i,t} \;\geq\; e_i \,, \tag{3.5}$$

where Δ_t is the time step duration (e.g., if each time-slot t represents $30\,\text{min}$, $\Delta_t = 0.5\,\text{h}$). Additionally, the power load of each i at time t, is assumed to be at least a **minimum power** $\ell_i^{\min} \in \mathbb{R}_0^+$ and cannot exceed the **maximal power** $\ell_i^{\max} \in \mathbb{R}^+$ at which i can be consuming:

$$\ell_i^{\min} \;\leq\; \ell_{i,t} \;\leq\; \ell_i^{\max} \,. \tag{3.6}$$

The minimum power refers when, e.g., i is switched off but is designed to draw some load in standby mode.

For instance, four sorts of charging models of an EV i are represented in Figure 3.2, namely: rectangular, continuous and discrete charging. In the first case, the controllable consumption of i at t can only take two values: either $\ell_{i,t} = \ell_i^{\min}$ or $\ell_{i,t} = \ell_i^{\max}$, and also when $\ell_{i,t} = \ell_i^{\max}$, the consumption is uninterrupted. This method is assumed in Section 3.3.1. In the second case, $\ell_{i,t}$ takes any arbitrary value between $\ell_i^{\min}$ and $\ell_i^{\max}$, see eq. (3.6). This method of consumption is assumed in Section 3.3.2 and Section 3.3.3. The last case is a discretization of the type of previous consumption, which is assumed in Section 3.4.1.

SYSTEM STATE

For the system of interest (the distribution network), the state is denoted by x_t (e.g., the monetary cost for consuming a given amount of energy, the DN-transformer temperature, etc.) and it is required to remain upper bounded by a given threshold $x^{\max} \in \mathbb{R}^+$ as follows:

If e.g., x_t represents the DN-transformer temperature, this must stays be at most a maximum prescribed temperature $x^{\max}$.

$$x_t \;\leq\; x^{\max} \,. \tag{3.7}$$

A general (non-linear) model assumed in this chapter for the **evolution law of the system state** is expressed by :

$$x_{t+1} \;=\; f(x_t, \boldsymbol{\ell}_t, \ell_{0,t}) \,. \tag{3.8}$$

for each $t = 1, ..., T$; where $x_1 \in \mathbb{R}^+$ is a given initial condition of the system state. A practical assumption in this chapter to make an effective calculation of controllable loads and to ensure convergence of algorithms, is the Assumption 2.5.1, which states that f is a function depending of the total load consumption (3.3).

SCHEDULING STRATEGIES

Based on all practical considerations, the controllable loads of consumers can be scheduled according to a function called power consumption scheduling strategy or decision rule. These are built and controlled by a single entity, i.e., a centralized system operator, which takes the decisions for all consumers at each time t. Once the scheduling strategies are fixed, a cost is incurred at each t and we can therefore compare the effectiveness of each strategy. The information that the centralized system operator takes into account to schedule is the history (also called path) of the system, represented by the visited states and the controllable loads chosen previously. More precisely, assume the following composed **history** or **path** of the system up to t is available to the centralized decision-maker:

$$\omega_t \;:=\; (x_1, \boldsymbol{\ell}_1, x_2, \boldsymbol{\ell}_2, \ldots, x_{t-1}, \boldsymbol{\ell}_{t-1}, x_t) \,. \tag{3.9}$$

A centralized scheduling **strategy** to select the controllable loads of consumers at time t, is defined by:

$$\boldsymbol{\pi}_t(\omega_t) \quad = \quad \boldsymbol{\ell}_t \ ,$$

and we let $\boldsymbol{\pi} := (\boldsymbol{\pi}_t)_{t\in\mathcal{T}}$ the profile of the latter. Note that this definition is according to the one of pure strategies seen in Section 2.6.3. Recall that, in this context, $\boldsymbol{\pi}$ is **memoryless** or **Markov** if for each $t \in \mathcal{T}$,

$$\boldsymbol{\pi}_t(\omega_t) \quad = \quad \boldsymbol{\pi}_t(\omega'_{t'}) \ , \tag{3.10}$$

where $\omega_t = (x_1, \boldsymbol{\ell}_1, x_2, ..., \boldsymbol{\ell}_{t-1}, x_t)$ and $\omega'_{t'} = (x_1, \boldsymbol{\ell}'_1, x'_2, ..., \boldsymbol{\ell}'_{t'-1}, x'_{t'})$ are any finite paths of the system, such that $x_t = x'_{t'}$. That is, the only relevant information that a Markov strategy needs is contained in the current state of the system.

SCHEDULING PROBLEM

The impact of the load consumption operations of consumers on the distribution network is measured as a composite cost over the whole time period under consideration. The centralized scheduling problem of interest can then be formulated as follows:

Centralized Power Consumption Scheduling Problem

$$\min_{\boldsymbol{\pi}} \quad \sum_{t\in\mathcal{T}} C^{\boldsymbol{\pi}}_{x_1}(x_t, \boldsymbol{\ell}_t, \ell_{0,t}) \tag{3.11}$$
$$\text{s.t.} \quad (3.3),\ (3.4),\ (3.5),\ (3.6),\ (3.7),\ (3.8)\ .$$

We aim to solve the scheduling problem (3.11), but at first glance, it is a difficult task. Indeed, it is known that determining an optimal solution of a scheduling problem with saturation constraints, e.g., one as (3.7), is generally difficult especially when the cost function is neither linear nor quadratic [19]. In addition, the way in which the controllable loads are chosen is according to a joint form by the centralized system operator, causing a high degree of information to handle with.

To solve the problem of power consumption scheduling (3.11), a forecast of the real noncontrollable load consumption (3.1) is assumed to be available, which can be deterministic or stochastic, see Section 2.4 for details. In this way, a strategy $\boldsymbol{\pi}$ is scheduled **offline** defining the controllable loads executed on the time under consideration. Once the strategy is determined, it can be effectively run **online**.

3.3 SOLUTION METHODOLOGY IN THE DETERMINISTIC CASE

In this section, a deterministic forecast on the real noncontrollable load consumption (3.1) is assumed to schedule the controllable loads to solve the scheduling problem (3.11). Following the description in Section 2.4.1, we represent such a **deterministic forecast** by:

$$\widetilde{\boldsymbol{\ell}}_0 \quad := \quad (\widetilde{\ell}_{0,t})_{t\in\mathcal{T}} \,. \tag{3.12}$$

Thus, the problem of power consumption scheduling to solve now can be rewritten as:

Centralized Deterministic Problem

$$\min_{\boldsymbol{\pi}} \quad \sum_{t\in\mathcal{T}} C^{\boldsymbol{\pi}}_{x_1}(\widetilde{x}_t, \boldsymbol{\ell}_t\,; \widetilde{\ell}_{0,t}) \tag{3.13}$$
$$\text{s.t.} \quad (3.3),\ (3.4),\ (3.5),\ (3.6),\ (3.7),\ (3.8),\ (3.12)\,.$$

Where, each noncontrollable load consumption at t is now a simple parameter, i.e., a value of the deterministic forecast. Note that the system state and the total load consumption are also affected by the forecast, because these depend on the noncontrollable loads, see eq. (3.8). Also note that choosing controllable loads, the system state is completely determined with the dynamics (3.8), and the scheduling strategy can therefore be found with the deterministic information of the system at each time. To solve the scheduling problem under the deterministic forecast assumption, we provide three sorts of strategies in the next Section.

3.3.1 *RECTANGULAR CONSUMPTION PROFILES*

In this section, we not only assume that the controllable loads $\ell_{i,t}$ can only take two values $\ell_i^{\min}$ or $\ell_i^{\max}$, but also that when consuming the load consumption has to be uninterrupted, which leads to rectangular consumption profiles. This is what Figure 3.3 shows.

Although this assumption seems to be restrictive, it can be strongly motivated (see e.g., [20]). Here we provide a couple of arguments of using such load consumption profiles. An important argument is that rectangular profiles are currently being used by existing electric vehicles. Second, for a given consumption start time, consuming at full power without interruption minimizes the delay to acquire the desired quantity of energy. Third, from an optimal control theory perspective, rectangular profiles may be optimal. This happens for instance when the state (e.g., the transformer HS temperature) is monotonically increasing with the control (i.e., the consumption power). Fourth, profiles without interruption are even required in some important practical scenarios encountered with home energy management. Last but not least, imposing the consumption profile to be rectangular makes the power consumption scheduling strategy robust to forecast errors [20][M4].

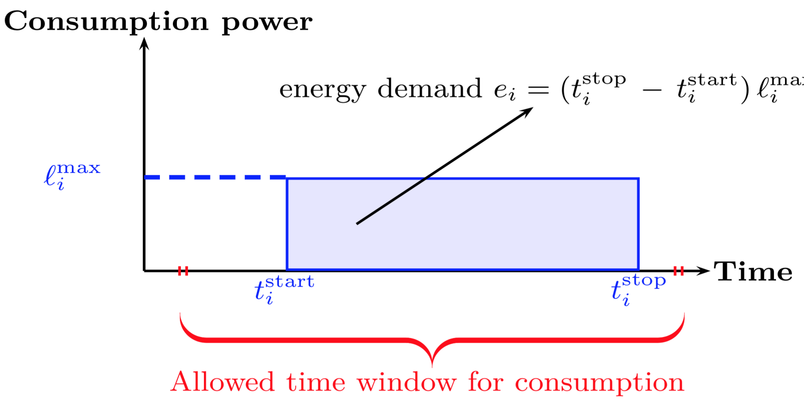


Figure 3.3: Power consumption scheduling strategy as rectangular profile charging for a consumer i.

Since we consider here rectangular consumption profiles, the power consumption scheduling strategy boils down to a simple decision namely: the time at which the consumption effectively starts. We denote the start time for i by $t_i^{\text{start}} \in \mathcal{T}$. Mathematically, rectangular profiles can be written as:

$$\boldsymbol{\ell}_i \in \left\{ (\ell_{i,t})_{t\in\mathcal{T}} \in \{\ell_i^{\min}, \ell_i^{\max}\}^T \;\middle|\; \begin{array}{l} \forall t \in \{t_i^{\text{start}}, ..., t_i^{\text{stop}}\} \subseteq \mathcal{T},\ \ell_{i,t} = \ell_i^{\max} \\ \forall t \in \mathcal{T} \setminus \{t_i^{\text{start}}, ..., t_i^{\text{stop}}\}\ ,\ \ell_{i,t} = \ell_i^{\min} \end{array} \right\} \tag{3.14}$$

In practice, from the constraint of energy demand (3.5), each t_i^{start} is limited to being:

$$t_i^{\text{start}} \quad \leq \quad T \;-\; \frac{e_i}{\ell_i^{\max}}\ , \tag{3.15}$$

and t_i^{stop} can be chosen being the minimum stopping time such that:

$$(t_i^{\text{stop}} - t_i^{\text{start}})\,\ell_i^{\max} \quad \geq \quad e_i\ , \tag{3.16}$$

where in eqs. (3.15) and (3.16) we do not take into account the load due to the minimum power $\ell_i^{\min}$, since it only refers when, e.g., i is switched off but is designed to draw some load in standby mode.

With a small abuse of notations, we write in the following the rectangular profiles strategy by:

$$\boldsymbol{\pi}^{\text{RP}} \quad = \quad \left(t_1^{\text{start}},\ ...,\ t_I^{\text{start}} \right)\ ,$$

which is the vector of starting time at which the load consumption of each i effectively starts, which is controlled by the centralized system operator. The scheduling problem (3.13) to solve can be rewritten in this case as:

Centralized Deterministic Problem - Rectangular Profiles

$$\min_{\boldsymbol{\pi}^{\text{RP}}} \quad \sum_{t\in\mathcal{T}} C_{x_1}^{\boldsymbol{\pi}^{\text{RP}}}(\widetilde{x}_t,\, \boldsymbol{\ell}_t\, ;\, \widetilde{\ell}_{0,t}) \tag{3.17}$$

s.t. (3.3), (3.4), (3.7), (3.8), (3.12), (3.14), (3.15), (3.16) .

Since the state of the system is completely determined under the assumption of the deterministic forecast, the problem (3.17) is reduced to a simple optimization problem. However, rectangular scheduling strategies are not well suited in presence of saturation constraints, such as (3.4) and (3.7), e.g., when the maximal power or the maximal temperature of the DN-transformer could be reached [19][M1]. A suitable scheduling method that can easily integrate this constraints is shown in the next section.

3.3.2 *DYNAMICAL CONSUMPTION STRATEGIES*

In contrast with the previous section, the consumption power does not need to be binary anymore but can take continuous value. Thus, the consumption profile is no longer rectangular and can be arbitrary. Therefore, the power consumption scheduling strategy for consumers does not boil down to a single scalars anymore namely, the time at which the consumption effectively starts. The motivation for this is to have a better performance for the consumers but also to be able to control the system state. In the previous section, the dynamical system was controlled in a one-shot manner, computing the system state values at each instant based on the current information. Here, the state evolution law is taken into account explicitly and the state can be controlled. For instance, it is possible to guarantee that the upper bound constraint on the system state is not violated, which is not well suited with the previous approach.

It turns out that in the problem under investigation (3.13), the system state can be expressed as a sole function of the initial condition x_1, jointly with the deterministic forecast (3.12) (vector of parameters), and the sequence $(\boldsymbol{\ell}_1, ..., \boldsymbol{\ell}_t)$ for each $t \in \mathcal{T}$. This observation allows us to convert the scheduling problem into a standard optimization problem [30], by defining the functions $(g_t)_{t\in\mathcal{T}}$ such that $g_1(x_1) := x_1$ and

$$\begin{aligned} & g_{t+1}(x_1, \boldsymbol{\ell}_1, ..., \boldsymbol{\ell}_t\,;\, \widetilde{\ell}_{0,1}, ..., \widetilde{\ell}_{0,t}) \\ & \qquad := \; f\big(g_t(x_1, \boldsymbol{\ell}_1, ..., \boldsymbol{\ell}_{t-1}\,;\, \widetilde{\ell}_{0,1}, ..., \widetilde{\ell}_{0,t-1}),\, \boldsymbol{\ell}_t;\, \widetilde{\ell}_{0,t}\big)\,. \end{aligned} \tag{3.18}$$

So that, we will have $g_{t+1}(x_1, \boldsymbol{\ell}_1, ..., \boldsymbol{\ell}_t\,;\, \widetilde{\ell}_{0,1}, ..., \widetilde{\ell}_{0,t}) = \widetilde{x}_{t+1}$. In this way, the constraint (3.7) can be rewritten as:

$$g_t(x_1, \boldsymbol{\ell}_1, ..., \boldsymbol{\ell}_{t-1}\,;\, \widetilde{\ell}_{0,1}, ..., \widetilde{\ell}_{0,t-1}) \;\; \leq \;\; x^{\max}\,. \tag{3.19}$$

Writing now the dynamic charging strategy by $\boldsymbol{\pi}^{\mathrm{DC}}$, the problem of power consumption scheduling (3.13) is expressed now as a standard optimization problem:

Centralized Deterministic Problem - Dynamic Charging

$$\begin{aligned} & \min_{\boldsymbol{\pi}^{\mathrm{DC}}} \sum_{t\in\mathcal{T}} C_{x_1}^{\boldsymbol{\pi}^{\mathrm{DC}}}\Big(g_t(x_1, \boldsymbol{\ell}_1, ..., \boldsymbol{\ell}_{t-1}\,;\, \widetilde{\ell}_{0,1}, ..., \widetilde{\ell}_{0,t-1}),\, \boldsymbol{\ell}_t\,;\, \widetilde{\ell}_{0,t}\Big) \\ & \quad \text{s.t.} \quad (3.3),\ (3.4),\ (3.5),\ (3.6),\ (3.12),\ (3.18),\ (3.19)\,. \end{aligned} \tag{3.20}$$

Note that formulating the problem as (3.20) has a potential disadvantage. If the finite horizon T is large, the dimension of the optimal scheduling strategy to be found might make any available numerical optimization routine impossible to be run, which would then necessitate to return to the initial problem formulation (3.13). However, for the numerical applications of interest in Section 3.5, T is typically equals 24 or 48 if the time horizon corresponds to a day and time-slots duration is respectively an hour or half an hour. Considering up to $I = 20$ controllable electric devices per distribution transformer is affordable computationally speaking. Solving the initial scheduling problem for an arbitrary T appears to be an interesting direction to explore. From now on, we consider the standard optimization problem formulation. The following property will be directly exploited in the Section 3.5 for the numerical applications, where standard convex optimization tools are used. Since the composition of convex functions is also convex, and the inequality constraints of the problem above define a convex and compact set, we simply quote the following proposition.

Proposition 3.3.1

The problem (3.20) is a convex optimization problem if $C_{x_1}, g_1, ..., g_t$, and f are convex.

Observe that in terms of information, all the system model parameters need to be known by the single entity that builds the scheduling strategies of consumers (e.g., a centralized scheduling operator). If this turns out to be a critical aspect in terms of identification in practice, e.g., if the distribution network operator (which is not necessarily the centralized scheduling operator) does not want to reveal physical parameters about its DN-transformer. Other techniques which only exploit directly measurable quantities such as the aggregated load consumption can be used. This is one of the purposes of the scheme proposed next.

3.3.3 *VALLEY-FILLING CONSUMPTION STRATEGY*

The method presented in this section replaces the minimization problem (3.13) by the valley-filling charging algorithm. This algorithm is a quite well-known technique (see e.g., [106]) to allocate a given additional energy demand (which corresponds here to the one induced by the consumers) over time, given a primary demand profile (which corresponds here to the noncontrollable load consumption expressed by the deterministic forecast). The idea is to charge when the noncontrollable demand is sufficiently low, e.g., filling the overnight valley, see Figure 3.4 for a graphical representation. This method, is also known as water-filling algorithm, which is traditionally used in communications theory (see e.g., [94]), wherein the method solves the problem of maximizing the mutual information of a communication channel.

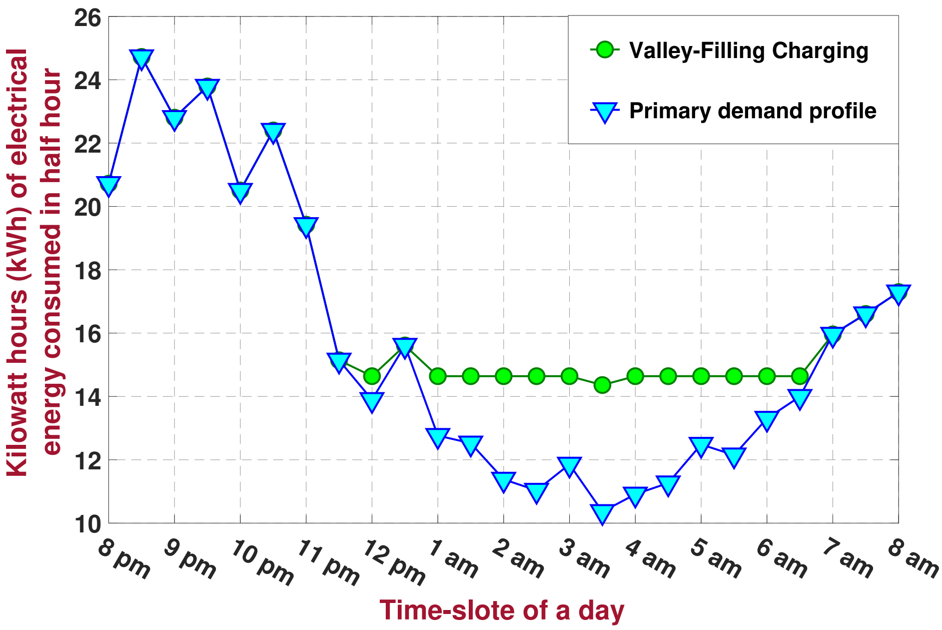


Figure 3.4: Example of filling the overnight valley. We allocate a given additional energy demand by the valley-filling algorithm, over the primary demand profile. The real data was taken from [3].

With a small abuse of notations, this algorithm controls the sequence of the aggregated load profile of consumers, denoted here by:

$$\pi^{\mathrm{VF}} \quad = \quad \left(\sum_{i \in \mathcal{I}} \ell_{i,1}, \; ..., \; \sum_{i \in \mathcal{I}} \ell_{i,T} \right) \, ,$$

to be allocated over the primary demand profile, i.e., the deterministic forecast of the noncontrollable load consumption. Since this method controls at each time the sum of controllable loads, we can write the energy demand constraint (3.5) here as:

$$\Delta_t \sum_{t \in \mathcal{T}} \left(\sum_{i \in \mathcal{I}} \ell_{i,t} \right) \quad \geq \quad \sum_{i \in \mathcal{I}} e_i \, . \tag{3.21}$$

Following the classical definition of this algorithm (see for instance [94, 106]) we write the scheduling problem under this context as follows:

Centralized Deterministic Problem - Valley-Filling

$$\max_{\pi^{\mathrm{VF}}} \quad \sum_{t \in \mathcal{T}} \Phi^{\pi^{\mathrm{VF}}} \left(\widetilde{\ell}_{0,t} + \sum_{i \in \mathcal{I}} \ell_{i,t} \right) \tag{3.22}$$
$$\text{s.t.} \quad (3.6), \; (3.12), \; (3.21) \, .$$

where Φ is any strictly convex function. So that, using Lagrange multipliers method [24, 30], the solution of this problem is given by:

$$\sum_{i \in \mathcal{I}} \ell_{i,t} \quad = \quad \min \left\{ \sum_{i \in \mathcal{I}} \ell_i^{\max}, \; \max \left\{ \sum_{i \in \mathcal{I}} \ell_i^{\min}, \; \mu - \widetilde{\ell}_{0,t} \right\} \right\} \, ,$$

for each time t, where μ is a threshold chosen to satisfy the energy demand constraint (3.21). Compared with the scheduling strategy of Section 3.3.2, an important practical advantage of valley-filling is that it relies only on the measure of the total load consumption. However, both solutions are based on continuous power levels. This assumption may not be met in some real applications. Also, but not least, the valley-filling algorithm does not take into account the impact (modeled by a cost function) on the DN that load consumption operations of consumers generate, this one is based only on the minimization of the aggregated load of consumers and not on any other measure of impact, e.g., the lifetime DN-transformer, electrical consumption payment, etc.

3.4 SOLUTION METHODOLOGY IN THE STOCHASTIC CASE

In the previous sections, the effects of forecast noises on the noncontrollable load consumption has not been taken into account. Indeed, the resulting strategies of rectangular profiles, dynamic charging and valley-filling method, have been designed by assuming a deterministic forecast, which is a single scenario of the noncontrollable loads. Here, the knowledge to schedule the controllable load consumption is based on a stochastic forecast (statistics) on the noncontrollable load consumption. The idea is to take into account forecast errors that have not been considered in the precedent sections, wherein a perfect/imperfect forecast (i.e., a single noiseless/noisy vector of) the noncontrollable load was available to scheduling. The principal motivation of using a stochastic forecast is that it provides a way to model the noncontrollable load consumption by generating several scenarios, so that a deterministic forecast can be seen as one of these. The resulting scheduling strategy can therefore adapt to different noncontrollable load events. Since the DN experiences an increased amount of variable load consumption depending on the controllable part, introducing this kind of (adaptable) scheduling strategy allows consumers to reduce their impacts on the DN [M1]. Otherwise, it could produce new load variations and possibly causing transformer overloading, power losses, or moreover increasing the transformer aging [19].

Following the description in Section 2.4.2, we represent the **stochastic forecast** of the noncontrollable load consumption profile $\boldsymbol{\ell}_0$ by:

$$\widetilde{\boldsymbol{L}}_0 \quad := \quad (\widetilde{L}_{0,t})_{t\in\mathcal{T}} \ , \tag{3.23}$$

which is a finite collection of i.i.d. random variables defined by[1] describing a probability distribution P_0. Suppose now that $(\boldsymbol{L}_t)_{t\in\mathcal{T}}$ is the sequence taking controllable load profiles $\boldsymbol{\ell}_t$, see eq. (3.2), and that for

1 Either the original probability measure of a probability space or describing a probability distribution P_0 can be used to compute probabilities of events involving $\widetilde{\boldsymbol{L}}_0$. Here, we use w.l.o.g. P_0 without needing to refer to a common underlying probability space.

a time t fixed, each component of the sequence (a controllable load vector) is expressed as:

$$\boldsymbol{L}_t \quad = \quad (L_{i,t})_{i\in\mathcal{I}}\ ,$$

where $L_{i,t}$ takes a controllable load value $\ell_{i,t}$, $i = 1, .., I$. In this way, the total load consumption (3.3) can be expressed (stochastically) now as:

$$\widetilde{L}_t \quad = \quad \widetilde{L}_{0,t} \; + \; \sum_{i\in\mathcal{I}} L_{i,t}\ . \tag{3.24}$$

Based on the latter, the induced system state process $(\widetilde{X}_t)_{t\in\mathcal{T}}$, see Definition 2.6.2 for details, has a stochastic behavior due to the forecast $\widetilde{\boldsymbol{L}}_0$ and then, under the probability distribution P_0, we can get an explicit representation of the transition probabilities between the system states. Considering that the system state is $\widetilde{x}_t$ at t and that a controllable load $\boldsymbol{\ell}_t$ has been selected, the probability that the state of the system is $\widetilde{x}_{t+1}$ at time $t+1$ (following the evolution law f of (3.8)), can be computed as follows:

$$\begin{aligned} & P\Big[\widetilde{X}_{t+1} = \widetilde{x}_{t+1} \;\Big|\; \widetilde{X}_t = \widetilde{x}_t,\; \boldsymbol{L}_t = \boldsymbol{\ell}_t\Big] \\ & \qquad = \; P_0\Big[\,\widetilde{L}_{0,t} \in \{\,\widetilde{\ell}_{0,t} \,|\, \widetilde{x}_{t+1} = f(\widetilde{x}_t, \boldsymbol{\ell}_t; \widetilde{\ell}_{0,t})\} \;\Big|\; \widetilde{X}_t = \widetilde{x}_t,\; \boldsymbol{L}_t = \boldsymbol{\ell}_t\Big]\ . \end{aligned} \tag{3.25}$$

Note that the deterministic case shown in Section 3.3, can be see as a particular case of the present one. Indeed, when $\boldsymbol{\ell}_t$ is selected by the centralized operator, we can make the set $\{\,\widetilde{\ell}_{0,t} \,|\, \widetilde{x}_{t+1} = f(\widetilde{x}_t, \boldsymbol{\ell}_t; \widetilde{\ell}_{0,t})\}$ to be a singleton at each t (a set with exactly one element) to represent the deterministic forecast (3.12), thus the next state of x_t is fully determined by making the transition probability to be equal to one for a (next) state and zero for all other candidate next states.

As we have seen in Section 2.6.5, some constraints under the stochastic forecast assumption have been discussed. Here, we accept that an event remains under a certain risk of probability that is coded by a given threshold. Under the same arguments for (3.25), the constraint (3.4) on the total load consumption, and the one on the system state (3.7), can be written resp. as (individual) probability constraints as follows:

$$\begin{aligned} P_0\Big[\,\widetilde{L}_t \;\; &\leq \; \ell^{\max} \;\Big|\; \boldsymbol{L}_t = \boldsymbol{\ell}_t\Big] && \geq \; 1 - \varepsilon_\ell\ , \\ P\Big[\widetilde{X}_{t+1} \; &\leq \; x^{\max} \;\Big|\; \widetilde{X}_t = \widetilde{x}_t,\; \boldsymbol{L}_t = \boldsymbol{\ell}_t\Big] && \geq \; 1 - \varepsilon_x\ , \end{aligned}$$

where $(\varepsilon_\ell, \varepsilon_x) \in [0,1]^2$ represents resp. the risk in probability of exceeding the upper bounds $\ell^{\max}$ and $x^{\max}$. However, the latter probability constraints refer to a situation wherein we wish to satisfy each individual constraint in the stochastic inequalities of the total load consumption and the system state with high enough probability, but we do not make any request in the system trajectory as a whole over the whole time period under consideration. For this reason, each of the above probability constraints is reformulated as one joint probability constraint over all the time period $\mathcal{T}$ and not considered more individually on each t as follows:

$$\prod_{t\in\mathcal{T}} P_0\Big[\widetilde{L}_t \;\leq\; \ell^{\max} \;\Big|\; \boldsymbol{L}_t = \boldsymbol{\ell}_t\Big] \qquad\qquad \geq\; 1-\varepsilon_\ell \tag{3.26}$$

$$\prod_{t\in\mathcal{T}} P\Big[\widetilde{X}_{t+1} \;\leq\; x^{\max} \;\Big|\; \widetilde{X}_t = \widetilde{x}_t,\, \boldsymbol{L}_t = \boldsymbol{\ell}_t\Big] \;\geq\; 1-\varepsilon_x \tag{3.27}$$

where we have used the i.i.d. assumption of the stochastic forecast.

To be consistent with the notation, we write the required energy (3.5) for the controllable electric device i as:

$$\Delta_t \sum_{t\in\mathcal{T}} L_{i,t} \;\geq\; e_i\ , \tag{3.28}$$

with Δ_t to be a time step duration. In addition, we express the constraint (3.6) of not exceeding the maximal power $\ell_i^{\max} \in \mathbb{R}^+$ and to be at least a minimum power $\ell_i^{\min} \in \mathbb{R}_0^+$ at which i can be charging, by:

$$\ell_i^{\min} \;\leq\; L_{i,t} \;\leq\; \ell_i^{\max}\ . \tag{3.29}$$

Based on all practical considerations, the scheduling problem (3.11) can be rewritten under this context as follows:

Centralized Stochastic Problem

$$\min_{\pi} \quad \sum_{t\in\mathcal{T}} \mathbb{E}^{\pi}_{x_1}\Big[C^{\pi}(\widetilde{X}_t,\, \boldsymbol{L}_t\,;\, \widetilde{L}_{0,t})\Big] \tag{3.30}$$

$$\text{s.t.}\;\; (3.23),\; (3.24),\; (3.25),\; (3.26),\; (3.27),\; (3.28),\; (3.29)\ .$$

To be computationally tractable solving this problem, a discretization over the consumption is considered. This problem can thus be modeled using finite Markov Decision Processes [96] when (suitable) statistics are available. This is shown in the next Section.

3.4.1 *MARKOV DECISION PROCESS - BASED APPROACH*

A suitable model to adapt the scheduling problem (3.30) in a discrete manner is a multi-weighted Markov Decision Process (MDP) shown in the Definition 2.6.9. Before to explain the weights or costs considered in the model, we start the analysis with the ordinary MDP model:

$$\mathcal{M} \;=\; \Big(\mathcal{X},\, \mathcal{A},\, P\Big)\ , \tag{3.31}$$

where $\mathcal{X}$ is a finite set of states, $\mathcal{A}$ is a finite space of actions, and P a transition probability between states.

Based on the latter structure, the problem of power consumption scheduling can (a priori) be modeled by considering $\mathcal{X}$ as the set of system states, $\mathcal{A}$ as the space of controllable loads of consumers, and P as the transition probability induced by the probability distribution P_0 of the stochastic forecast (3.23). However, it remains to discuss two things:

(*i*) the way in which the actions are chosen, defining the decisions, the transitions between states, the cost defining the minimization criterion and also the constraints.

(*ii*) how the scheduling strategy for consumers are found and constructed to solve effectively the scheduling problem under consideration.

Consider for instance that a joint action is chosen for consumers. If it is chosen by a single entity (centralized controller), the scheduling problem becomes centralized and no individual decisions are taken into account. That is what we are looking for in this chapter. In fact, our definition of MDP can be seen as something better known as multi-agent MDP [75], which generalizes an ordinary MDP in which several agents exist. However, when we compute an optimal scheduling strategy, this kind of model is indistinguishable from the MDP defined previously if a centralized scheduling operator is considered [85].

On the other hand, we can express several requirements on an MDP according to [M3]. For instance, an upper bound on the total load consumption as a safety condition (that can be encoded as a reachability condition in our MDP model), a constraint of satisfying the energy demand of consumers as quantitative reachability objectives, and various optimization criteria (e.g., minimizing the sum of costs between transitions). In the following, we show how the scheduling problem (3.30) can be modeled and solved using an MDP with several cost functions, namely multi-weighted MDP.

MDP MODELING

Consider that a second component is defined on the state space $\mathcal{X}$ of the MDP (3.31) to keep the information of the total load consumption at each time-step. More specifically, let $\mathcal{L}$ be a (finite) set representing the possible values of the total load (3.24) and define the aggregate space of states as $\mathcal{X} \times \mathcal{L}$. To simplify the analysis, let $\varepsilon_\ell = \varepsilon_x = 0$ from the constraints (3.26) and (3.27). We can thus define each state at time t, $(\widetilde{x}_t, \widetilde{\ell}_t) \in \mathcal{X} \times \mathcal{L}$, as a **deadlock-state** if $\widetilde{x}_t > x^{\max}$ or $\widetilde{\ell}_t > \ell^{\max}$, meaning that the evolution of the system is not propagate if the thresholds $x^{\max}$ or $\ell^{\max}$ are reached for the joint state process $(\widetilde{X}_t, \widetilde{L}_t)_{t\in\mathcal{T}}$, ensuring that the constraints under consideration are satisfied. This can be easily taken into account by defining, in such a case, the probability:

$$P[(\widetilde{X}_{t+1}, \widetilde{L}_{t+1}) = (\widetilde{x}_t, \widetilde{\ell}_t) \mid (\widetilde{X}_t, \widetilde{L}_t) = (\widetilde{x}_t, \widetilde{\ell}_t),\, \boldsymbol{L}_t = \boldsymbol{\ell}_t\,] \;=\; 1\,.$$

Second, since a constraint on the energy demand of consumers (3.28) has to be considered, cost functions on the state transitions can be defined as the energy due by controllable loads, i.e.,

$$C_i(\widetilde{X}_t,\, \widetilde{L}_t,\, \boldsymbol{L}_t,\, \widetilde{L}_{0,t}) \;=\; \Delta_t\, L_{i,t}\;, \tag{3.32}$$

so that each one is defined for each $i \in \mathcal{I}$, and then a (joint) scheduling strategy must satisfy the energy demand constraint on each accumulated sum of such cost. In addition, another cost function can be defined as the cost of the scheduling problem (3.30) to minimize. As we will see in the following, finding an optimal scheduling strategy is reduced to solve the so-called Stochastic Shortest Path (SSP) problem [25], see Definition 2.6.23 for details. Thus globally, an MDP with multiple cost functions can be used to model the scheduling problem under consideration.

While single-weighted MDPs are well-know to be solved in POLYNOMIAL-TIME for the SSP problem, see for instance [14, 25], multi-weighted MDPs subject to constraints imposed on several objectives (that is the case here, e.g., the energy demand for each consumer), take POLYNOMIAL-TIME in the size of the model and EXPONENTIAL in the size of the requirements [50]. To reduce the multi-requirements of energy demand of consumers (3.28) and find an optimal scheduling strategy, we will do two things. First, we reduce the cost functions to a single cost function representing the aggregated sum of energy due to the loads of consumers, so that now the requirement is simplified to satisfy the accumulated energy demand. Thus, the multi-weighted MDP is simplified to be a doubly-weighted MDP, wherein one cost function represents the cost of the scheduling problem (3.30) and the second one the aggregated sum of actions (charge of energy) taken in each time-step. Second, we build a Markov scheduling strategy solving the problem, i.e., a strategy that depends only on the current state and not on the history of the system, see (3.10). This is possible to do by unfolding the doubly-weighted MDP by adding recursively the information of the second cost function on the states reducing the doubly-weighted MDP into a simply-weighted MDP, see Definition 2.6.10. Thus, the new constraint of the aggregated energy demand of consumers is reduced to be a stochastic reachability objective in such a simply-weighted MDP, see Definition 2.6.16. Because we compute an optimal strategy in offline mode, and since by Assumption 2.5.1 the law of the system state depends on the total load consumption, we do not lose any optimality. In addition, once this problem is solved, the allocation problem of the sum energy due to the load of consumers can be tackled. The latter problem is a transportation problem, for which there is an feasible allocation of resources [19]. Possible flow search techniques will not be detailed here. More details can be found, e.g., in [58].

From the Definition 2.6.9, the multi-weighted MDP model to use in our case is of the form:

$$\mathcal{M} \quad = \quad \Big(\mathcal{X} \times \mathcal{L},\, (x_1, 0),\, \mathcal{A},\, P,\, C,\, (C_i)_{i \in \mathcal{I}}\Big)\,, \tag{3.33}$$

where we are fixed the initial state to be $(x_1, 0)$. In the following, we simplify the cost functions $(C_i)_{i\in\mathcal{I}}$, defined individually in (3.32), to be a single cost function between transitions as the aggregated energy due to the sum of controllable loads (actions):

$$C_{\mathcal{I}}(\widetilde{X}_t,\, \widetilde{L}_t,\, \boldsymbol{L}_t,\, \widetilde{L}_{0,t}) \quad = \quad \Delta_t \sum_{i \in \mathcal{I}} L_{i,t}\,. \tag{3.34}$$

In this way, we can define a doubly-weighted MDP with the cost functions C and $C_{\mathcal{I}}$ to reduce the MDP (3.33) to be a doubly-weighted MDP:

$$\mathcal{M} \;=\; \left(\mathcal{X}\times\mathcal{L},\,(x_1,0),\,\mathcal{A},\,P,\,C,\,C_{\mathcal{I}}\right). \tag{3.35}$$

Now, we define the unfolding of the latter MDP model, to simplify the constraint of the aggregated energy demand of the consumers to a quantitative reachability objective into an MDP with only one cost function. This is possible by using the unfolding of $\mathcal{M}$, see Definition 2.6.10.

UNFOLDING THE MDP

The unfolding of $\mathcal{M}$ with depth T used here is the following structure:

$$\mathcal{M}_T \;=\; \left(\mathcal{S}_T,\,s_1,\,\mathcal{A},\,P_T,\,\boldsymbol{C}\right), \tag{3.36}$$

where the space of states is:

$$\mathcal{S}_T \;=\; \mathcal{X}\times\mathcal{L}\times\mathcal{E}\times\mathcal{T}\,,$$

with $\mathcal{E}$ representing the space of the accumulated (and aggregated) energy of consumers, i.e., the set:

$$\mathcal{E} \;=\; \left[0,\,T\sum_{i\in\mathcal{I}}\ell_i^{\max}\right].$$

The given initial state is $s_1 = (x_1, 0, 0, 1)$, the set of actions is defined as the set of sum of loads:

$$\mathcal{A} \;=\; \bigcup_{t\in\mathcal{T}}\left\{\ell_{\mathcal{I},t} \;=\; \sum_{i\in\mathcal{I}}\ell_{i,t}\;\middle|\;\sum_{i\in\mathcal{I}}\ell_i^{\min} \;\le\; \ell_{\mathcal{I},t} \;\le\; \sum_{i\in\mathcal{I}}\ell_i^{\max}\right\}. \tag{3.37}$$

The state transition probability is $P_T : \mathcal{S}_T\times\mathcal{A}\to\mathcal{D}(\mathcal{S}_T)$ is defined as:

$$P_T(s_t,\ell_{\mathcal{I},t})(s_{t+1}) \;=\; P(\mathrm{proj}_1(s_t),\ell_{\mathcal{I},t})(\mathrm{proj}_1(s_{t+1}))$$

if (i) $\mathrm{proj}_1(s_{t+1}) \;=\; f(\mathrm{proj}_1(s_t),\,\ell_{\mathcal{I},t}\,;\,\widetilde{\ell}_{0,t})\,,$
(ii) $\mathrm{proj}_2(s_{t+1}) \;=\; \widetilde{\ell}_{0,t} + \ell_{\mathcal{I},t}\,,$
(iii) $\mathrm{proj}_3(s_{t+1}) \;=\; \min\{\,e_{\mathcal{I}},\,\mathrm{proj}_3(s_t) + \Delta_t\,\ell_{\mathcal{I},t}\,\}$, with

$$e_{\mathcal{I}} \;=\; \sum_{i\in\mathcal{I}} e_i\,, \tag{3.38}$$

(iv) $\mathrm{proj}_4(s_{t+1}) \;\le\; T,$

and $P_T(s_t,\ell_{\mathcal{I},t})(s_{t+1}) \;=\; 0$ otherwise. Note that (i) is implicitly included in the definition (3.25) of P. Finally, the cost function $\boldsymbol{C}$ is the same as C.

Note that we do not have write explicitly the dependence over the cost $C_{\mathcal{I}}$ from the MDP (3.35), because this one represents the aggregated energy due to the sum of controllable loads (actions), see eq. (3.34). Thus, this can simply be written as $\Delta_t\ell_{\mathcal{I},t}$. On the other hand, since we are interested in reaching the accumulated energy demand of consumers, we can naturally define a **set of goal states** $\mathcal{G}$ by:

$$\mathcal{G} \quad := \quad \Big\{ s_t \in \mathcal{S}_T \ \mid \ \mathrm{proj}_3(s_t) = e_{\mathcal{I}} \ \text{ and } \ \mathrm{proj}_4(s_t) = T \Big\} \ . \tag{3.39}$$

This set is of our interest and we want to find a scheduling strategy for which $\mathcal{G}$ is reached, that is to say that the aggregated energy demand $e_{\mathcal{I}}$ is achieved. Note that there is a natural one-to-one correspondence between the "histories or paths" in the doubly-weighted MDP (3.35) and the unfolding (3.36), and therefore, strategies can equivalently be seen in both MDPs[2].

3.5 NUMERICAL APPLICATION

The numerical analysis will be separated in two main sections. The first application is based on a technical approach [70], and the second one on an economical approach [8]. Both are analyzed in the noisy deterministic and stochastic forecast cases. Other method to schedule the power consumption based on the deterministic forecast, is called Plug-and-Charge (PaC), which is one of the most known to schedule power consumption strategies, which is obtained by assuming that the consumption of the controllable electric devices start as soon as they plugin to the grid, minimizing the time needed to reach the cumulative energy demanded. This method is used in this section to compare the performance of the others methods.

SYSTEM STATE

From the evolution law of the system state (3.8), we consider a function parametrized by two values to represent the two approaches jointly. In this section, the system state evolves with the function f defined with a single period time-lag in the load consumption, according to:

$$x_{t+1} \quad = \quad f(x_t,\, \boldsymbol{\ell}_t,\, \ell_{0,t},\, \boldsymbol{\ell}_{t-1},\, \ell_{0,t-1}) \ ,$$

where $x_1 \in \mathbb{R}^+$ is the given initial condition for the system state, $\boldsymbol{\ell}_{t-1}$ is the zero vector in $\mathbb{R}^I$ when $t = 1$ (do not confuse with the real noncontrollable load profile (3.1)), and $\ell_{0,0}$ can be understood as the noncontrollable load consumption of base at $t = 0$. Now, for two parameters $p, q \in \mathbb{N}$, a general representation of the latter function f showing the two approaches that we are interested, has the form:

$$\begin{aligned} & f(x_t, \boldsymbol{\ell}_t, \ell_{0,t}, \boldsymbol{\ell}_{t-1}, \ell_{0,t-1}) \\ & \qquad = \quad \alpha x_t + \beta\Big(\ell_{0,t} + \sum_{i\in\mathcal{I}} \ell_{i,t} \Big)^p + \gamma\Big(\ell_{0,t-1} + \sum_{i\in\mathcal{I}} \ell_{i,t-1} \Big)^q + z_t \ , \end{aligned} \tag{3.40}$$

where $\alpha, \beta, \gamma \in \mathbb{R}$ are constants of the (technical or economical) model, and $z_t \in \mathbb{R}_0^+$ is completely deterministic making always x_t positive. In

2 A formal proof is made in Chapter 4.

particular, the parameters p and q will be of the form $p = q = 1$ for the economical model, and $p = q = 2$ for the technical one.

The simulations are performed over the chosen time (slot) corresponds to $30\,\text{min}$, so that $\Delta_t = 0.5\,\text{h}$, and the DN-transformer is a $20\,\text{kV}/410\,\text{V}$ transformer whose apparent power is $100\,\text{kVA}$ and nominal (active) power is $90\,\text{kW}$, which approximately corresponds to a district of 30 households.

CONTROLLABLE LOAD CONSUMPTION

The controllable load operations occur within the time window from $5\,\text{pm}$ to $7\,\text{am}$ of the next day, i.e., $T = 30$. During the rest of the day (from $7\,\text{am}$ to $5\,\text{pm}$ of the next day), the total load consumption on the DN-transformer only consists of the noncontrollable loads. The controlled devices (consumers) are considered here be a set of Electric Vehicles (EVs), and each $i \in \mathcal{I}$ represents one of them. Unless specified otherwise, the minimum and maximum controllable load induced by one EV i are resp.

$$\ell_i^{\min} = 0\,\text{kW} \qquad \text{and} \qquad \ell_i^{\max} = 3\,\text{kW}\,. \tag{3.41}$$

This is a standard assumption for home charging without any additional connectors to plugin [19, 56]. The arrival and depart time of the EVs are fixed for the simulations and are chosen randomly to be the closest integers of realizations of Gaussian random variables $\mathcal{N}(4, 1.5)$ and $\mathcal{N}(28, 0.75)$ resp. Additionally, the total energy demand of each EV is obtained from the distance to be covered for each of them in the next trip[3]:

$$e_i = \lambda_i \frac{24\,\text{kWh}}{150\,\text{km}} (29.4 + 8)\,\text{km}\,,$$

where $24\,\text{kWh}$ is the capacity of a Renault Zoe battery, $150\,\text{km}$ is the corresponding (average) distance covered, $(29.4 + 8)\,\text{km}$ is the average daily distance covered ($29.4\,\text{km}$ to commute and $8\,\text{km}$ for another trip), and λ_i is the taken margin by EV-users to be confident not running out of energy when driving. The latter is generated randomly once between $\{1.5,\ 2,\ 2.5,\ 3\}$.

NONCONTROLLABLE LOAD CONSUMPTION

To represent the real noncontrollable loads profile $(\ell_{0,t})_{t\in\mathcal{T}}$, four scenarios are used. First, based on historical data taken from the Ausgrid Australian DN-operator for Sydney [3], we choose randomly 30 households representing a district. Second, we use a subtractive clustering method [40] with an influence range of 0.9 to generate four representative clusters, where each one stands for a scenario of the real noncontrollable loads profile, e.g., representing a season of the year. The Figure 3.5 shows a graphical representation of the four scenarios.

One important component is how to assess the impact of not being able to forecast the noncontrollable load consumption perfectly, i.e.,

3 In this way, the energy demand represents approx. the 40%–80% of capacity of a RENAULT Zoe or Fluence EV, similarly to those in [82].

the values of the profile $(\ell_{0,t}^k)_{t\in\mathcal{T}}$ in each scenario $k = 1, ..., 4$. Here, we consider that the forecast can be either deterministic or stochastic, see Section 2.4 for details. When the forecast is deterministic (resp. stochastic), it turns into a noise vector (resp. a random variable), one for each scenario of the noncontrollable loads. In both cases, the noise of the forecast is assumed to be a zero-mean additive white Gaussian noise with known variance σ_k^2. Thus, the models (2.4) and (2.6) are used for the numerical purposes. When the stochastic approach is assumed, a discretization is considered (this will be explained a little further). Considering k fixed, the variance is obtained on the time under consideration $\mathcal{T}$ by **Signal-to-Noise Ratio** (SNR) expressed in decibel (dB), which allows one to measure to what extent the noncontrollable load consumption of the scenario k can be forecasted:

$$\text{SNR} \quad := \quad 10\log_{10}\left(\frac{1}{T\sigma_k^2}\sum_{t\in\mathcal{T}}(\ell_{0,t}^k)^2\right).$$

For example, fixing a $\text{SNR} = 7\,\text{dB}$, we compute σ_k^2 for the scenario k and then, we obtain

(i) a noisy vector of length T for the deterministic forecast wherein each component is a single value from $\mathcal{N}(0, \sigma_k^2)$, and

(ii) a random variable distributing $\mathcal{N}(0, \sigma_k^2)$ to construct a stochastic forecast.

For computational aspects, a discretization over is made on the latter, wherein a normalized histogram is used to obtain relative frequencies. A graphical representation is shown in the Figure 3.6.

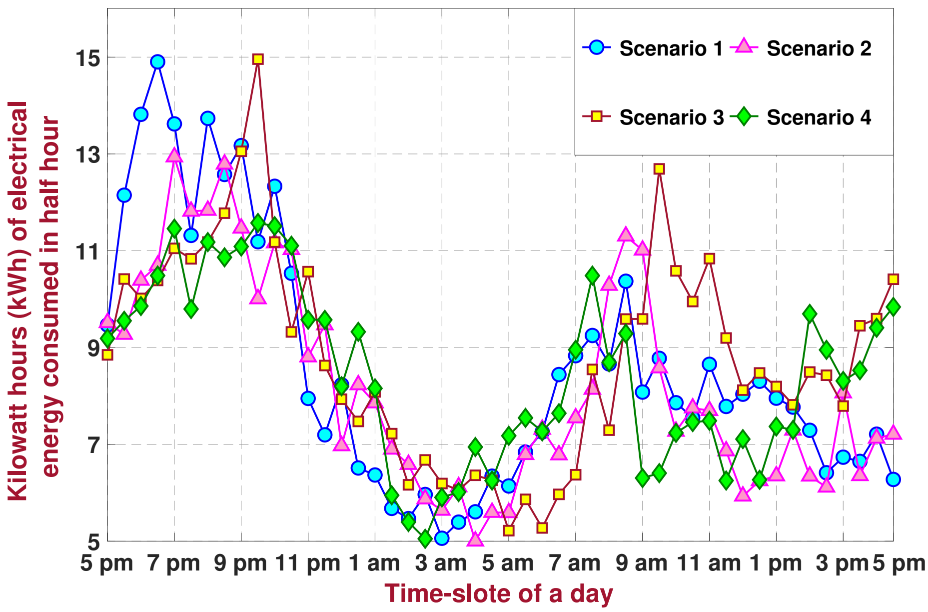


Figure 3.5: Four scenarios of the noncontrollable consumption in a day.

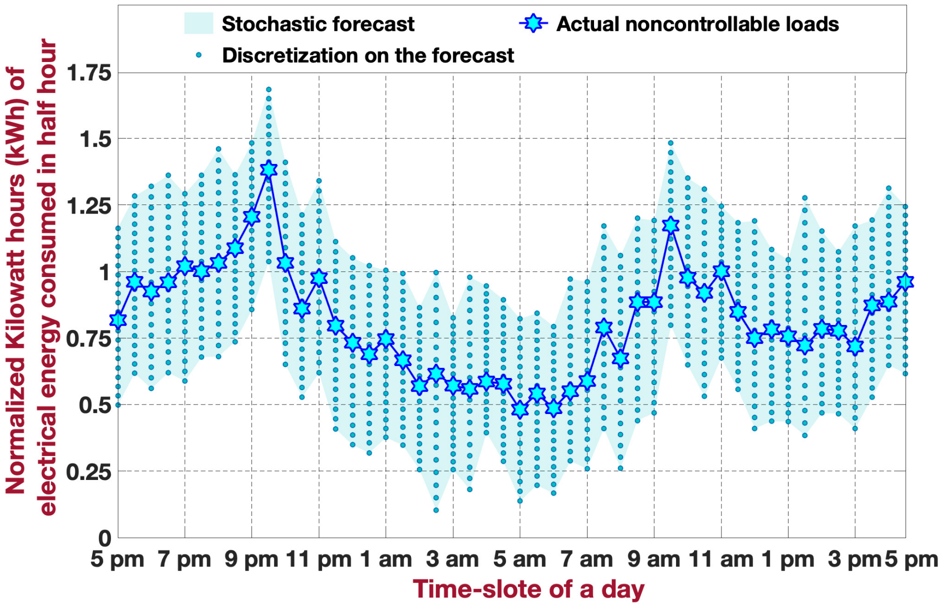


Figure 3.6: Based on a SNR of 7 dB, a stochastic forecast (area filled) of one scenario of the noncontrollable consumption for a day (continuous line), where a discretization is considered (small circles). For a fixed time-slot, each value forescated (a circle) has an associated probability.

DYNAMIC CHARGING SETTINGS

To solve the problem (3.20) to find the dynamic charging strategies, we need the explicit representation of the functions $(g_t)_{t\in\mathcal{T}}$ of (3.18). It is straightforward to show that $(g_t)_{t\in\mathcal{T}}$ with the dynamic law (3.40) for the parameters $p = q = 1$ (economical case) and $p = q = 2$ (technical case), are of the form: $g_1(x_1) = x_1$ and

$$g_{t+1}(x_1, \boldsymbol{\ell}_1, ..., \boldsymbol{\ell}_t\,;\, \widetilde{\ell}_{0,1}, ..., \widetilde{\ell}_{0,t})$$
$$= \alpha^t x_1 + \sum_{\tau=1}^{t} \alpha^{t-\tau}\Big(\beta\Big(\widetilde{\ell}_{0,\tau} + \sum_{i\in\mathcal{I}} \ell_{i,\tau}\Big)^p + \gamma\Big(\widetilde{\ell}_{0,\tau-1} + \sum_{i\in\mathcal{I}} \ell_{i,\tau-1}\Big)^q + z_\tau\Big)$$

Under realistic values of α, β and γ (that we gives in next for the two cases analyzed here), these functions are convex. In particular when $p = q = 2$, the convexity is guaranteed if $\alpha\,\beta + \gamma \geq 0$, which is the case here.

MARKOV DECISION PROCESS SETTINGS

Concerning the MDP used to build the scheduling strategy when a stochastic forecast is considered, we focus on $\mathcal{M}_T$ defined in (3.36). Wherein, the aggregated energy demand (3.38) is modeled as a qualitative objective, more precisely in reaching a set of goal states $\mathcal{G}$ defined in (3.39). On the other hand, the space of actions $\mathcal{A}$ defined in eq. (3.37), is discretized according to the minimum and maximum controllable load induced by EVs (3.41) with a parameter $\Delta_{\mathcal{A}} = 3$, so that this stands as the set $\mathcal{A} = \{0,\, \Delta_{\mathcal{A}}, ...,\, I\Delta_{\mathcal{A}}\}$. The discretization over the stochastic forecast is made with $\Delta_b = 0.05$ to adjust the bin width for the (discrete) probability distribution, wherein normalized histograms are considered to obtain relative frequencies. See Figure 3.6 for a graphical representation of the discretization on the stochastic forecast. Under this practical considerations, the space of states in $\mathcal{M}_T$ is discrete as well.

To solve the MDP under consideration, we use PRISM tool [4, 78], which is a probabilistic model checker, having direct support for MDPs and incorporates the majority of the techniques from [53] to quantify properties specified on MDPs. One of the advantages of PRISM is that it includes multiple efficient engines to describe models that are impractical for the user to explicitly list every state and transitions. In such a way, we just need to specify the parameters and the dynamic (3.40), to generate the MDP $\mathcal{M}_T$.

3.5.1 DN-TRANSFORMER LIFETIME

The goal of this section is to quantify the performance of the different scheduling strategies on the DN-transformer aging. For this, the dynamic law (3.40) represents the Hot-Spot (HS) temperature of the DN-transformer, whose nominal temperature is assumed to be $x_1 = 98\,°\mathrm{C}$, and the shut-down HS temperature is $x^{\max} = 150\,°\mathrm{C}$. The corresponding values of the parameters in (3.40) are as in [70]: $p = q = 2$, $\alpha = 0.83$, $\beta = 30.91\,°\mathrm{C.kW}^{-2}$, $\gamma = -19.09\,°\mathrm{C.kW}^{-2}$ and,

$$z_t \quad = \quad 0.17\,(8.47 + x_t^{\mathrm{amb}})\;,$$

where x_t^{amb} denotes the ambient temperature at time t. Data corresponding to the latter temperature is obtained from the Australian Bureau of Meteorology for the New South Wales territory [5].

To model the DN-transformer lifetime (in years), we consider that when a scheduling strategy $\boldsymbol{\pi}$ is used, it is given by:

$$\mathrm{Lifetime} \quad := \quad T\,\frac{\mathrm{Lifetime}_0}{\sum_{t\in\mathcal{T}} C^{\boldsymbol{\pi}}_{x_1}(x_t, \boldsymbol{\ell}_t, \ell_{0,t})}\;,$$

where the noncontrollable loads consumption is normalized so that without the consumption of the EVs, the DN-transformer lifetime is $\mathrm{Lifetime}_0 = 40\,\mathrm{years}$. In addition, the cost function in this section is considered as the instantaneous Accelerated Aging Factor (AAF), which measures the speed of degradation relatively to the given nominal transformer temperature. A well admitted model for AAF is [68]:

$$C^{\boldsymbol{\pi}}_{x_1}(x_t) \quad = \quad \exp(a\,x_t + b)\;, \tag{3.42}$$

where $a = 0.12\,°\mathrm{C}^{-1}$ and $b = -11.32$.

PERFORMANCE OF THE SCHEDULING STRATEGIES

The Figure 3.7 shows the performance of the power consumption strategies in function of the DN-transformer lifetime (mean over scenarios) against the number of EVs, where the forecast of the noncontrollable part of the total load consumption was assumed to have a noise based on SNR $= 15\,\text{dB}$ in each scenario. Although this noise value is a very ambitious hypothesis (since it represents an "almost-perfect" forecast, see Figure 2.3 for instance), the PaC strategy is seen to be non-acceptable. This strategy induces a significant overload of the transformer, causing a rapid decline in DN-transformer lifetime. This is mainly because the AAF model in eq. (3.42) is exponential in the HS temperature. In addition, under the same scenario for $I = 10$ EVs, the maximum shut-down HS temperature at which the transformer can operate in safe conditions is quickly exceeded as it is shown in the Figure 3.8. The Figure 3.7 also shows the performances of the other power consumption strategies. All of them (with the exception of the PaC strategy) have approximately the same performance in terms of the DN-transformer lifetime, where the black top dashed line corresponds to the case under the real noncontrollable consumption and without EVs.

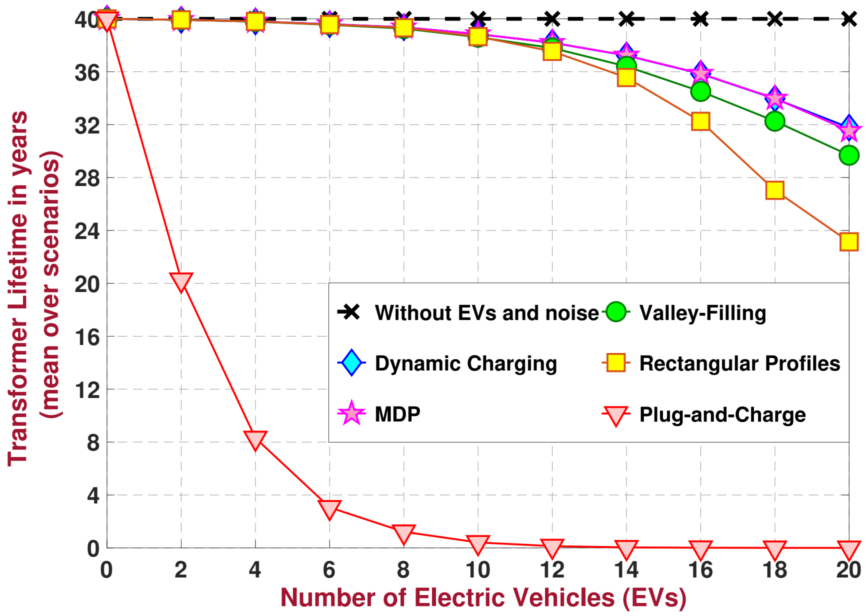


Figure 3.7: DN-transformer lifetime (mean over the scenarios) against the number of EVs (I), based on a forecast of the noncontrollable consumption with a noise (in each scenario) based on SNR $= 15\,\text{dB}$ (centralized case).

We have observed that in the (ambitious hypothesis) of almost-perfect forecast, the dynamic charging becomes optimal. However, for more realistic cases, the strategy built by the MDP method is almost insensitive to forecast noises compared to the other methods. This is confirmed by the Figure 3.9, which represents the DN-transformer lifetime (mean over scenarios) against different values of the forecast noise, under the assumption of $I = 10$ EVs. In particular, we observe that from a forecast noise based on SNR $= 6\,\mathrm{dB}$ upwards, the strategy of rectangular profiles and dynamic charging are robust (in average) to noise. This is also confirmed in detail in each scenario plotted in the Figure 3.10. However, when the forecast noise is based on values lower than SNR $= 6\,\mathrm{dB}$ (i.e., high forecast noises), the performances of these two strategies decreased and are considerably "chaotic". The presence of oscillations on certain curves is here due to the fact that the scenarios are in themselves not to be "smooth curves", i.e., they are stochastic. Although the strategy of rectangular profiles in more robust than the one of dynamic charging (in line with [20]), the strategy built by the MDP method is globally much more robust and the stability of the performance is remarkably guaranteed. This is in line with our practical results of [M1]. On the other hand, the strategy built by the valley-filling method is the more sensitive to amplitude errors. This is caused mainly because the valley-filling algorithm relies only on the minimization of the total consumption and not on any other measure of impact over the DN, like the DN-transformer lifetime. Note that when this strategy "filled the valleys" of the noncontrollable consumption (forecast), the consumption tends to be uniform over time.

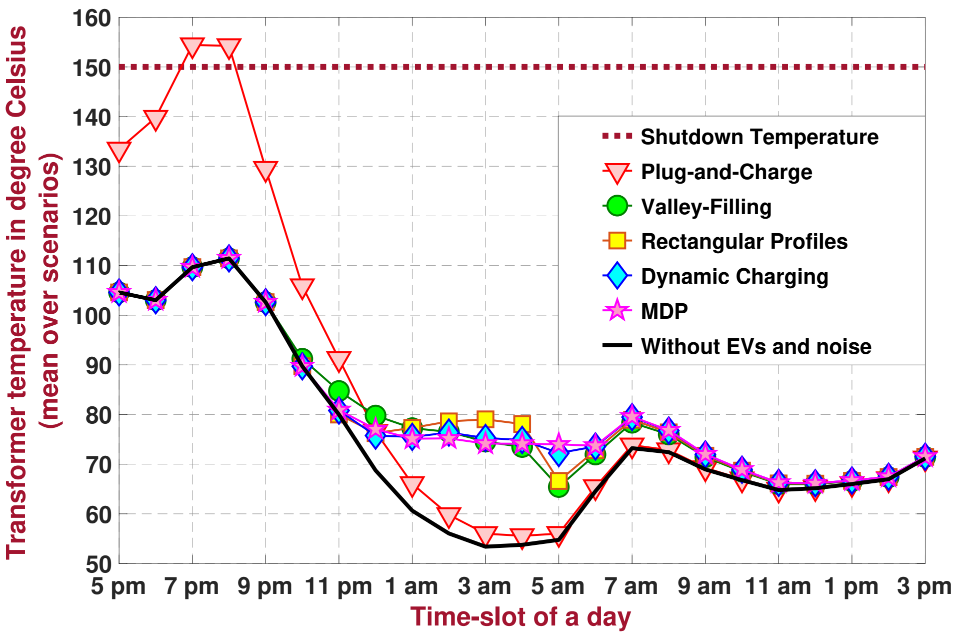


Figure 3.8: HS temperature of the DN-transformer (mean over the scenarios) for $I = 10$ EVs over time, based on a forecast of the noncontrollable consumption with a noise (in each scenario) based on SNR $= 15\,\mathrm{dB}$ (centralized case).

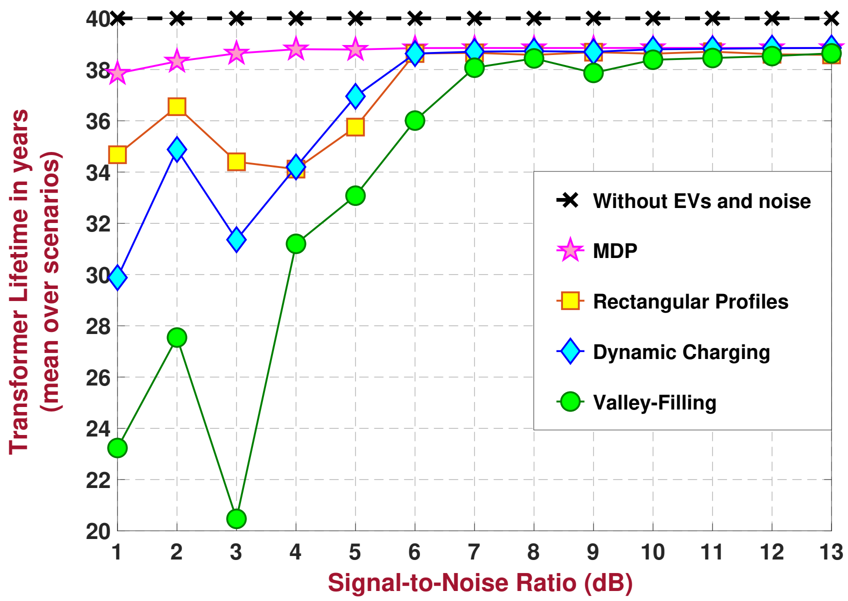


Figure 3.9: DN-transformer lifetime (mean over the scenarios) against forecast noises on the noncontrollable consumption, for $I = 10$ EVs (centralized case).

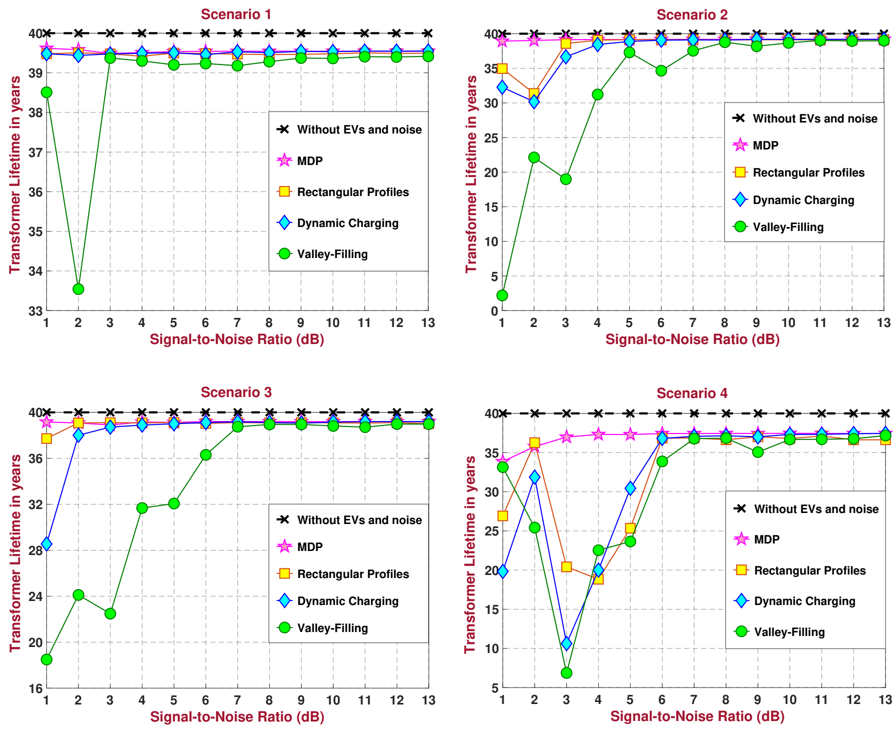


Figure 3.10: DN-transformer lifetime (in each scenario) against forecast noises on the noncontrollable consumption, for $I = 10$ EVs (centralized case).

We also notice that, since the strategy of rectangular profiles is less flexible (since the consumption is uninterrupted), it is also possible to reach the maximum shut-down HS temperature of the DN-transformer by this method as PaC does. This is shown in particular for the scenario 4, which is plotted in Figure 3.11 for $I = 20$ EVs and a forecast noise based on $\text{SNR} = 6\,\text{dB}$. We do not observe this with the others methods.

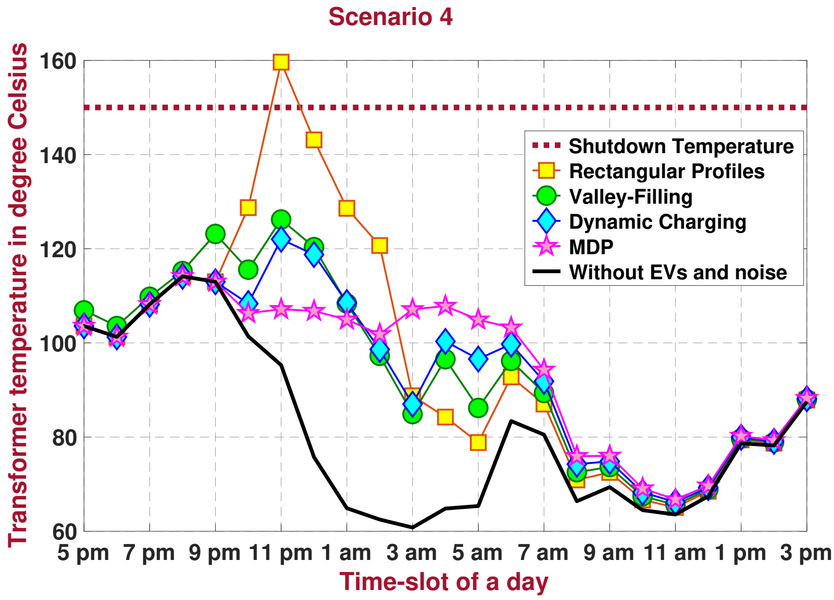


Figure 3.11: HS temperature of the DN-transformer (mean over the scenarios) for $I = 20$ EVs over time, based on a forecast of the noncontrollable consumption with a noise (in each scenario) based on $\mathrm{SNR} = 6\,\mathrm{dB}$ (centralized case).

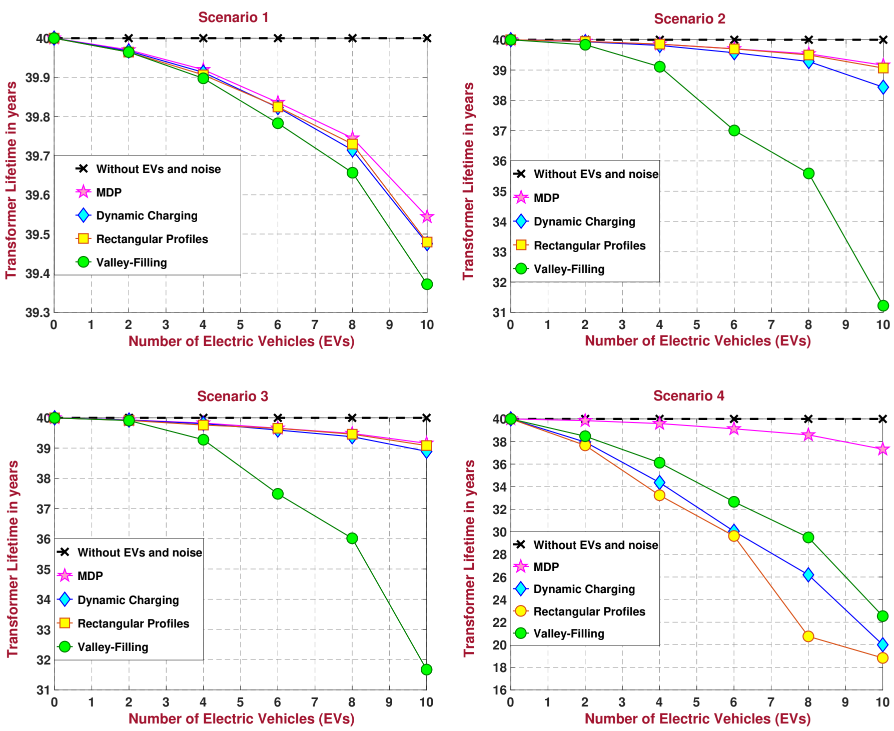


Figure 3.12: DN-transformer lifetime (in each scenario) against the number of EVs, for a forecast noise based $\mathrm{SNR} = 5\,\mathrm{dB}$ (centralized case).

Figure 3.12 shows how the DN-transformer lifetime (in each scenario) decreases when the number of EVs increases. Again, the strategy built by the MDP method is much more robust than the other ones. However, in the case where the number of EVs is high the complexity of the problem increases considerably, since there is a single decision-maker to build the strategies for the consumers. Then, all variables and information are controlled by this centralized controller. To handle this issue, in the Chapter 6 we develop decentralized strategies.

To conclude with the performance analysis of this section, we observe in the Figure 3.13 the probability value of satisfying the constraint (3.27), here for the shut-down HS temperature of the DN-transformer against forecast noises on the noncontrollable consumption. These values were computed a posteriori for the verification of the strategy built by the MDP model. In this case, we have observed that the constraint of exceeding the upper bound of the maximal power is satisfied always with probability one. This is due because the value of the maximal power in the simulations ($\ell^{\max} = 90\,\mathrm{kW}$) was sufficient for all energy demands of EVs. However, the probability of exceeding the maximum HS temperature of the DN-transformer is only satisfied with probability one from a forecast noise based on SNR $= 7\,\mathrm{dB}$, as it is shown in Figure 3.13 which is based on $I = 10$ EVs. This is due mainly because the variance of the stochastic forecast is very high when SNR is approaching zero (i.e., when the forecast noise is high). This is confirmed by the probability values of the dashed line in Figure 3.13, which represents the the values without any EV, i.e., even when there is no EVs, the forecast model of the noncontrollable consumption makes the maximum temperature of the DN-transformer is reached with some probability. However, adding EVs in the consumption, the loss in terms of probability of satisfying such a constraint, is minimal as it is shown in the Figure 3.14.

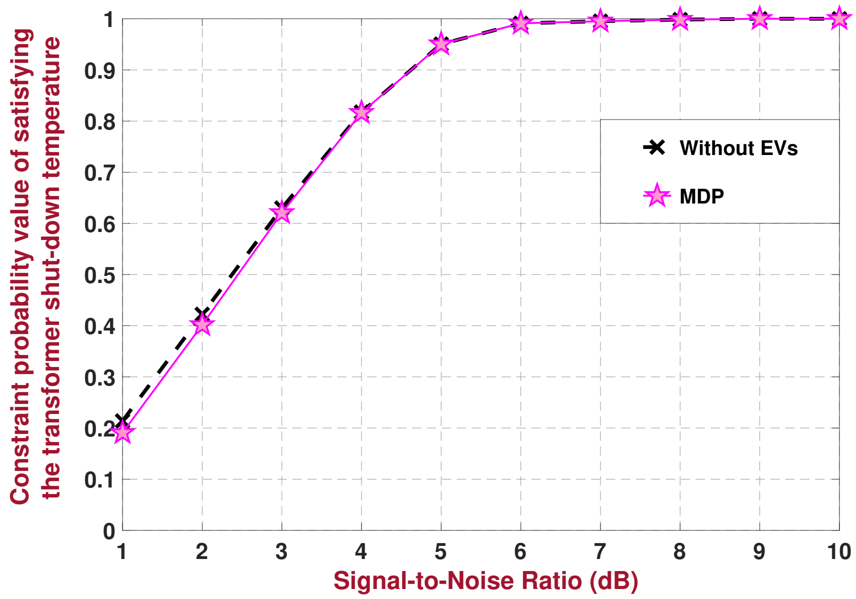


Figure 3.13: Probability value (mean over the scenarios) of satisfying the constraint (3.27) of shut-down HS temperature of the DN-transformer against forecast noises on the noncontrollable consumption, for $I = 10$ EVs (centralized case).

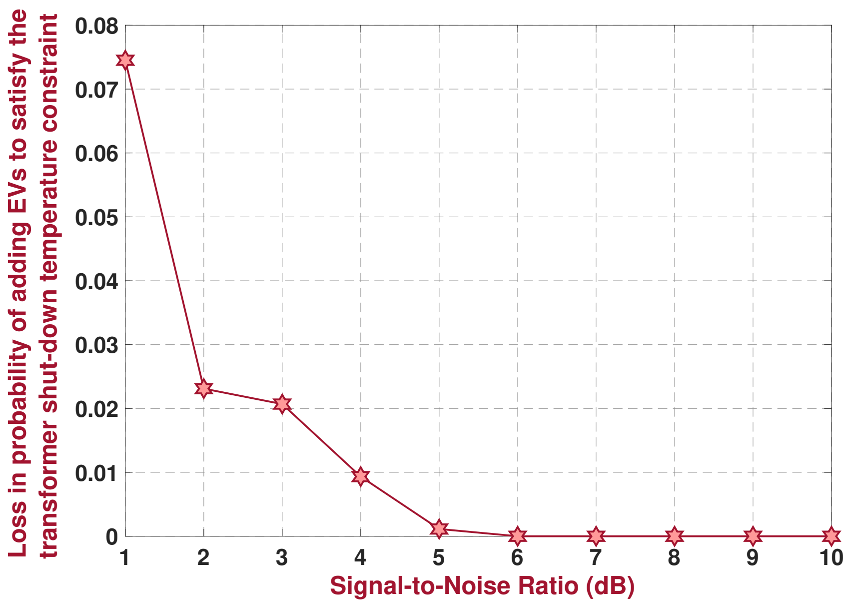


Figure 3.14: Difference between the values of Figure 3.13. It is assumed to have $I = 10$ EVs (centralized case).

3.5.2 *ELECTRICAL CONSUMPTION PAYMENT*

Here, we want to quantify the performance of different scheduling strategies on the Electricity Consumption Payment (ECP). The dynamic of the system state (3.40) represents the EVs total energy monetary cost, whose parameters are [8]: $p = q = 1$, and the other values were estimated by solving a data-fitting problem in least-squares sense from the national electricity market of Australia for New South Wales territory [6]: $\alpha = 1$, $\beta = 11.91\,$¢$.\text{kWh}^{-2}$, $\gamma = -\beta$ and $z_t = 0$. A convenient way of measuring the ECP is formulated by the cost function as [45, 105]:

$$C_{x_1}^{\pi_i}(x_t, \boldsymbol{\ell}_t, \ell_{0,t}) \quad := \quad \left(\ell_{0,t} \;+\; \sum_{t \in \mathcal{T}} \ell_{i,t}\right)^{\eta} x_t\,,$$

where $\eta = 3$ is a coefficient indicating the impact of nonlinearity of the total load consumption. See [26] for details, and the initial electrical price is assumed to be $x_1 = 39.65\,$¢ taken from the data.

PERFORMANCE OF THE SCHEDULING STRATEGIES

Here, we start with the analysis of the Figure 3.15 and Figure 3.16. These figures show that the ratio of the electrical consumption payment of the different strategies for the EVs, with respect to the case without EVs and noise, under the assumption of a forecast noise based on SNR $= 5\,$dB, which is a realistic hypothesis of noise (with high values of SNR the forecast becomes “perfect”). The Figure 3.15 shows the performances in average over the four scenarios plotted in the Figure 3.16. Of course, when the number of EVs increases, the electrical consumption payment also increases. We observe that the strategy built by the MDP method has the best performance (in average and in each particular scenario) for each pool of EVs considered. Remarkably, the performance by the valley filling algorithm decreases faster than the other methods. This could be caused because this strategy tends to be a uniform consummation when it has already "filled the valleys" of the noncontrollable consumption (forecast). This is mainly confirmed by the Figure 3.17 and Figure 3.18, which show the performances resp. for $I = 10$ and $I = 20$ EVs. In both cases, the strategy by valley-filling becomes independent of the noise, consuming uniformly over time.

We notice that the PaC strategy is seen to be non-acceptable, because the payment is the worst, as it is shown in the Figure 3.17. Remarkably, in the Figure 3.17 and Figure 3.18, we observe that the performance of the strategy built by the MDP method is stable and the strategy is much more robust than the others. Interesting, when the forecast noise begins to decrease (high values of the SNR), the dynamic charging strategy becomes optimal. This is also observed for the strategy pf rectangular profiles for $I = 10$ EVs. However, this is not the case when the number of EVs increases to $I = 20$. In such a case, the performance of the strategy by rectangular profiles begins to oscillate and is not stable. This can occur, e.g., because the scenarios are in themselves stochastic and not be "smooth curves" to schedule the rectangular profiles.

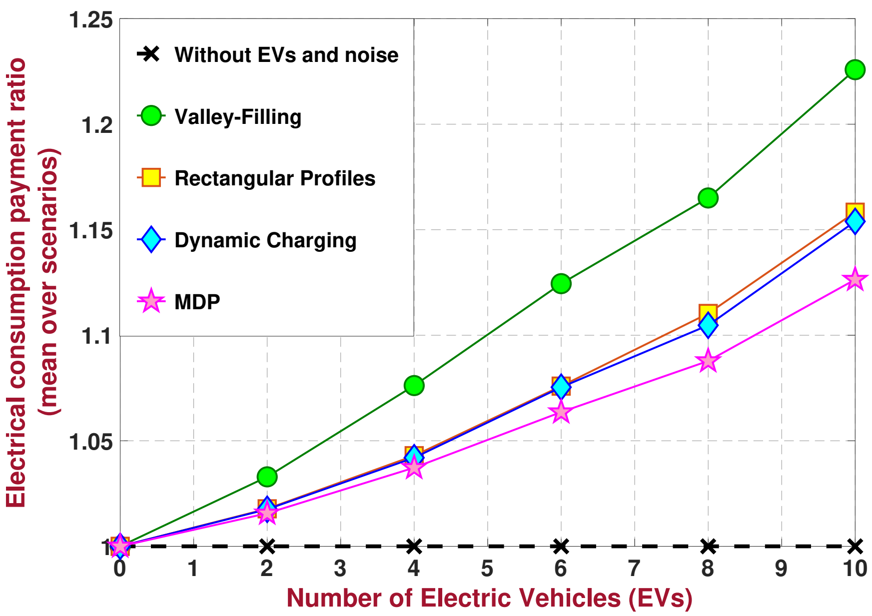


Figure 3.15: Ratio of the electrical consumption payment (mean over the scenarios) against the number of EVs, with respect to the case without EVs and noise, under the assumption of a forecast noise based on SNR = 5 dB (centralized case).

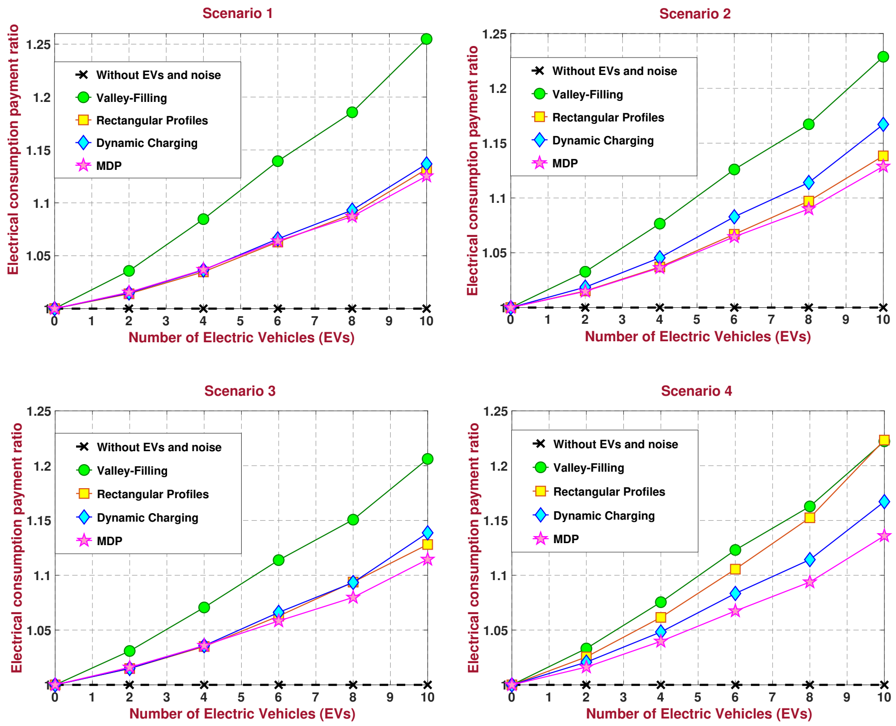


Figure 3.16: Ratio of the electrical consumption payment (in each scenario) against the number of EVs, with respect to the case without EVs and noise, under the assumption of a forecast noise based on SNR = 5 dB (centralized case).

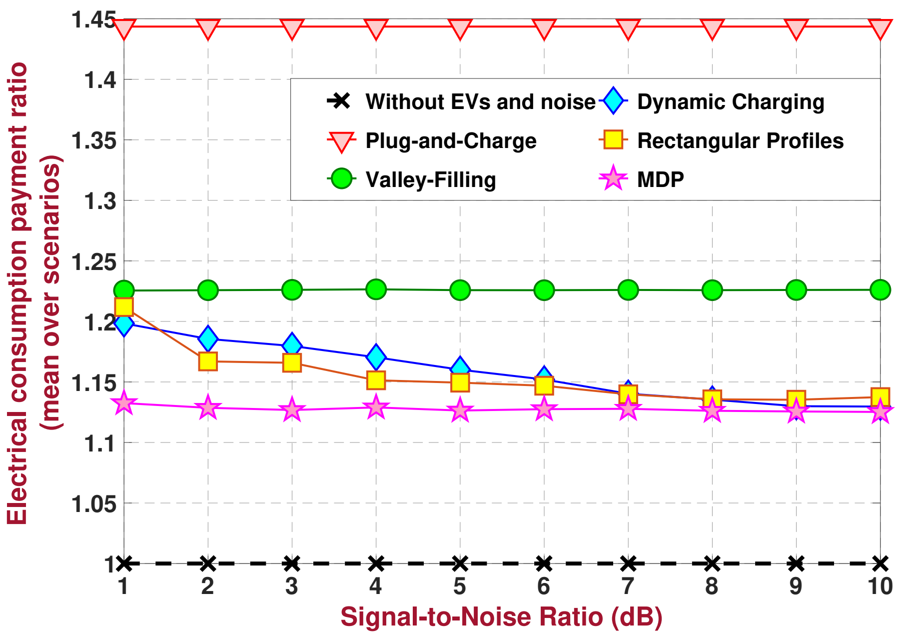


Figure 3.17: Ratio with respect to the case without EVs and noise, of the electrical consumption payment (mean over the scenarios) against forecast noises on the noncontrollable consumption, for for $I = 10$ EVs (centralized case).

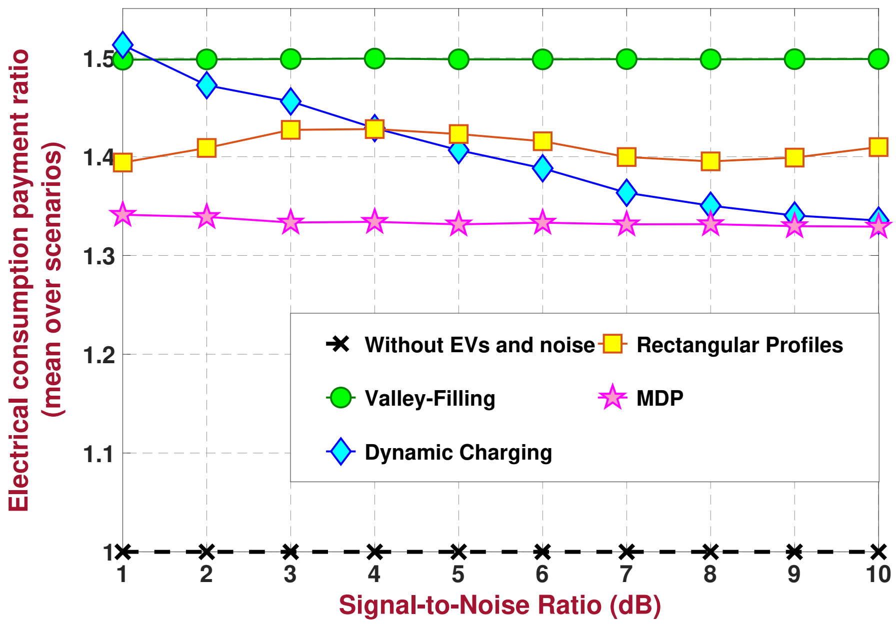


Figure 3.18: Ratio with respect to the case without EVs and noise, of the electrical consumption payment (mean over the scenarios) against forecast noises on the noncontrollable consumption, for for $I = 20$ EVs (centralized case).

3.6 DISCUSSION

A useful observation of the dynamic charging scheme in Section 3.3.2 together with the (2.5.1), is that such a problem formulation (3.20) can be seen as that it only depends on the sequence of the sums of the controllable loads of consumers, i.e., the sequence:

$$\left(\sum_{i \in \mathcal{I}} \ell_{i,1}, \ \ldots, \ \sum_{i \in \mathcal{I}} \ell_{i,T} \right) . \tag{3.43}$$

Then the centralized decision-maker can be equivalently solve the problem (3.20) in two steps:

(*i*) Find an optimal sequence of controllable load sum (3.43).

(*ii*) Allocate each sum of the sequence among the consumers.

The optimization problem associated with the determination of an optimal sequence of controllable load sum is directly derived from the optimization problem (3.20) by introducing for each $t \in \mathcal{T}$ the variables:

$$\ell_{\mathcal{I},t} \ = \ \sum_{i \in \mathcal{I}} \ell_{i,t}$$

and, from the eqs. (3.18), the functions $g_{\mathcal{I},1}(x_1) = g_1(x_1)$ and

$$g_{\mathcal{I},t+1}(x_1, \ell_{\mathcal{I},1}, \ldots, \ell_{\mathcal{I},t}, ; \widetilde{\ell}_{0,1}, \ldots, \widetilde{\ell}_{0,t}) \ = \ g_{t+1}(x_1, \boldsymbol{\ell}_1, \ldots, \boldsymbol{\ell}_t, ; \widetilde{\ell}_{0,1}, \ldots, \widetilde{\ell}_{0,t}) \ .$$

Thus, the constraints (3.5) and (3.6) are also replaced resp. by:

$$\Delta_t \sum_{t \in \mathcal{T}} \ell_{\mathcal{I},t} \;\geq\; \sum_{i \in \mathcal{I}} e_i$$

and

$$\sum_{i \in \mathcal{I}} \ell_i^{\min} \;\leq\; \ell_{\mathcal{I},t} \;\leq\; \sum_{i \in \mathcal{I}} \ell_i^{\max} \,.$$

For the applications studied in Section 3.5, the cost function minimized in the optimization problem (3.20) is continuous and convex, and the inequality constraints define a convex and compact set. Thus, there is an unique solution of the optimization problem associated to the sequence of controllable load sum.

Note that the controllable variables (sum of loads) for the dynamic charging problem in this case, are the same as the ones of the valley-filling algorithm in the Section 3.3.3 and the MDP method in the Section 3.4.1. Once this problems are solved, the allocation problem of the sums can be tackled. The latter is a transportation problem where the "sources" are the T time-slots with $\ell_{\mathcal{I},t}$ supply units, the "destinations" are the I consumers with e_i units received and each $\ell_{i,t}$ represents the "flow" from the time t (the source) to consumer i (destination), see e.g., [110] for details.

All these resulting schemes developed in the deterministic and stochastic models have a complexity issue due to the degree of information used to schedule in a centralized way, in particular when the time-horizon and the number of consumers are large. Remarkably, the decentralized solutions we propose in the Chapter 6 solve the power consumption scheduling problem by construction, and transportation theoretic tools are not necessary.

To end the performance comparison analysis and influence of the forecast noise, studied in the numerical applications of the Section 3.5, we can conclude that the strategy built by the MDP model are almost insensitive to forecast noise. Particularly, when this is high, the performance of the others methods decreased considerably and the MDP method is globally much more robust. When the forecasting noise is close to zero, i.e., under a "perfect forecast" (which is a very ambitious hypothesis), the dynamic charging method becomes optimal.

4

UNDERLYING GENERAL PROBLEM

Abstract:

In this chapter, we focus on the existence of scheduling strategies in a multi-weighted Markov decision process, which satisfy both a probability constraint over a reachability condition, and a quantitative constraint on the expected value of a random variable defined also using the reachability condition. This problem is inspired by the modelization of the power consumption scheduling problem of the previous chapter (typically the problem of charging electric vehicles). Focused on a general version of such a scheduling problem, we investigate the "cartography" of the problem when one parameter varies (a threshold), i.e., the set of values of such a parameter for which the problem has or not a solution; and show how a partial cartography can be obtained via two sequences of optimization problems. We also discuss completeness and feasibility of the resulting approach.

LIST OF ABBREVIATIONS AND SYMBOLS

MDP	Markov Decision Process	SSP	Stochastic Shortest Path

a_t	action at t
$\mathcal{A}$	finite action space in $\mathcal{M}$
$\mathcal{B}(\Omega_{x_0})$	Borel sigma-algebra over Ω_{x_0}
C_i	i^{th} cost function
$C_i^{\min}$	minimum C_i cost in the transitions of $\mathcal{M}$
$C_i^{\max}$	maximum C_i cost in the transitions of $\mathcal{M}$
δ	randomized strategy
δ_ε	randomized strategy for problem $\mathbf{Pb}(\varepsilon)$
δ_γ^*	γ-optimal randomized strategy for $\mathbb{J}_{\alpha,T}^*$
Δ	set of randomized strategies
$\mathcal{D}(\mathcal{S})$	set of probability distributions over a finite set $\mathcal{S}$
$\mathbb{E}_{x_0}^\delta$	expectation operator under a fixed δ and x_0
$\mathcal{E}_T$	set of paths reaching $\mathcal{G}$ in at most T-steps
$\overline{\mathcal{E}}_T$	complement set $\mathcal{E}_T$
$\overline{\mathcal{E}}_{\alpha,T}$	to be $\overline{\mathcal{E}}_T$ for $\alpha = 1$ and $\varnothing$ otherwise
ε	non-negative risk in probability (parameter)
$\mathcal{G}$	set of goal states in an MDP model
$\mathrm{id}_{\mathcal{A}}$	identity function on a set $\mathcal{A}$
$\mathbb{J}_{\alpha,T}$	objective function of a minimization problem
$\mathbb{J}_{\alpha,T}^*$	optimal value of $\mathbb{J}_{\alpha,T}$
$\mathcal{M}$	MDP model
$\mathcal{M}_T$	unfolding of $\mathcal{M}$ with depth T
ω	history or path of the system
Ω_{x_0}	set of histories or paths of a system from an initial state x_0
proj_j	projection function on the j-component of a sequence
$\mathrm{proj}_t^{\mathcal{X}}$	projection function on $\mathcal{X}$ in the t-component of a path
π	pure strategy
$\underline{\Pi}$	set of pure Markov strategies
P, P_T	transition probability between states in $\mathcal{M}$ and $\mathcal{M}_T$, resp.
$\mathbb{P}_{x_0}^\delta$	probability measure on $\mathcal{B}(\Omega_{x_0})$ under a fixed δ and x_0
$\mathbf{Pb}(\varepsilon)$	general problem in function of ε
$\mathcal{S}_T$	augmented space of states in $\mathcal{M}_T$
s_t	state in $\mathcal{S}_T$ at t
T	finite horizon, or depth
$\mathrm{TS}_i^{\mathcal{G}}$	truncated sum function for $\mathcal{G}$ under C_i
x_t	system state at t
$\mathcal{X}$	finite set of states in $\mathcal{M}$

4.1 MOTIVATION AND CONTRIBUTIONS

The principal motivation of the present chapter is founded on formal methods, which can help providing algorithmic solutions and guarantees for control designs and applications, as in Smart grids. The controllable electric devices scheduling problem is an example of such an application area. This problem, usually presented as a control problem (see for instance [20, 103] for the electric vehicle charging context), can actually be modeled using Markov Decision Processes (MDP) [M1][49]. In addition, probabilities provide a way to model the non-controllable part of the total load consumption on the distribution network, i.e., the consumption outside the controllable electric devices, for which large databases exist in order to extract precise statistics (see for example [3]). We can then express an upper bound on the peak load consumption as a safety condition (encoded as a reachability condition in our finite-horizon model), a constraint on charging all the controllable electric devices as a quantitative reachability objective, and various optimization criteria (e.g., minimizing the aging of the distribution network transformer [19], or the electrical consumption payment [74]) as an optimization of random variables (cost functions). However, the computability of an optimal strategy in such a general context, as well as the corresponding decision problem, are unexplored.

The general multi-constrained problem, arising from the scheduling problem of Section 3.4, takes as input a double-weighted MDP, and requires the existence of a strategy ensuring:

(i) a **reachability objective** in the MDP, i.e., some set of goal states is reached,

(ii) a **probability constraint** on the accumulated sum of the first cost function (lower bound parametrized by some threshold), and

(iii) an **expectation objective** on the accumulated sum of the second cost function (upper bound parametrized by another threshold).

The initial scheduling problem corresponds to the instance of this problem when the first threshold in the probability constraint is zero, and the cost functions for the constraints represent resp. the energy needed of the controllable electric devices and the impact of the load consumption operation on the distribution network.

Our problem integrates both the “beyond worst-case” paradigm of [34], and the mixture of probability and expectation constraints as in [37]. While for the latter, linear programs are used for solving the problem, we need different techniques in our general problem. Here, we develop a methodology to describe the cartography of the problem, for which it has a solution. Our approach is based on two sequences of optimization problems which, in some cases we characterize, allow us to have an almost-complete full picture of the solution. We then discuss computability issues.

4.2 PROBLEM FORMULATION

In this chapter, we focus on doubly-weighted MDPs (see Definition 2.6.9 taking into account two cost functions) with a combination of:

(i) a reachability objective to a certain set of goal states, see Definition 2.6.16,

(ii) a probability constraint on the proportion of paths having a high accumulated value under the first cost function, see Definition 2.6.20, and

(iii) a constraint on the expected value of the truncated sum defined by the second cost function, see Definition 2.6.23.

We aim to design an algorithmic technique to identify when the general problem presented in the following has or has not a solution, we consider then all the variables of the problem as rational numbers, due to computational problems arising with irrational ones.

Definition 4.2.1: General Problem

Given a doubly-weighted MDP $\mathcal{M} = (\mathcal{X}, \mathcal{A}, P, (C_i)_{i=1}^2)$, an initial state $x_0 \in \mathcal{X}$, goal set[a] $\mathcal{G} \subset \mathcal{X}$, and $\nu_1, \nu_2 \in \mathbb{Q}$ two thresholds. For each $\varepsilon \in [0,1] \cap \mathbb{Q}$, the general problem $\mathbf{Pb}(\varepsilon)$ is defined as[b]: there exists $\delta_\varepsilon \in \Delta$, such that each $\omega \in \Omega_{x_0}^{\delta_\varepsilon}$ reaches $\mathcal{G}$,

$$\mathbb{P}_{x_0}^{\delta_\varepsilon}[\, \mathrm{TS}_1^{\mathcal{G}} \geq \nu_1 \,] \geq 1 - \varepsilon \qquad \text{and} \qquad \mathbb{E}_{x_0}^{\delta_\varepsilon}[\, \mathrm{TS}_2^{\mathcal{G}} \,] < \nu_2$$

a The goal set $\mathcal{G}$ is absorbent by Assumption 2.6.8.
b To be more precise, the problem should be denoted as $\mathbf{Pb}_{\mathcal{M},\nu_1,\nu_2}(\varepsilon)$, because this one depends on the fixed $\mathcal{M}$, ν_1 and ν_2. However, to relax the notation, we simply write $\mathbf{Pb}(\varepsilon)$.

We aim at computing the values of ε for which $\mathbf{Pb}(\varepsilon)$ has a solution. For readability, we present the following example.

Example 4.2.2. *Consider the double-weighted MDP depicted in the Figure 4.1, and* $\varepsilon = 0.5$*,* $\nu_1 = 1$*, and* $\nu_2 = 4.3$*. Let* δ *the strategy that selects* a_0 *or* b_0 *uniformly at random in* x_0 *and always selects* a_1 *in* x_1*. Then, the probability of reaching* x *is one, i.e.,* $\mathbb{P}^{\delta}_{x_0}[\,T_x < +\infty\,] = 1$*, where* T_x *is the reachability time (2.23), and*

$$\mathbb{P}^{\delta}_{x_0}[\,\mathrm{TS}^{\mathcal{G}}_1 \geq \nu_1\,] = \frac{1}{2} \quad \text{and} \quad \mathbb{E}^{\delta}_{x_0}[\,\mathrm{TS}^{\mathcal{G}}_2\,] = \frac{1}{2}5 + \frac{1}{2}\sum_{t=1}^{+\infty}\frac{-t}{2^t} = \frac{3}{2} < \nu_2$$

The latter strategy is Markov, because it does not require any memory of finite paths. On the contrary, if we consider now the strategy that makes at least half of the paths reach x *(impacting* 2.5 *over the expectation of* $\mathrm{TS}^{\mathcal{G}}_2$*). The other paths have to go to* x_1*, and then take* a_1 *for some time (provided the path goes back to* x_1*) in order to decrease the expectation of* $\mathrm{TS}^{\mathcal{G}}_2$*, before it becomes possible to take* b_1 *and then* a_2 *(so that the strategy is surely winning). This strategy uses both randomization (at* x_0*) and memory (counting the number of times* a_1 *is taken before* b_1 *can be taken).* □

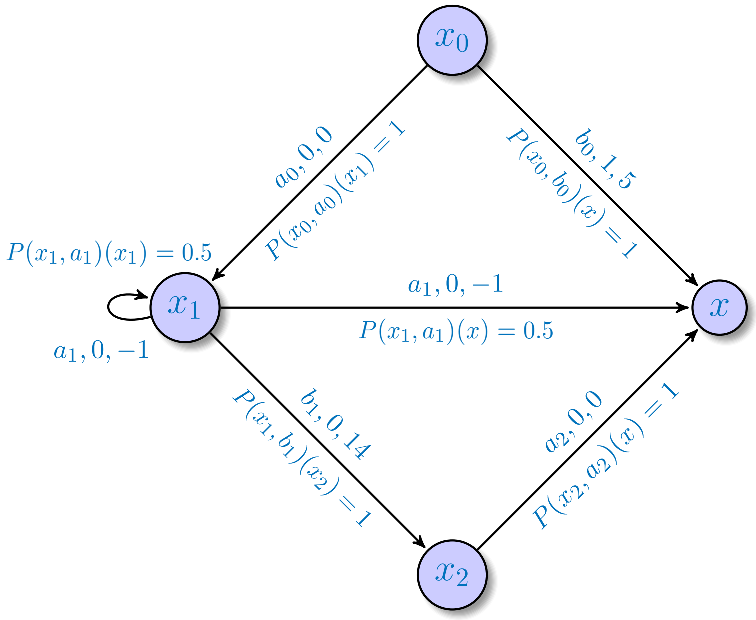


Figure 4.1: An example of a doubly-weighted MDP, where the initial state is x_0 and the goal set is $\mathcal{G} = \{x\}$. All the transitions have probability one, except the transition from x_1 using the action a_1, which has a uniform distribution between the two possible next states. On the edges, the 3-tuple is composed by an action, the first and the second cost function.

For the rest of this section, we make the following assumption.

Assumption 4.2.3: Feasibility

There exists a strategy $\pi \in \underline{\Pi}$, such that $\mathbb{E}^{\pi}_{x_0}[\mathrm{TS}^{\mathcal{G}}_2] < \nu_2$.

Otherwise, the problem $\mathbf{Pb}(\varepsilon)$ trivially has no solution for each ε. Note that the latter assumption is reduced to the so-called Stochastic Shortest Path (SSP) problem, wherein the existence of a pure Markov strategy $\pi \in \Pi$ can be built in POLYNOMIAL-TIME [14].

Definition 4.2.4: Cartography

We call **cartography** of the problem $\mathbf{Pb}(\varepsilon)$ the function which associates to each $\varepsilon \in [0, 1]$, either TRUE if $\mathbf{Pb}(\varepsilon)$ has a solution, or FALSE otherwise.

We will describe an algorithmic technique to map the cartography function of the problem $\mathbf{Pb}(\varepsilon)$ on the interval $[0, 1]$. As we explain below, our approach partially characterizes such an interval. That is why we call such a map **approximated cartography** in what follows.

4.3 APPROXIMATED CARTOGRAPHY

Let $\mathcal{M} = (\mathcal{X}, \mathcal{A}, P, (C_i)_{i=1}^2)$ be a fixed doubly-weighted MDP, $x_0 \in \mathcal{X}$ an initial state, $\mathcal{G} \subset \mathcal{X}$ be a goal set, and two thresholds $\nu_1, \nu_2 \in \mathbb{Q}$. We will introduce two optimization problems related to the problem $\mathbf{Pb}(\varepsilon)$, from which we derive information on the values of a parameter ε, more precisely, for which such a problem has or has not a solution. As we explain below, our approach partially characterizes the interval $[0, 1]$, however under certain conditions in the structure of $\mathcal{M}$, the cartography is almost-complete [M3]. In what follows, we use a solution of the underlying optimization problems, to characterize the solution of $\mathbf{Pb}(\varepsilon)$. About a method to find such solutions is mentioned below and developed in Chapter 5.

4.3.1 *OPTIMIZATION PROBLEMS*

From the Definition 4.2.1 of $\mathbf{Pb}(\varepsilon)$, we are looking for an event over the paths that are reaching the goal set $\mathcal{G}$ with a truncated sum on the first cost function of at least ν_1, and satisfying an expected value on the truncated sum on the second cost function lower than ν_2. Moreover, we want the probability of such an event is be at least $1 - \varepsilon$. To compute the values of such ε, we consider the induced probability space from $\mathcal{M}$ as $(\Omega_{x_0}, \mathcal{B}(\Omega_{x_0}), \mathbb{P}_{x_0})$, see Section 2.6.4 for details. For $T \in \mathbb{N}_0$ we define $\mathcal{E}_T \in \mathcal{B}(\Omega_{x_0})$ by:

$$\mathcal{E}_T \;:=\; \left\{ \omega \in \Omega_{x_0} \;\mid\; \exists t \leq T \;:\; \mathrm{proj}_t^{\mathcal{X}}(\omega) \in \mathcal{G} \right\}, \tag{4.1}$$

i.e., $\mathcal{E}_T$ represents the measurable set of paths reaching the goal set $\mathcal{G}$ in at most T-steps. Additionally, since $\mathcal{G}$ is absorbent by assumption, any path $\omega \in \mathcal{E}_T$ also belongs to $\mathcal{E}_{T+1}$. Thus,

$$\mathcal{E}_0 \;\subseteq\; \mathcal{E}_1 \;\subseteq\; \dots \;\subseteq\; \mathcal{E}_T \;\subseteq\; \mathcal{E}_{T+1} \;\subseteq\; \dots \tag{4.2}$$

If the limit of $\mathcal{E}_T$ exists as $T \to +\infty$, we let:

$$\mathcal{E}_\infty \;:=\; \lim_{T \to +\infty} \mathcal{E}_T \,.$$

On the other hand, since $\mathcal{B}(\Omega_{x_0})$ is a well-defined sigma-algebra, the complement set of $\mathcal{E}_T$ exists within $\mathcal{B}(\Omega_{x_0})$ and then, it is measurable under the probability distribution $\mathbb{P}_{x_0}$. We denote the complement of $\mathcal{E}_T$ as $\overline{\mathcal{E}}_T$. The latter is understood as the set of paths that are not reaching $\mathcal{G}$ in the first T-steps, i.e.,

$$\overline{\mathcal{E}}_T \;=\; \left\{ \omega \in \Omega_{x_0} \;\mid\; \mathrm{proj}_t^{\mathcal{X}}(\omega) \notin \mathcal{G} \;,\;\; \forall\, t \leq T \right\} . \tag{4.3}$$

Analogously, if a path $\omega \in \overline{\mathcal{E}}_{T+1}$, then it is also not reaching $\mathcal{G}$ in the first T-steps. Thus,

$$\overline{\mathcal{E}}_0 \;\supseteq\; \overline{\mathcal{E}}_1 \;\supseteq\; \dots \;\supseteq\; \overline{\mathcal{E}}_T \;\supseteq\; \overline{\mathcal{E}}_{T+1} \;\supseteq\; \dots \tag{4.4}$$

If the limit of $\overline{\mathcal{E}}_T$ exists as $T \to +\infty$, we let:

$$\overline{\mathcal{E}}_\infty \;:=\; \lim_{T \to +\infty} \overline{\mathcal{E}}_T \;,$$

which represents the set of states that never visit the goal set $\mathcal{G}$.

Remark 4.3.1: Reaching the Goal Surely

Note that if there is a strategy δ allows us to reach $\mathcal{G}$ surely, see Definition 2.6.14, then taking the limit of $\mathcal{E}_T$ as $T \to +\infty$ is well defined. Moreover, in such a case we have $\mathcal{E}_\infty = \Omega_{x_0}^\delta$ and $\overline{\mathcal{E}}_\infty = \varnothing$.

It is straightforward to show that for each $T \in \mathbb{N}$, any path ω of length at least T, is such that[1]:

$$\omega \;\in\; (\mathcal{E}_T \cap \mathrm{TS}_1^{\mathcal{G}} \geq \nu_1) \;\cup\; (\mathcal{E}_T \cap \mathrm{TS}_1^{\mathcal{G}} < \nu_1) \;\cup\; \overline{\mathcal{E}}_T \tag{4.5}$$

where for instance, the measurable event $(\mathcal{E}_T \cap \mathrm{TS}_1^{\mathcal{G}} < \nu_1)$ stands for all the paths reaching the goal set $\mathcal{G}$ in at most T-steps and having a truncated sum $\mathrm{TS}_1^{\mathcal{G}} < \nu_1$ at the first visit of $\mathcal{G}$. Furthermore, the following also holds:

$$\begin{aligned} &\mathbb{P}_{x_0}[\mathcal{E}_T \cap \mathrm{TS}_1^{\mathcal{G}} \geq \nu_1] \\ &\qquad\qquad \leq \mathbb{P}_{x_0}[\mathcal{E}_\infty \cap \mathrm{TS}_1^{\mathcal{G}} \geq \nu_1] \\ &\qquad\qquad\qquad\qquad \leq \mathbb{P}_{x_0}[(\mathcal{E}_T \cap \mathrm{TS}_1^{\mathcal{G}} \geq \nu_1) \cup \overline{\mathcal{E}}_T] \end{aligned} \tag{4.6}$$

We are looking to maximize the probability of $(\mathcal{E}_\infty \cap \mathrm{TS}_1^{\mathcal{G}} \geq \nu_1)$, subject to the constraint on the expectation of $\mathrm{TS}_2^{\mathcal{G}}$ to characterize the values of ε for which $\mathbf{Pb}(\varepsilon)$ has or has not a solution. Such an event represents all the paths that are reaching $\mathcal{G}$ at some time-step ($T \to +\infty$) and having a truncated sum $\mathrm{TS}_1^{\mathcal{G}} \geq \nu_1$ at such a moment. Mathematically, we are looking for:

$$\begin{aligned} \sup_{\delta} \quad & \mathbb{P}_{x_0}^{\delta}[\,\mathcal{E}_\infty \cap \mathrm{TS}_1^{\mathcal{G}} \geq \nu_1\,] \\ \text{s.t.} \quad & \mathbb{E}_{x_0}^{\delta}[\,\mathrm{TS}_2^{\mathcal{G}}\,] \;<\; \nu_2 \end{aligned}$$

However, such a probability objective is bounded by the two probabilities in the inequality (4.6). Equivalently, maximizing the probability

1 Well understood that when we write $\mathrm{TS}_1^{\mathcal{G}} \geq \nu_1$, we refer to the event on the r.v. $\mathrm{TS}_1^{\mathcal{G}}$, i.e., to the event $\{\omega \in \Omega_{x_0} \,|\, \mathrm{TS}_1^{\mathcal{G}}(\omega) \geq \nu_1\}$.

of $(\mathcal{E}_\infty \cap \mathrm{TS}_1^{\mathcal{G}} \geq \nu_1)$ is the same as to minimizing the probability of its complement[2], i.e., $(\mathcal{E}_\infty \cap \mathrm{TS}_1^{\mathcal{G}} < \nu_1) \cup \overline{\mathcal{E}}_\infty$, see eq. (4.5) for $T \to +\infty$. This event represents all the paths that are either reaching $\mathcal{G}$ at some time-step $(T \to +\infty)$ and having a truncated sum $\mathrm{TS}_1^{\mathcal{G}} < \nu_1$ at such a moment (i.e., belonging to $\mathcal{E}_\infty \cap \mathrm{TS}_1^{\mathcal{G}} < \nu_1$), or never reaching $\mathcal{G}$ (i.e., belonging to $\overline{\mathcal{E}}_\infty$).

Applying the complement in the inequality (4.6), the following holds:

$$\begin{aligned} &\mathbb{P}_{x_0}[\mathcal{E}_T \cap \mathrm{TS}_1^{\mathcal{G}} < \nu_1] \\ &\qquad \leq \mathbb{P}_{x_0}[(\mathcal{E}_\infty \cap \mathrm{TS}_1^{\mathcal{G}} < \nu_1) \cup \overline{\mathcal{E}}_\infty] \\ &\qquad\qquad \leq \mathbb{P}_{x_0}[(\mathcal{E}_T \cap \mathrm{TS}_1^{\mathcal{G}} < \nu_1) \cup \overline{\mathcal{E}}_T] \end{aligned} \tag{4.7}$$

In this way, the two underlying optimization problems come naturally from the precedent inequality to bound the probability value that we are looking for, i.e., to solve the problem:

Randomized Strategy Problem

$$\begin{aligned} \inf_{\delta} \quad & \mathbb{P}^{\delta}_{x_0}[\,(\mathcal{E}_\infty \cap \mathrm{TS}_1^{\mathcal{G}} < \nu_1) \cup \overline{\mathcal{E}}_\infty\,] \\ \text{s.t.} \quad & \mathbb{E}^{\delta}_{x_0}[\,\mathrm{TS}_2^{\mathcal{G}}\,] \ < \ \nu_2 \end{aligned}$$

by minimizing the left and right side of (4.7) subject to the constraint in expectation, i.e.,

$$(i) \inf_{\delta} \ \mathbb{P}^{\delta}_{x_0}[\mathcal{E}_T \cap \mathrm{TS}_1^{\mathcal{G}} < \nu_1] \quad \text{s.t. } \mathbb{E}^{\delta}_{x_0}[\,\mathrm{TS}_2^{\mathcal{G}}\,] < \nu_2 \quad , \quad (ii) \inf_{\delta} \ \mathbb{P}^{\delta}_{x_0}[(\mathcal{E}_T \cap \mathrm{TS}_1^{\mathcal{G}} < \nu_1) \cup \overline{\mathcal{E}}_T] \quad \text{s.t. } \mathbb{E}^{\delta}_{x_0}[\,\mathrm{TS}_2^{\mathcal{G}}\,] < \nu_2$$

With a small abuse of notations, we represent both problems in a single one, parametrized by an $\alpha \in \{0, 1\}$ as follows. Let $\overline{\mathcal{E}}_{\alpha,T} \in \mathcal{B}(\Omega_{x_0})$ defined as:

$$\overline{\mathcal{E}}_{\alpha,T} \ := \ \begin{cases} \overline{\mathcal{E}}_T & \text{if } \alpha = 1 \\ \varnothing & \text{if } \alpha = 0 \end{cases} \tag{4.8}$$

For $T \in \mathbb{N}$ fixed, we call **objective function** for randomized strategies $\delta \in \Delta$ the next function[3]:

For pure strategies $\pi \in \Pi$, the definition is equivalent for $\mathbb{J}_{\alpha,T}(\pi)$.

$$\mathbb{J}_{\alpha,T}(\delta) \ := \ \mathbb{P}^{\delta}_{x_0}[\,(\mathcal{E}_T \cap \mathrm{TS}_1^{\mathcal{G}} < \nu_1) \cup \overline{\mathcal{E}}_{\alpha,T}\,]\,, \tag{4.9}$$

and so, the **optimization problems** are defined for each α by:

Jointly Randomized Strategy Problem

$$\begin{aligned} \inf_{\delta} \quad & \mathbb{J}_{\alpha,T}(\delta) \\ \text{s.t.} \quad & \mathbb{E}^{\delta}_{x_0}[\,\mathrm{TS}_2^{\mathcal{G}}\,] \ < \ \nu_2 \end{aligned} \tag{4.10}$$

Note that, by Assumption 4.2.3, this problem is feasible.

2 Here, it is possible that $\overline{\mathcal{E}}_\infty = \varnothing$, see Remark 4.3.1, but we continue to write $(\mathcal{E}_\infty \cap \mathrm{TS}_1^{\mathcal{G}} < \nu_1) \cup \overline{\mathcal{E}}_\infty$ with some abuse, because there is no sense that a path belongs to such an event if it is the case.

3 To be more precise, this function should be denoted as $\mathbb{J}_{x_0,\alpha,T,\nu_1}$ due to the fixed parameters. However, to relax the notation, we simply write $\mathbb{J}_{\alpha,T}$.

We denote the **optimal value** of the latter problem as:

$$\mathbb{J}^*_{\alpha,T} \;:=\; \inf_{\delta} \left\{ \mathbb{J}_{\alpha,T}(\delta) \;\mid\; \mathbb{E}^{\delta}_{x_0}[\,\mathrm{TS}^{\mathcal{G}}_2\,] \;<\; \nu_2 \right\}, \tag{4.11}$$

and if there exists the limit of $\mathbb{J}^*_{\alpha,T}$ as $T \to +\infty$, we denote it by:

$$\mathbb{J}^*_{\alpha,\infty} \;:=\; \lim_{T\to+\infty} \mathbb{J}^*_{\alpha,T} \,.$$

Note that it is possible that there is no strategy reaching the minimum on the objective function (4.9). Thus, for each $\gamma > 0$ we can consider by definition of infimum a strategy δ^*_γ that is γ-optimal for the problem (4.10), i.e.,

$$\mathbb{J}_{\alpha,T}(\delta^*_\gamma) \;\le\; \mathbb{J}^*_{\alpha,T} + \gamma \qquad \text{and} \qquad \mathbb{E}^{\delta^*_\gamma}_{x_0}[\,\mathrm{TS}^{\mathcal{G}}_2\,] \;<\; \nu_2 \,. \tag{4.12}$$

We mentioned that a technique based on polynomial optimization problems for solving the optimisation problem (4.10) for each $\alpha \in \{0,1\}$, is developed in [M3]. Such a proposed method is in general hard to solve, and the authors have not been able to exploit the particular shape of the optimization problems to get efficient specialized algorithms. A suitable method to find a solution of each optimization problem is developed in Chapter 5, which is based on duality theory [23, 24].

Coming back to the inequality (4.7), an iterative approach to compute the probability values from the left and the right in such an inequality is shown in the following.

Proposition 4.3.2

The sequence $(\mathbb{J}^*_{\alpha,T})_{T\in\mathbb{N}}$ is nonincreasing for $\alpha = 1$ and nondecreasing for $\alpha = 0$. Moreover, $\mathbb{J}^*_{0,T} \le \mathbb{J}^*_{1,T}$ for each $T \in \mathbb{N}$.

Proof. See Proof B.1.2 in Appendix B.1. ∎

First, we analyse theoretically when the problem $\mathbf{Pb}(\varepsilon)$ has or not a solution for each $\varepsilon \in [0,1]$. For that, we use the limit $\mathbb{J}^*_{\alpha,\infty}$ for each $\alpha = 0,1$. Thereafter, we construct an iterative approach using the nonincreasing sequence $(\mathbb{J}^*_{1,T})_{T\in\mathbb{N}}$ and the nondecreasing sequence $(\mathbb{J}^*_{0,T})_{T\in\mathbb{N}}$, to approximate the limit $\mathbb{J}^*_{1,T}$ and $\mathbb{J}^*_{0,T}$ as $T \to +\infty$, resp. Each iteration is made for each $T \in \mathbb{N}$ and thus the cartography of $\mathbf{Pb}(\varepsilon)$ is mapped on the interval $[0,1]$ at each iteration-step T.

Theorem 4.3.3

If $\varepsilon < \mathbb{J}^*_{1,\infty} \Rightarrow$ the problem $\mathbf{Pb}(\varepsilon)$ has no solution.

Proof. See Proof B.1.3 in Appendix B.1. ∎

Because the inequality $\mathbb{J}^*_{0,T} \le \mathbb{J}^*_{1,T}$ is true for each $T \in \mathbb{N}$ from the Proposition 4.3.2, the next corollary holds as a consequence of the precedent theorem.

Corollary 4.3.4

If $\varepsilon < \mathbb{J}^*_{0,\infty} \Rightarrow$ the problem $\mathbf{Pb}(\varepsilon)$ has no solution.

Based on the precedent results, the theoretical analysis is shown at the top of the Figure 4.2. On the other hand, to make an effective calculation of the limits of the sequences, we analyze the existence of the solution of $\mathbf{Pb}(\varepsilon)$ in function of each iteration-step T as follows.

Theorem 4.3.5

For each $T \in \mathbb{N}$, the problem $\mathbf{Pb}(\varepsilon)$ has a solution $\forall\, \varepsilon > \mathbb{J}^*_{1,T}$, and has no solution $\forall\, \varepsilon < \mathbb{J}^*_{0,T}$.

Proof. See Proof B.1.4 in Appendix B.1. ∎

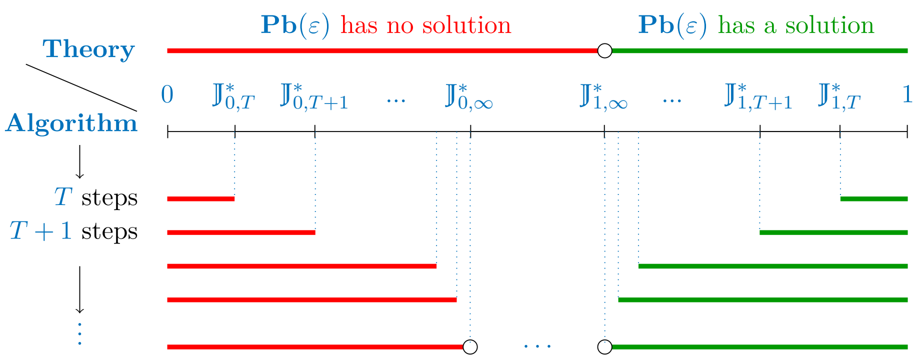


Figure 4.2: A (partial) cartography of our problem $\mathbf{Pb}(\varepsilon)$.

4.3.2 *SUMMARY*

Figure 4.2 summarizes the analysis of the previous section. Theoretically (Theorem 4.3.3), the picture seems rather complete since the only status of the problem $\mathbf{Pb}(\varepsilon)$ that remains uncertain is with $\varepsilon = \mathbb{J}^*_{1,\infty}$. However, it remains to discuss two things in a numerical sense (Theorem 4.3.5):

(i) the limits $\mathbb{J}^*_{0,\infty}$ and $\mathbb{J}^*_{1,\infty}$ are a priori unknown (by numerical calculation) and distinct. Hence the cartography is not effective and not complete so far.

(ii) the idea is then to use the sequence $(\mathbb{J}^*_{\alpha,T})_{T\in\mathbb{N}}$ for each α to approximate the limits. We will therefore discuss cases where the two limits coincide (we then say that the approach is *almost-complete*), allowing us for a converging scheme and hence an algorithm to almost cover the interval $[0,1]$ with either the red line (where there are no solutions) or the green line (where there is a solution), i.e., to almost compute the full cartography of $\mathbf{Pb}(\varepsilon)$.

4.4 ALMOST-COMPLETENESS OF THE APPROACH

In this section, we discuss the almost-completeness of our approach, and describe situations where one can show that

$$\mathbb{J}^*_{0,\infty} \quad = \quad \mathbb{J}^*_{1,\infty} \tag{4.13}$$

which allows us to reduce the unknown part of the cartography to a singleton, i.e., ε equal to (4.13). The situations for completeness we describe below are conditions over cycles in the doubly-weighted MDP $\mathcal{M} = (\mathcal{X}, \mathcal{A}, P, (C_i)_{i=1}^2)$, with fixed initial state $x_0 \in \mathcal{X}$.

Theorem 4.4.1

If all cycles have positive costs[a] under C_i for some $i \in \{1, 2\}$, then $\mathbb{J}^*_{0,\infty} = \mathbb{J}^*_{1,\infty}$.

[a] When we assume that cycles have positive costs, we mean it for every cycle, except for cycles at $\mathcal{G}$, which we assumed are self-loops with costs equal to zero.

Proof. See Proof B.1.5 in Appendix B.1. ∎

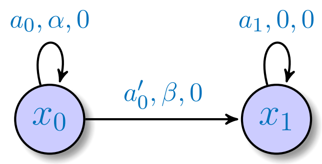


Figure 4.3: A doubly-weighted MDP parametrized by α and β, with $x_1 \in \mathcal{G}$. Over the edges, the first component represents the actions, and the other two are resp. the values for the costs C_1 and C_2.

Note that the result does not hold without the assumption on cycles. Indeed, considering the doubly-weighted MDP depicted in Figure 4.3 with $\alpha = 0$ and $\beta = -1$. If the first threshold is $\nu_1 = 0$, we have that for every $T \in \mathbb{N}$, $\mathbb{J}^*_{0,T} = 0$ and $\mathbb{J}^*_{1,T} = 1$, then the limits of the sequences are different. Also, one could think that assuming negativity of cycles under the cost function C_1 would be a similar result, but this is not the case as witnessed by the doubly-weighted MDP defined in the same Figure 4.3 with $\alpha = -1$ and $\beta = 1$. In such a case, for every $T \geq 2$, we have that $\mathbb{J}^*_{0,T} = 0$, while $\mathbb{J}^*_{1,T} = 1$.

4.5 PARTICULAR CASE

While the previous developments cannot give a solution to the problem $\mathbf{Pb}(\varepsilon)$ for $\varepsilon = 0$, since it requires not only to show that the limits of the sequences are $\mathbb{J}^*_{0,\infty} = \mathbb{J}^*_{1,\infty} = 0$, but also that there is a solution to the limit points (which we do not have in general). We dedicate special developments to that problem, which has been applied in [M1]. The particular problem $\mathbf{Pb}(0)$ can be rephrased as follows.

Definition 4.5.1: Particular Problem

Given a double-weighted MDP $\mathcal{M} = (\mathcal{X}, \mathcal{A}, P, (C_i)_{i=1}^2)$, an initial state $x_0 \in \mathcal{X}$, a goal set $\mathcal{G} \subset \mathcal{X}$ and two thresholds $\nu_1, \nu_2 \in \mathbb{Q}$. The particular problem $\mathbf{Pb}(0)$ is defined as[a]: there exists $\delta_0 \in \Delta$, such that each $\omega \in \Omega_{x_0}^{\delta_0}$ reaches $\mathcal{G}$,

$$\mathbb{P}_{x_0}^{\delta_0}[\,\mathrm{TS}_1^{\mathcal{G}} \geq \nu_1\,] \geq 1 \qquad \text{and} \qquad \mathbb{E}_{x_0}^{\delta_0}[\,\mathrm{TS}_2^{\mathcal{G}}\,] \;<\; \nu_2$$

a To be more precise, the problem should be denoted as $\mathbf{Pb}_{\mathcal{M},\nu_1,\nu_2}(0)$, because this one depends on the fixed $\mathcal{M}$, ν_1 and ν_2. However, to relax the notation, we simply write $\mathbf{Pb}(0)$.

Note that this problem is somehow a "beyond worst-case problem", as defined in [102], with a strong constraint on all outcomes, and a stochastic constraint (here defined using expected value). We describe a solution in the case all cycles of $\mathcal{M}$ have nonnegative cost under C_1. As we explain below, our solution extends to multiple costs with non-negative cycles and strong constraints, like the one for $\mathrm{TS}_1^{\mathcal{G}}$. However, it is not correct when C_1 may have negative cycles as well. In that case, the status of the problem remains open [M3].

Assumption 4.5.2: Nonnegative Cycles

All cycles in $\mathcal{M}$ have nonnegative costs under C_1.

UNFOLDING THE MDP

To solve the problem $\mathbf{Pb}(0)$, we fix a double-weighted MDP

$$\mathcal{M} \;=\; \Big(\mathcal{X},\, \mathcal{A},\, P,\, (C_i)_{i=1}^2\Big)\,,$$

an initial state $x_0 \in \mathcal{X}$, a goal set $\mathcal{G} \subset \mathcal{X}$ and two thresholds $\nu_1, \nu_2 \in \mathbb{Q}$. For a fixed $T \in \mathbb{N}_0$, we define the unfolding $\mathcal{M}_T$ of $\mathcal{M}$, see Definition 2.6.10, explicitly keeping track of the cost C_1 in the states of $\mathcal{M}_T$, but here, keeping the second cost function C_2 on the transitions. More precisely, the unfolding of $\mathcal{M}$ used to solve $\mathbf{Pb}(0)$ is the following structure:

$$\mathcal{M}_T \;=\; \Big(\mathcal{S}_T,\, s_0,\, \mathcal{A},\, P_T,\, C\Big)\,,$$

where the space of states is:

$$\mathcal{S}_T \;=\; \mathcal{X} \times \Big([T\,C_1^{\min},\, T\,C_1^{\max}] \cap \mathbb{Q}\Big) \times \{0,\, 1,\, \ldots,\, T\}\,,$$

with $C_1^{\min}, C_1^{\max} \in \mathbb{Q}$ to be resp. the (possibly negative) minimum and maximum cost appearing in the transitions of $\mathcal{M}$ due to C_1. The initial state is $s_0 = (x_0, 0, 0)$, the set of actions is $\mathcal{A}$ from $\mathcal{M}$, the probability transition function between states is $P_T : \mathcal{S}_T \times \mathcal{A} \to \mathcal{D}(\mathcal{S}_T)$, defined as:

$$P_T(s_t, a_t)(s_{t+1}) \;=\; \Big(P \circ (\mathrm{proj}_1, \mathrm{id}_{\mathcal{A}})(s_t, a_t)\Big)(s_{t+1})$$

if $\text{proj}_2(s_{t+1}) = \min\left\{\nu_1, \ \text{proj}_2(s_t) + C_1(\text{proj}_1(s_t), a_t, \text{proj}_1(s_{t+1}))\right\}$ and $\text{proj}_3(s_{t+1}) \leq T$; and $P_T(s_t, a_t)(s_{t+1}) = 0$ otherwise. Finally, the cost function over transitions is $C : \mathcal{S}_T \times \mathcal{A} \times \mathcal{S}_T \to \mathbb{Q}$, such that:

$$C = C_2 \circ (\text{proj}_1, \text{id}_{\mathcal{A}}, \text{proj}_1) \ .$$

In particular, if the current state is $s_t = (x_t, y_t, t)$ and the action selected is a_t, then a transition occurs to $s_{t+1} = (x_{t+1}, y_{t+1}, t+1)$ with probability $P_T(s_t, a_t)(s_{t+1}) = P(x_t, a_t)(x_{t+1})$ if $t + 1 \leq T$ and $y_{t+1} = y_t + C_1(x_t, a_t, x_{t+1})$; and the cost $C(s_t, a_t, s_{t+1}) = C_2(x_t, a_t, x_{t+1})$ is incurred.

We also consider each state s_T of the form $\text{proj}_3(s_T) = T$, as an initial state of the MDP $\mathcal{M}$. More precisely, we keep a copy of $\mathcal{M}$ below each leaf, i.e., $\text{proj}_1(s_T)$ is an initial state, as is shown in the Figure 4.4. Since we are interested in reaching the goal set $\mathcal{G}$ defined from $\mathcal{M}$, such that the truncated sum $\text{TS}_1^{\mathcal{G}}$ be at least ν_1, we consider the **augmented goal set** in the unfolding $\mathcal{M}_T$, defined as:

$$\mathcal{G}_{\nu_1} \ := \ \mathcal{G} \times \{\, \nu_1 \,\} \times \{0, 1, ..., T\} \ \subset \ \mathcal{S}_T \ . \tag{4.14}$$

In addition, since $\mathcal{G}$ is absorbent, see Assumption 2.6.8, then $\mathcal{G}_{\nu_1}$ is absorbent as well. Note that there is a natural one-to-one correspondence between paths in $\mathcal{M}$ and its unfolding, and therefore, strategies in $\mathcal{M}$ can equivalently be seen as strategies in $\mathcal{M}_T$. Also, since the set of states $\mathcal{X}$ in $\mathcal{M}$ is finite and C_1 takes rational numbers, there exists a maximum depth of interest for the unfolding of $\mathcal{M}$. We denote such a maximum depth as $T^{\max} \in \mathbb{N}$. Concerning the constraint in expectation, we denote by $\text{TS}^{\mathcal{G}_{\nu_1}}$ the truncated sum due to the cost function C in the unfolding.

Theorem 4.5.3

There exists a solution to $\mathbf{Pb}(0)$ if, and only if, there is a strategy δ^* in the unfolding $\mathcal{M}_{T^{\max}}$, such that each $\omega \in \Omega_{s_0}^{\delta^*}$ reaches $\mathcal{G}_{\nu_1}$ and $\mathbb{E}_{s_0}^{\delta^*}[\,\text{TS}^{\mathcal{G}_{\nu_1}}\,] \ < \ \nu_2$.

Proof. See Proof B.1.6 in Appendix B.1. ∎

We can further show that $\mathbf{Pb}(0)$ has a solution if, and only if, the stochastic shortest path problem in $\mathcal{M}_{T^{\max}}$ with the cost function C, for reaching $\mathcal{G}_{\nu_1}$ from s_0 is less than ν_2, where a solution can be constructed in POLYNOMIAL-TIME [14]. However, the size of the unfolding $\mathcal{M}_{T^{\max}}$ is exponential, more precisely it is PSEUDO-POLYNOMIAL [M3] in the size of the initial double-weighted MDP $\mathcal{M}$ because of the thresholds.

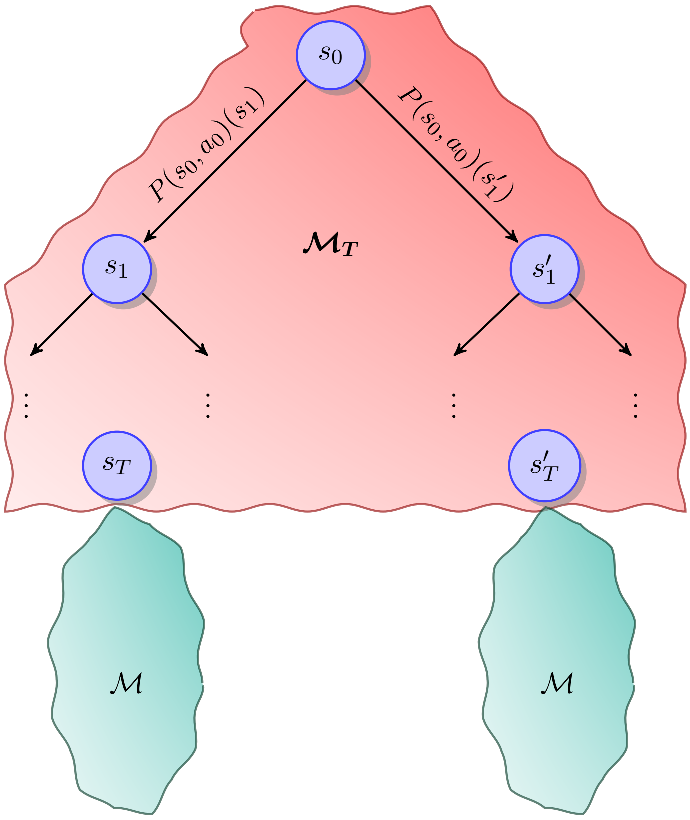


Figure 4.4: Unfolding $\mathcal{M}_T$ of the MDP $\mathcal{M}$, where we keep a copy of $\mathcal{M}$ below each leaf s_T.

4.6 DISCUSSION

In this chapter, we have investigated a multi-constrained reachability problem over Markov decision process (MDP), which originated in the context of controllable electric devices scheduling problem (e.g., the electric vehicle charging problem [20][M1]). In Section 4.2, the problem formulation 4.2.1 consists in finding a strategy that surely reaches a quantitative goal (e.g., all vehicles are fully charged and the load of the network remains below a given bound at any time), while satisfying a condition on the expected value of some random variable (e.g., minimizing the accelerated ageing factor of the distribution network transformer, or the expected monetary cost of charging all electric vehicles). Here, we have developed a partial solution to the general problem in Section 4.3, by providing a cartography, see Definition 4.2.4, of the solution to a relaxed version (4.10). In Section 4.4, we identified realistic conditions on the structure of the double-weighted MDP under which the cartography is almost-complete. However, the case of MDPs not satisfying these conditions remains also open, but we believe that our approximation techniques may give interesting informations which suffice for practical applications such as the electric vehicle charging problem.

Our approach for the particular case shown in Section 4.5, which amounts to explicitly keep track of the worst-case constraint on the first cost function, immediately extends to multi-weights with worst-case constraints (with the same assumptions on cycles for the multidimensional probability constraints problem for truncated sums). Note that the more general setting could not be solved, which has to be put in parallel with the undecidability result of [102]. The cartography for the relaxed problem (4.10) requires solving sequences of intermediary optimization problems, for which a method will be developed in the next chapter. This work could be extended to several such costs as well, e.g., by putting an assumption on the cycles under the costs functions. A nice continuation of our work would consist in computing (approximations of) Pareto-optimal solutions in such a setting. Improving the complexity and practicality of our approach is also on our agenda for future work.

5

SOLVING THE OPTIMIZATION PROBLEMS

Abstract:

In the previous Chapter 4, we have investigated the "cartography" (set of values of a parameter) for which the general problem (modeled in a multi-weighted Markov decision process) has or has not an optimal solution. While a partial cartography has been obtained via two sequences of optimization problems for randomized strategies, in this chapter we focus on solving them. The key idea is that for each fixed index of the sequences, we unify the optimization problems into one and define its mixed strategy counterpart, which is a more easy problem to solve (the randomization lies on pure strategies and not on local actions at every time-step). Finding an optimal mixed strategy we can define here its randomized counterpart, solving the unified problem and hence, each optimization problem. An optimal mixed strategy is built by finding two pure strategies by Lagrangian-based approach, namely: one strategy satisfying the constraint and another that does not. The synthesis of an optimal solution for the general problem can be thus guaranteed.

LIST OF ABBREVIATIONS AND SYMBOLS

MDP	Markov Decision Process
SSP	Stochastic Shortest Path
P-PS	Pure Strategy Primal Problem
P-MS	Mixed Strategy Primal Problem
D-PS	Pure Strategy Dual Problem
D-MS	Mixed Strategy Dual Problem

a_t	action at t
$\mathcal{A}$	finite action space in $\mathcal{M}$
$\mathcal{B}(\Omega_{x_0})$	Borel sigma-algebra over Ω_{x_0}
$\mathrm{Conv}(S)$	convex hull of a discrete set S
C	cost function in $\mathcal{M}_T$
C_i	i^{th} cost function
$C_i^{\min}$	minimum C_i cost in the transitions of $\mathcal{M}$
$C_i^{\max}$	maximum C_i cost in the transitions of $\mathcal{M}$
δ	randomized strategy
Δ	set of randomized strategies
$\Delta[\underline{\Pi}_\gamma]$	set of mixed strategies over $\underline{\Pi}_\gamma$
$\mathcal{D}(\mathcal{S})$	set of probability distributions over a finite set $\mathcal{S}$
$\mathbb{E}_{x_0}$	expectation operator under a fixed x_0
$\mathcal{E}_T$	set of paths reaching $\mathcal{G}$ in at most T-steps
$\boldsymbol{\mathcal{E}}_T$	set of paths reaching $\mathcal{G}_T$ in at most T-steps
$\overline{\mathcal{E}}_T$	complement set $\mathcal{E}_T$
$\overline{\boldsymbol{\mathcal{E}}}_T$	complement set $\boldsymbol{\mathcal{E}}_T$
$\overline{\mathcal{E}}_{\alpha,T}$	to be $\overline{\mathcal{E}}_T$ for $\alpha = 1$ and $\varnothing$ otherwise
$\overline{\boldsymbol{\mathcal{E}}}_{\alpha,T}$	to be $\overline{\boldsymbol{\mathcal{E}}}_T$ for $\alpha = 1$ and $\varnothing$ otherwise
$\mathcal{G}$	set of goal states in an MDP model
$\mathcal{G}_T$	set of goal states in an unfolded-MDP model
$\mathrm{id}_{\mathcal{A}}$	identity function on a set $\mathcal{A}$
$\mathbb{J}_{\alpha,T}$	objective function of a minimization problem
$\mathbb{J}^{\mathrm{p}*}_{\alpha,T}$	optimal value of $\mathbb{J}_{\alpha,T}$ under strategies $\overline{\pi}$
$\mathbb{J}^{\mathrm{m}*}_{\alpha,T}$	optimal value of $\mathbb{J}_{\alpha,T}$ under strategies σ
K	number of pure strategies in $\underline{\Pi}_\gamma$

λ	dual variable
λ^*	optimal dual variable
$\mathbb{L}_{\alpha,T}$	Lagrange function
$\mathbb{L}^{\mathrm{p}}_{\alpha,T}$	Lagrange dual function for pure strategies
$\mathbb{L}^{\mathrm{p}*}_{\alpha,T}$	optimal dual value of $\mathbb{L}^{\mathrm{p}}_{\alpha,T}$
$\mathbb{L}^{\mathrm{m}}_{\alpha,T}$	Lagrange dual function for mixed strategies
$\mathbb{L}^{\mathrm{m}*}_{\alpha,T}$	optimal dual value of $\mathbb{L}^{\mathrm{m}}_{\alpha,T}$
$\mathcal{M}$	MDP model
$\mathcal{M}_T$	unfolding of $\mathcal{M}$ with depth T
Ω_{x_0}	set of histories or paths of a system from an initial state x_0
proj_j	projection function on the j-component of a sequence
$\mathrm{proj}^{\mathcal{X}}_t$	projection function on $\mathcal{X}$ in the t-component of a path
π	pure strategy
π_γ	γ-optimal pure strategy for $\mathbb{SP}^*_{s_t}$
$\overline{\pi}$	pure strategy playing after T-steps as π_γ
$\overline{\pi}^*_\alpha$	optimal pure strategy for $\mathbb{J}^{\mathrm{p}*}_{\alpha,T}$
$\overline{\pi}^*_{\alpha,\lambda}$	optimal pure strategy for $\mathbb{L}^{\mathrm{p}}_{\alpha,T}$ for fixed λ
$\underline{\Pi}$	set of pure Markov strategies
$\underline{\Pi}_\gamma$	subset of strategies $\overline{\pi}$
P	transition probability between states in $\mathcal{M}$
P_T	transition probability between states in $\mathcal{M}_T$
$\mathbb{P}_{x_0}$	probability measure on $\mathcal{B}(\Omega_{x_0})$
σ	mixed strategy
$\sigma^*_{\alpha,\kappa}$	κ-optimal mixed strategy for $\mathbb{J}^{\mathrm{m}*}_{\alpha,T}$
$\sigma^*_{\alpha,\vartheta,\lambda}$	ϑ-optimal mixed strategy for $\mathbb{L}^{\mathrm{m}}_{\alpha,T}$ for fixed λ
$\sigma^*_{\alpha,\lambda}$	optimal mixed strategy for $\mathbb{L}^{\mathrm{m}}_{\alpha,T}$ for fixed λ
σ^*_α	optimal mixed strategy for $\mathbb{L}^{\mathrm{m}}_{\alpha,T}$ under λ^*
$\sigma^\star_{\alpha,\varsigma}$	strategy σ^*_α perturbed by a constant ς
$\mathbb{SP}^*_{s_t}$	expected SSP-value from a state s_t
$\mathbb{SP}^{\pi_\gamma}_{s_t}$	expected SSP-value from a state s_t under π_γ
$\mathcal{S}_T$	augmented space of states in $\mathcal{M}_T$
s_t	state in $\mathcal{S}_T$ at t
$\overline{s}_T$	vector of leafs in $\mathcal{M}_T$
T	finite horizon, or depth
$\mathrm{TS}^{\mathcal{G}}_i$	truncated sum function for $\mathcal{G}$ under C_i
x_t	system state at t
$\mathcal{X}$	finite set of states in $\mathcal{M}$

5.1 MOTIVATION AND CONTRIBUTIONS

The problem of interest in this chapter, comes from the two sequences of optimization problems defined in the previous Chapter 4. We focus on the existence and synthesis of a mixed strategy of the problem, in which its objective function is a probability and it has a constraint in expectation bounded above strictly by a fixed threshold. Obtaining an optimal mixed strategy allows us to define a randomized strategy solution for the general multi-constrained problem of Chapter 4. This chapter explains our research reported in [M2] and in a forthcoming paper [M6].

The optimization problem, arising from the general problem of Section 4.2, takes as input a single-weighted unfolded-MDP, and requires the existence and synthesis of a strategy such that:

(*i*) it minimizes a probability of a measurable set of paths (defined with some condition over a finite number of steps),

(*ii*) it satisfies a constraint (strict inequality) between the expectation of the accumulated sum of costs with a reachability condition and a fixed threshold.

Such a problem is defined from the beginning by pure strategies, which we call Pure Strategy Primal (P-PS) problem, and after we define its mixed strategy counterpart, which we call Mixed Strategy Primal (P-MS) problem. The latter is a convexification of the initial P-PS problem. Finding an optimal mixed strategy can provide optimal values that are at least as good as the values obtained with pure strategies. Interestingly, it is shown in this section that an optimal mixed strategy solving the P-MS problem, only needs to mix up to two pure strategies in order to achieve optimality. These two pure strategies can be built iteratively by an algorithmic technique provided in this chapter, which solve an unconstrained problem. Our approach is based on Lagrangian-based method to convert the problem with the constraint in expectation into a problem without constraint by adding the constraint in the objective function, pondered by a parameter so-called dual variable. Since the objective function (a probability) and the constraint (an expectation) are different structures, we transform the objective function into an expectation of an accumulated sum of indicator random variables to solve (e.g., by dynamic programming) the unconstrained problem. To find the two pure strategies defining an optimal mixed strategy for the problem P-MS, we need to find an optimal dual variable, which is computed here by bisection method. After an optimal mixed strategy is built, its randomized counterpart can be defined from an interesting result in game theory, namely: the Kuhn's theorem.

5.1.1 *STRUCTURE*

The main contributions and structure of this section can be summarized as follows.

(i) In Section 5.2, we revisit the main definitions of the Chapter 4 to place the work of this chapter in the dynamic of research on the general multi-constrained problem. We focus mainly on the synthesis of an optimal randomized strategy for each constrained optimization problem defined in the previous chapter, in which the objective function is a probability and the constraint is an expectation strictly upper bounded by a threshold. Here, the two optimization problems are unified into one.

(ii) In Section 5.3, the key idea is introduced. Since randomized and mixed strategies are equivalent here (see Section 2.6.3), We focus on the unified optimization problem for mixed strategies. Because the latter are merely a convexification of pure strategies, we introduce thus the pure and the mixed strategy (primal) problems.

(iii) in Section 5.4, we present the solution methodology which is based on dual optimization [23, 24]. We show that there exit two pure strategies (built by duality), which define an optimal mixed strategy for the primal (mixed) problem. This strategy gives the equality in the constraint in expectation (with the upper bound threshold). Next, we perturb such a mixed strategy to have another one that satisfies the strict inequality and to be near to the optimum. Finally, we define a randomized strategy solution of the initial problem through the (perturbed) mixed strategy.

5.2 BACKGROUND

First, we fix a doubly-weighted MDP $\mathcal{M} = (\mathcal{X}, \mathcal{A}, P, (C_i)_{i=1}^2)$, an initial state $x_0 \in \mathcal{X}$, a goal set $\mathcal{G} \subset \mathcal{X}$ (which is absorbent by Assumption 2.6.8), and two thresholds $\nu_1, \nu_2 \in \mathbb{Q}$. From $\mathcal{M}$, we write $(\Omega_{x_0}, \mathcal{B}(\Omega_{x_0}), \mathbb{P}_{x_0})$ the induced probability space, see Section 2.6.2 for details.

Let $T \in \mathbb{N}_0$. From the Section 4.3.1, we consider $\mathcal{E}_T \in \mathcal{B}(\Omega_{x_0})$ as the set of paths reaching the goal set $\mathcal{G}$ in at most T-steps, which is defined as:

$$\mathcal{E}_T \quad = \quad \left\{ \omega \in \Omega_{x_0} \quad | \quad \exists\, t \leq T \quad : \quad \mathrm{proj}_t^{\mathcal{X}}(\omega) \in \mathcal{G} \right\}, \tag{4.1}$$

and has the property $\mathcal{E}_T \subseteq \mathcal{E}_{T+1}$ for each $T \in \mathbb{N}_0$. If the limit of $\mathcal{E}_T$ exists as $T \to +\infty$, we write it $\mathcal{E}_\infty$. Since $\mathcal{B}(\Omega_{x_0})$ is a well-defined sigma-algebra, the complement of $\mathcal{E}_T$ is measurable and represents the set of paths that are not reaching $\mathcal{G}$ in the first T-steps, i.e.,

$$\overline{\mathcal{E}}_T \quad = \quad \left\{ \omega \in \Omega_{x_0} \quad | \quad \mathrm{proj}_t^{\mathcal{X}}(\omega) \notin \mathcal{G} \;, \quad \forall\, t \leq T \right\}, \tag{4.3}$$

which has the property $\mathcal{E}_T \supseteq \mathcal{E}_{T+1}$ for each $T \in \mathbb{N}_0$. Again, if the limit of $\overline{\mathcal{E}}_T$ exists as $T \to +\infty$, we write it $\overline{\mathcal{E}}_\infty$. In addition, we have seen that any path ω of length at least T satisfies the following[1]:

$$\omega \;\in\; (\mathcal{E}_T \cap \mathrm{TS}_1^{\mathcal{G}} \geq \nu_1) \;\cup\; (\mathcal{E}_T \cap \mathrm{TS}_1^{\mathcal{G}} < \nu_1) \;\cup\; \overline{\mathcal{E}}_T\,, \tag{4.5}$$

where for instance, $(\mathcal{E}_T \cap \mathrm{TS}_1^{\mathcal{G}} < \nu_1)$ stands for all the paths reaching the goal set $\mathcal{G}$ in at most T-steps and having a truncated sum $\mathrm{TS}_1^{\mathcal{G}} < \nu_1$ at the first visit of $\mathcal{G}$. For a precise definition of the $\mathrm{TS}_1^{\mathcal{G}}$, see Definition 2.6.7. On the other hand, using the probability measure $\mathbb{P}_{x_0}$, the following holds:

$$\begin{aligned} &\mathbb{P}_{x_0}[\mathcal{E}_T \cap \mathrm{TS}_1^{\mathcal{G}} < \nu_1] \\ &\qquad \leq \mathbb{P}_{x_0}[(\mathcal{E}_\infty \cap \mathrm{TS}_1^{\mathcal{G}} < \nu_1) \cup \overline{\mathcal{E}}_\infty] \\ &\qquad\qquad \leq \mathbb{P}_{x_0}[(\mathcal{E}_T \cap \mathrm{TS}_1^{\mathcal{G}} < \nu_1) \cup \overline{\mathcal{E}}_T] \end{aligned} \tag{4.7}$$

In this way, the two underlying optimization problems come naturally from the precedent inequality to bound the probability value that we are looking for, i.e., to solve the problem under randomized strategies $\delta \in \Delta$:

Randomized Strategy Problem

$$\begin{aligned} \inf_{\delta} \quad & \mathbb{P}_{x_0}^{\delta}[\,(\mathcal{E}_\infty \cap \mathrm{TS}_1^{\mathcal{G}} < \nu_1) \cup \overline{\mathcal{E}}_\infty\,] \\ \text{s.t.} \quad & \mathbb{E}_{x_0}^{\delta}[\,\mathrm{TS}_2^{\mathcal{G}}\,] \quad < \quad \nu_2 \end{aligned}$$

1 Well understood that when we write $\mathrm{TS}_1^{\mathcal{G}} \geq \nu_1$, we refer to the event on the r.v. $\mathrm{TS}_1^{\mathcal{G}}$, i.e., to the event $\{\omega \in \Omega_{x_0} \,|\, \mathrm{TS}_1^{\mathcal{G}}(\omega) \geq \nu_1\}$.

We minimize the left and right side of (4.7) subject to the constraint in expectation, i.e.,

$$(i)\ \inf_{\delta}\ \mathbb{P}^{\delta}_{x_0}[\mathcal{E}_T \cap \mathrm{TS}^{\mathcal{G}}_1 < \nu_1] \quad , \quad (ii)\ \inf_{\delta}\ \mathbb{P}^{\delta}_{x_0}[(\mathcal{E}_T \cap \mathrm{TS}^{\mathcal{G}}_1 < \nu_1) \cup \overline{\mathcal{E}}_T]$$
$$\text{s.t. } \mathbb{E}^{\delta}_{x_0}[\,\mathrm{TS}^{\mathcal{G}}_2\,] < \nu_2 \qquad\qquad \text{s.t. } \mathbb{E}^{\delta}_{x_0}[\,\mathrm{TS}^{\mathcal{G}}_2\,] < \nu_2$$

where $\mathbb{P}^{\delta}_{x_0}$ stands for the probability measure from the induced probability space $(\Omega^{\delta}_{x_0}, \mathcal{B}(\Omega^{\delta}_{x_0}), \mathbb{P}^{\delta}_{x_0})$ that is induced from the MDP $\mathcal{M}$ to the MC $\mathcal{M}^{\delta}$ when δ is fixed, and $\mathbb{E}^{\delta}_{x_0}$ is the respective expectation operator, see Section 2.6.4 for details.

With a small abuse of notations, we represent both problems in a single one, which is parametrized by $\alpha \in \{0, 1\}$ as follows. We define an event $\overline{\mathcal{E}}_{\alpha,T} \in \mathcal{B}(\Omega_{x_0})$ by:

$$\overline{\mathcal{E}}_{\alpha,T} \;=\; \begin{cases} \overline{\mathcal{E}}_T & \text{if } \alpha = 1 \\ \varnothing & \text{if } \alpha = 0 \end{cases} \tag{4.8}$$

We are interested in solving the following problem for α and T fixed:

Note that, by Assumption 4.2.3, this problem is feasible.

Jointly Randomized Strategy Problem

$$\inf_{\delta}\quad \mathbb{P}^{\delta}_{x_0}[\,(\mathcal{E}_T \cap \mathrm{TS}^{\mathcal{G}}_1 < \nu_1) \cup \overline{\mathcal{E}}_{\alpha,T}\,]$$
$$\text{s.t.}\quad \mathbb{E}^{\delta}_{x_0}[\,\mathrm{TS}^{\mathcal{G}}_2\,] \;<\; \nu_2 \tag{5.1}$$

To solve this problem, we construct the unfolding of $\mathcal{M}$ with depth T, and keeping a copy of $\mathcal{M}$ in each leaf. The idea is to optimize from the initial state up to the leafs. Formally, it is explained in the following.

UNFOLDING THE MDP

For $T \in \mathbb{N}$ fixed, we construct the unfolding $\mathcal{M}_T$ of the MDP

$$\mathcal{M} \;=\; \Big(\mathcal{X},\, \mathcal{A},\, P,\, (C_i)_{i=1}^{2}\Big)\,,$$

see Definition 2.6.10, by explicitly keeping track of the cost C_1 in the states of $\mathcal{M}_T$, but here, keeping the cost function C_2 on the transitions. On the contrary to the particular case seen in Section 4.5, we do not suppose here that all cycles in $\mathcal{M}$ under the cost function C_1 have nonnegative costs, see Assumption 4.5.2. Here, we stop at depth T and we keep a copy of $\mathcal{M}$ below each leaf of $\mathcal{M}_T$. Mathematically, let $x_0 \in \mathcal{X}$ a fixed initial state of $\mathcal{M}$. The unfolding of $\mathcal{M}$ used here is the following structure:

$$\mathcal{M}_T \;=\; \Big(\mathcal{S}_T,\, s_0,\, \mathcal{A},\, P_T,\, C\Big)\,, \tag{5.2}$$

where the space of states is:

$$\mathcal{S}_T \;=\; \mathcal{X} \times \Big([T\,C_1^{\min},\, T\,C_1^{\max}] \cap \mathbb{Q}\Big) \times \{0,\, 1,\, \ldots,\, T\}\,,$$

with $C_1^{\min}, C_1^{\max} \in \mathbb{Q}$ to be resp. the minimum and maximum cost appearing in the transitions of $\mathcal{M}$ due to C_1. In the unfolding $\mathcal{M}_T$, the initial state is $s_0 = (x_0, 0, 0)$, the set of actions is $\mathcal{A}$ from $\mathcal{M}$, the probability transition function between states is $P_T : \mathcal{S}_T \times \mathcal{A} \to \mathcal{D}(\mathcal{S}_T)$, defined by:

$$P_T(s_t, a_t)(s_{t+1}) \quad = \quad \Big(P \circ (\mathrm{proj}_1, \mathrm{id}_{\mathcal{A}})(s_t, a_t)\Big)(s_{t+1})$$

if $\mathrm{proj}_2(s_{t+1}) = \mathrm{proj}_2(s_t) + C_1(\mathrm{proj}_1(s_t), a_t, \mathrm{proj}_1(s_{t+1}))$ and $\mathrm{proj}_3(s_{t+1}) \leq T$; and $P_T(s_t, a_t)(s_{t+1}) = 0$ otherwise. Finally, the cost function over transitions is $C : \mathcal{S}_T \times \mathcal{A} \times \mathcal{S}_T \to \mathbb{Q}$, defined by:

$$C \quad = \quad C_2 \circ (\mathrm{proj}_1, \mathrm{id}_{\mathcal{A}}, \mathrm{proj}_1) \; . \tag{5.3}$$

It is easy to see that there is a natural one-to-one correspondence between paths in $\mathcal{M}$ and its unfolding, and therefore, the strategies in $\mathcal{M}$ can equivalently be seen as strategies in $\mathcal{M}_T$.

By Assumption 4.2.3, we assume that there exists a strategy allowing us to reach the goal set $\mathcal{G}$ in $\mathcal{M}$ from each state, and that it satisfies the expected shortest path objective for the truncated sum $\mathrm{TS}_2^{\mathcal{G}}$ and the threshold ν_2 from the initial state x_0 in $\mathcal{M}$, see Definition 2.6.23 for details. As we have seen in the Section 2.6.5, the existence of such a strategy is reduced to the so-called SSP problem, where a pure memoryless strategy $\pi \in \underline{\Pi}$ can be computed in POLYNOMIAL-TIME [14, 25]. Thus, we will consider the following.

From each leaf s_T in $\mathcal{M}_T$, more precisely, from each state $\mathrm{proj}_1(s_T)$ in $\mathcal{M}$, we minimize the total expected cost with C_2 until $\mathcal{G}$ is reached. Because we are interested in such an accumulated cost and C_1 is kept in the states in the unfolding $\mathcal{M}_T$, we write also C to refer to the cost C_2 to be consistent with the single cost in $\mathcal{M}_T$, see eq. (5.3). Also for readability, we write ν for the threshold ν_2, and $\mathrm{TS}^{\mathcal{G}}$ for the truncated sum $\mathrm{TS}_2^{\mathcal{G}}$. The expected SSP-value from each leaf s_T is denoted by:

$$\mathbb{SP}^*_{s_T} \quad := \quad \inf_{\pi} \; \Big\{ \; \mathbb{E}^{\pi}_{x_T}[\mathrm{TS}^{\mathcal{G}}] \quad | \quad x_T \; = \; \mathrm{proj}_1(s_T) \; \Big\} \; . \tag{5.4}$$

When the context is clear, we do not make any difference between x_T and s_T. In the following, we say that for $\gamma > 0$, a strategy $\pi_\gamma \in \underline{\Pi}$ is γ-optimal for the SSP problem from s_T if:

$$\mathbb{SP}^{\pi_\gamma}_{s_T} \quad := \quad \mathbb{E}^{\pi_\gamma}_{s_T}[\mathrm{TS}^{\mathcal{G}}] \quad \leq \quad \mathbb{SP}^*_{s_T} \quad + \quad \gamma \; , \tag{5.5}$$

which is assumed to be such that:

$$\mathbb{SP}^{\pi_\gamma}_{s_0} \quad < \quad \nu \; . \tag{5.6}$$

We consider that each s_T is labeled with the corresponding SSP-value. In this way, we know how much we increase (in expected cost under C) to reach $\mathcal{G}$ from each leaf of the unfolding. We write $m \in \mathbb{N}$ for the number of leafs in the unfolding and denote the vector of these as:

$$\overline{s}_T \quad := \quad (s_{T,1}, \; \ldots \; , \; s_{T,m}) \; . \tag{5.7}$$

We are able now to define the optimization problem concerning the two events $\mathcal{E}_T$ and $\overline{\mathcal{E}}_{\alpha,T}$, see eqs. (4.1) and (4.8) resp., but in this case on the unfolding keeping of $\mathcal{M}$ in the leafs. First, so that the problem (5.1) and the new structure $\mathcal{M}_T$ are consistent, we let:

$$\boldsymbol{\mathcal{E}}_T \;:=\; \Big\{ \omega \in \Omega_{s_0} \;\mid\; \exists\, t \leq T \;:\; \mathrm{proj}_t^{\mathcal{S}_T}(\omega) \in \mathcal{G}_T \Big\}\,, \tag{5.8}$$

as the set of paths reaching a set $\mathcal{G}_T \subset \mathcal{S}_T$ in at most T-steps, where $\mathcal{G}_T$ is defined as:

$$\mathcal{G}_T \;:=\; \mathcal{G} \times \{\, y \in \mathbb{Q} \mid y < \nu_1 \,\} \times \{0, 1, ..., T\}\,. \tag{5.9}$$

That is, the set of states s_t such that its first component is in $\mathcal{G}$, i.e., $\mathrm{proj}_1(s_t) \in \mathcal{G}$, and the second one being less than ν_1 in such a moment $t \leq T$. Since $\mathcal{G}$ is absorbent, then $\mathcal{G}_T$ is absorbent as well, see Assumption 2.6.8. On the other hand, the event $\overline{\mathcal{E}}_T$ in the eq. (4.3), stays the same but here, in the context of the unfolding, we write it as:

$$\overline{\boldsymbol{\mathcal{E}}}_T \;:=\; \Big\{ \omega \in \Omega_{s_0} \;\mid\; (\mathrm{proj}_1 \circ \mathrm{proj}_t^{\mathcal{S}_T})(\omega) \notin \mathcal{G} \;,\;\; \forall\, t \leq T \Big\}\,,$$

so that $\overline{\mathcal{E}}_{\alpha,T}$ of the eq. (4.8) is considered now as:

$$\overline{\boldsymbol{\mathcal{E}}}_{\alpha,T} \;:=\; \begin{cases} \overline{\boldsymbol{\mathcal{E}}}_T & \text{if } \alpha = 1 \\ \varnothing & \text{if } \alpha = 0 \end{cases} \tag{5.10}$$

Thus, the problem of interest (for randomized strategies) can be expressed as:

Jointly Randomized Strategy Problem - Unfolding

$$\begin{aligned} \inf_{\delta} \quad & \mathbb{P}^{\delta}_{s_0}[\, \boldsymbol{\mathcal{E}}_T \cup \overline{\boldsymbol{\mathcal{E}}}_{\alpha,T} \,] \\ \text{s.t.} \quad & \mathbb{E}^{\delta}_{s_0}[\, \mathrm{TS}^{\mathcal{G}} \,] \;<\; \nu \end{aligned} \tag{5.11}$$

5.3 PROBLEM FORMULATION

To solve the problem (5.11) under randomized strategies, we will define the mixed strategy counterpart. Finding an optimal mixed strategy, we can define the optimal randomized strategy from it. For that, we define first the pure strategy problem related to (5.11).

For $\gamma > 0$ fixed, we suppose that from each leaf in the unfolding $\mathcal{M}_T$, a γ-optimal strategy of the expected SSP problem, π_γ, is fixed and always played, see eq. (5.6). The idea is to optimize from $t = 0$ up to the depth T. We let $\underline{\Pi}_\gamma$ the set of pure memoryless strategies, such that $\overline{\pi} \in \underline{\Pi}_\gamma$ plays from $t = 0$ as a pure strategy $\pi \in \Pi$ in $\mathcal{M}_T$, and after T-steps as the strategy π_γ from each leaf s_T. Thus, we represent the strategies in $\underline{\Pi}_\gamma$ as:

$$\overline{\pi} \;:=\; \pi\pi_\gamma \;\in\; \underline{\Pi}_\gamma\,.$$

5.3.1 *PURE STRATEGY PROBLEM*

We define the optimization problem for pure strategies as follows. We call **objective function for pure strategies** $\overline{\pi} \in \underline{\Pi}_\gamma$, parametrized by $\alpha \in \{0,1\}$ and fixed T, the function[2]:

$$\mathbb{J}_{\alpha,T}(\overline{\pi}) \quad := \quad \mathbb{P}^{\overline{\pi}}_{s_0}[\,\boldsymbol{\mathcal{E}}_T \cup \overline{\boldsymbol{\mathcal{E}}}_{\alpha,T}\,]\ , \tag{5.12}$$

and so, the **pure strategy primal problem** related to (5.11) is defined by:

Pure Strategy Primal Problem

$$\begin{array}{lll} [\textbf{P-PS}]_{<\nu} & \displaystyle\min_{\overline{\pi}} & \mathbb{J}_{\alpha,T}(\overline{\pi}) \\ & & \text{s.t.}\quad \mathbb{E}^{\overline{\pi}}_{s_0}[\,\mathrm{TS}^{\mathcal{G}}\,] \;<\; \nu \end{array} \tag{5.13}$$

When the context is clear, we write P-PS to refer to the problem (5.13) without mentioning the inequality and the threshold ν. A pure strategy $\overline{\pi} \in \underline{\Pi}_\gamma$ is said **feasible** if it satisfies the constraint in expectation. In the following, the **pure strategy optimal value** will be denoted as:

$$\mathbb{J}^{\mathrm{p}*}_{\alpha,T} \;:=\; \min_{\overline{\pi}} \left\{ \mathbb{J}_{\alpha,T}(\overline{\pi}) \;\middle|\; \mathbb{E}^{\overline{\pi}}_{s_0}[\,\mathrm{TS}^{\mathcal{G}}] \;<\; \nu \right\}. \tag{5.14}$$

Recall that there is a finite number of strategies. We say that a strategy $\overline{\pi}^*_\alpha \in \underline{\Pi}_\gamma$ is optimal for the P-PS problem if $\overline{\pi}^*_\alpha$ is feasible and

$$\mathbb{J}_{\alpha,T}(\overline{\pi}^*_\alpha) \quad = \quad \mathbb{J}^{\mathrm{p}*}_{\alpha,T}\,.$$

We represent in the Figure 5.1 the region generated by the values of the objective function for pure strategies (5.12) and the respective constraint in expectation. Denoting by $K \in \mathbb{N}$ the number of pure strategies, this region is a discrete set of points, each one assigned to one pure strategy $\overline{\pi}_k \in \underline{\Pi}_\gamma$, $k = 1, ..., K$. More precisely, it is the set:

$$\bigcup_{k=1}^{K} \left\{ \left(\mathbb{J}_{\alpha,T}(\overline{\pi}_k),\, \mathbb{E}^{\overline{\pi}_k}_{s_0}[\mathrm{TS}^{\mathcal{G}}] \right) \right\} \;\subseteq\; ([0,1] \cap \mathbb{Q}) \times \mathbb{Q}\,. \tag{5.15}$$

Since by hypothesis the set $\underline{\Pi}_\gamma$ is discrete, the P-PS problem (5.13) is non-convex for $K > 1$. On the other hand, if there is a single strategy in $\underline{\Pi}_\gamma$, i.e., if $K = 1$, then it is the strategy of the SSP problem π_γ which is played from the initial state s_0 and satisfies the constraint in expectation by Assumption 4.2.3. Thus, the solution of the problem (5.11) will be π_γ in such a case. In the following, we suppose that $K > 1$. We introduce the problem under mixed strategies in the next section, to extend the solution set of the pure strategy problem (5.13). This can be understood as a convexification of pure strategies.

2 To be more precise, this function should be denoted as $\mathbb{J}_{s_0,\alpha,T}$ due to the fixed parameters. However, to relax the notation, we simply write $\mathbb{J}_{\alpha,T}$.

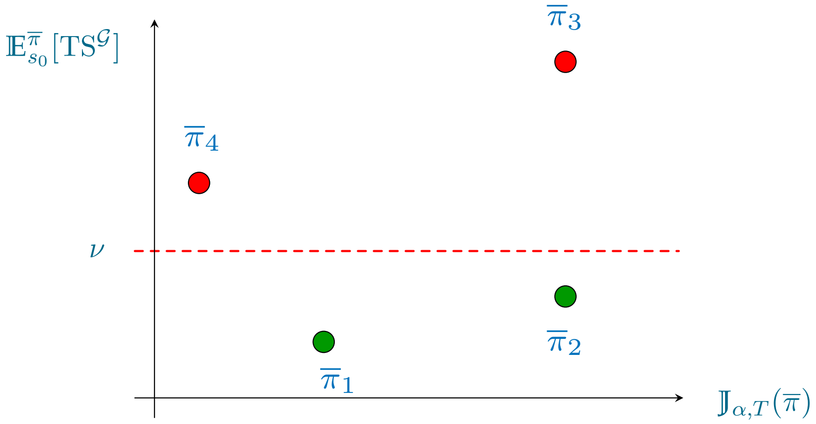


Figure 5.1: Points representing the values of the objective function and the constraint in expectation of the P-PS problem (5.13) under pure strategies. In this figure, $\overline{\pi}_1$ and $\overline{\pi}_2$ satisfy the constraint for the threshold ν, contrary to strategies $\overline{\pi}_3$ and $\overline{\pi}_4$.

5.3.2 *MIXED STRATEGY PROBLEM*

As we have seen in Section 2.6.3, a **mixed strategy** is a function of the form $\sigma : \underline{\Pi} \to [0,1]$, assigning a probability distribution over pure strategies. We define the set of mixed strategies over $\underline{\Pi}_\gamma$ by:

$$\Delta[\,\underline{\Pi}_\gamma] \;\coloneqq\; \left\{ \sigma : \underline{\Pi}_\gamma \longrightarrow [0,1] \;\;\middle|\;\; \sum_{k=1}^{K} \sigma(\overline{\pi}_k) \;=\; 1 \right\}. \tag{5.16}$$

This is also called a simplex in $\mathbb{R}^K$, which is compact, convex, and of dimension $K-1$. Considering the region generated by pure strategies, see eq. (5.15), we can represent graphically the convex combinations of its points as is shown in the Figure 5.2. Specifically, it is the **convex hull** of its K vertices, where each one is generated by a pure strategy $\overline{\pi}_k \in \underline{\Pi}_\gamma$, $k = 1, \dots, K$; i.e.,

$$\mathrm{Conv}\left(\bigcup_{k=1}^{K} \left\{ \left(\mathbb{J}_{\alpha,T}(\overline{\pi}_k),\, \mathbb{E}^{\overline{\pi}_k}_{s_0}[\mathrm{TS}^{\mathcal{G}}]\, \right) \right\} \right) \;\subseteq\; [0,1] \times \mathbb{R}\,. \tag{5.17}$$

Each point $(\,\mathbb{J}_{\alpha,T}(\overline{\pi}_k),\, \mathbb{E}^{\overline{\pi}_k}_{s_0}[\mathrm{TS}^{\mathcal{G}}]\,)$ has assigned a weight, in such a way that the weights are all non-negative and sum to one. This weights correspond to the values assigned by a mixed strategy. For each choice of weights, the resulting convex combination is a point in the convex hull (either in the interior, or on a vertex, or on an edge), and the whole convex hull can be formed by choosing weights in all possible ways, i.e., choosing mixed strategies from $\Delta[\,\underline{\Pi}_\gamma]$.

Based on this, we can define an extended probability space with a new probability measure representing the joint probability distribution due to the mixed strategies and the distribution between states, see Section 2.6.4 for details. Thus, we can write $(\Omega^\sigma_{s_0}, \mathcal{B}(\Omega^\sigma_{s_0}), \mathbb{P}^\sigma_{s_0})$ the induced probability space when a mixed strategy σ is fixed. In addition, from the definition of the expectation operator, see eq. (2.18), it is straightforward to verify that for a measurable function F and a mixed strategy $\sigma \in \Delta[\underline{\Pi}_\gamma]$, it holds:

$$\mathbb{E}^{\sigma}_{s_0}[\,F\,] \;=\; \sum_{k=1}^{K} \sigma(\overline{\pi}_k)\, \mathbb{E}^{\overline{\pi}_k}_{s_0}[\,F\,]\,. \tag{5.18}$$

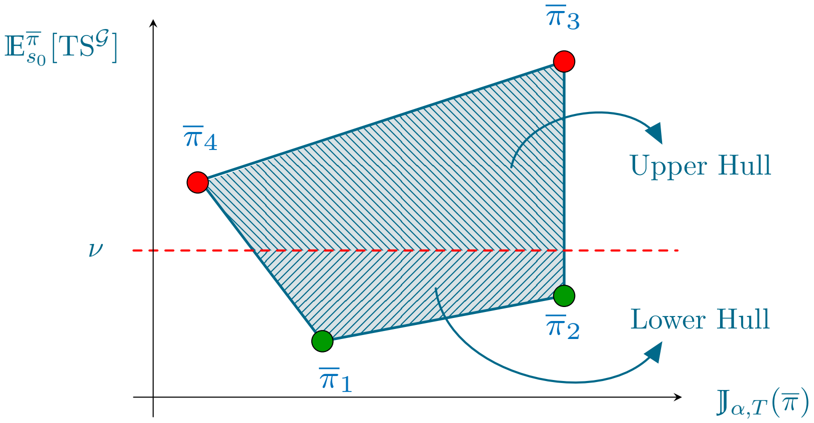


Figure 5.2: Graphical representation of the Convex Hull (5.17).

Proposition 5.3.1

The expectation $\mathbb{E}^{\sigma}_{s_0}[\mathrm{TS}^{\mathcal{G}}]$ and $\mathbb{J}_{\alpha,T}(\sigma)$ are continuous for each $\sigma \in \Delta[\underline{\Pi}_\gamma]$.

Proof. See Proof B.2.1 in Appendix B.2. ■

Note that the convex hull (5.17) can be partitioned into two convex polygons with the continuous line $[0,1] \times \{\nu\}$, generating the **upper hull** and the **lower hull**. The upper and lower hulls represent resp. the strategies that are satisfying and are not satisfying the constraint in the expectation. Note that the lower hull is not empty by Assumption 4.2.3.

Coming back to the problem under consideration for pure strategies, specifically the P-PS problem (5.13), we define the **mixed strategy primal problem** for P-PS, as:

Mixed Strategy Primal Problem

$$[\textbf{P-MS}]_{<\nu} \qquad \inf_{\sigma} \quad \mathbb{J}_{\alpha,T}(\sigma) \qquad \text{s.t.} \quad \mathbb{E}^{\sigma}_{s_0}[\,\mathrm{TS}^{\mathcal{G}}\,] \;<\; \nu \tag{5.19}$$

where $\mathbb{J}_{\alpha,T}(\sigma)$ is understood as the objective function (5.12) for mixed strategies. Again, when the context is clear we write P-MS to refer to the problem (5.19) without mentioning the inequality and the threshold ν. A mixed strategy $\sigma \in \Delta[\underline{\Pi}_\gamma]$ is said **feasible** if it satisfies the constraint in expectation. In the following, the **mixed strategy optimal value** will be denoted as:

$$\mathbb{J}^{\mathrm{m}*}_{\alpha,T} \;:=\; \inf_{\sigma}\;\Big\{\; \mathbb{J}_{\alpha,T}(\sigma) \;\;\big|\;\; \mathbb{E}^{\sigma}_{s_0}[\,\mathrm{TS}^{\mathcal{G}}\,] \;<\; \nu \;\Big\}\,, \tag{5.20}$$

and, for $\kappa > 0$, we say that a mixed strategy $\sigma^*_{\alpha,\kappa} \in \Delta[\,\underline{\Pi}_\gamma]$ is κ-optimal for the P-MS problem if $\sigma^*_{\alpha,\kappa}$ is feasible and

$$\mathbb{J}_{\alpha,T}(\sigma^*_{\alpha,\kappa}) \;\leq\; \mathbb{J}^{\mathrm{m}*}_{\alpha,T} \;+\; \kappa\,.$$

Note that optimizing with mixed strategies $\sigma \in \Delta[\,\underline{\Pi}_\gamma]$ means to find marginal probabilities $\sigma(\overline{\pi}_k)$ assigned to each pure strategy $\overline{\pi}_k \in \underline{\Pi}_\gamma$, $k = 1, ..., K$; while the convex combination of the objective function of the pure strategies is minimized and the convex combination of the expected constraint is satisfied. Thus, based on eq. (5.18), the P-MS problem (5.19) can also be written as[3]:

$$\begin{aligned} \inf_{\sigma} \quad & \sum_{k=1}^{K} \sigma(\overline{\pi}_k)\, \mathbb{J}_{\alpha,T}(\overline{\pi}_k) \\ \text{s.t.} \quad & \sum_{k=1}^{K} \sigma(\overline{\pi}_k)\, \mathbb{E}^{\overline{\pi}_k}_{s_0}[\mathrm{TS}^{\mathcal{G}}] \;<\; \nu \end{aligned} \tag{5.21}$$

In addition, we have the following proposition relating the optimal values for mixed and pure strategies.

Proposition 5.3.2

Mixed strategies always provide optimal values that are at least as good as the values obtained with pure strategies, i.e.,

$$\mathbb{J}^{\mathrm{m}*}_{\alpha,T} \;\leq\; \mathbb{J}^{\mathrm{p}*}_{\alpha,T}$$

Proof. See Proof B.2.2 in Appendix B.2. ■

5.4 SOLUTION METHODOLOGY

Here, the solution will be obtained by solving a dual optimization problem, see for instance [23, 24]. Recall that the P-PS problem (5.13) is not convex under the assumption of $K > 1$, and using mixed strategies we may enable to obtain an optimal value that is better than under pure strategies, see Proposition 5.3.2. On the other hand, we will see that for the P-MS problem (5.19), it is possible to construct an optimal solution by duality theory [24]. Furthermore, we will see that the optimal dual solution related to the problems P-PS and P-MS are the same.

5.4.1 *LAGRANGIAN-BASED APPROACH*

We denote the **Lagrange function** for pure strategies $\overline{\pi} \in \underline{\Pi}_\gamma$ as $\mathbb{L}_{\alpha,T}(\overline{\pi}, \lambda)$, and for mixed strategies $\sigma \in \Delta[\,\underline{\Pi}_\gamma\,]$ as $\mathbb{L}_{\alpha,T}(\sigma, \lambda)$, where $\lambda \in \mathbb{R}^+_0$. Since pure strategies can be seen as a particular case of mixed

3 For a general (continuous) formulation with mixed strategies, see [9].

strategies (assigning probability one), we only define $\mathbb{L}_{\alpha,T}(\sigma,\lambda)$. From the P-MS problem in eq. (5.19), we define:

$$\mathbb{L}_{\alpha,T}(\sigma,\lambda) \;\; := \;\; \mathbb{J}_{\alpha,T}(\sigma) \;\; + \;\; \lambda\Big(\mathbb{E}^{\sigma}_{s_0}[\mathrm{TS}^{\mathcal{G}}] \; - \; \nu\Big)\,. \tag{5.22}$$

Note that if we take the maximum above with respect to $\lambda \geq 0$, we recover the primal problem. Indeed, if a strategy is feasible, i.e., if it satisfies the constraint in the expectation, then the best we can do is to set $\lambda \to 0$, since the expected value in the constraint is strictly less than ν. Thus, the Lagrange function will be equal to the objective function. On the other hand, if the constraint is violated, i.e., the expected value in the constraint is higher than ν for some strategy, then the supremum on the Lagrange function will be infinite. The latter is obtained by choosing a very large $\lambda \to +\infty$. Based on the above, finding an optimal pure or mixed strategy will be resp. equivalent to finding:

$$\min_{\overline{\pi}} \; \sup_{\lambda \geq 0} \; \mathbb{L}_{\alpha,T}(\overline{\pi},\lambda) \qquad \text{or} \qquad \inf_{\sigma} \; \sup_{\lambda \geq 0} \; \mathbb{L}_{\alpha,T}(\sigma,\lambda)\,. \tag{5.23}$$

However, we have no certainty of finding the solution in a simple way, because this problem is not easy to solve. On the other hand, if we reversed the order of the maximization over $\lambda \geq 0$ and the minimization over the strategies, then the problem is more tractable. Thus, we define for $\lambda \geq 0$ fixed the **Lagrange dual function** for pure and mixed strategies resp. by:

$$\begin{aligned} \mathbb{L}^{\mathrm{p}}_{\alpha,T}(\lambda) \;\; &:= \;\; \min_{\overline{\pi}} \;\; \mathbb{L}_{\alpha,T}(\overline{\pi},\lambda) \\ \mathbb{L}^{\mathrm{m}}_{\alpha,T}(\lambda) \;\; &:= \;\; \inf_{\sigma} \;\; \mathbb{L}_{\alpha,T}(\sigma,\lambda) \end{aligned} \tag{5.24}$$

Concerning the optimal strategies for the above problems, we say first that a pure strategy $\overline{\pi}^{*}_{\alpha,\lambda} \in \underline{\Pi}_{\gamma}$ is optimal for $\mathbb{L}^{\mathrm{p}}_{\alpha,T}(\lambda)$ if:

$$\mathbb{L}_{\alpha,T}(\overline{\pi}^{*}_{\alpha,\lambda},\lambda) \quad = \quad \mathbb{L}^{\mathrm{p}}_{\alpha,T}(\lambda)\;,$$

and second, for $\vartheta > 0$, we say that a mixed strategy $\sigma^{*}_{\alpha,\vartheta,\lambda} \in \Delta[\,\underline{\Pi}_{\gamma}\,]$ is ϑ-optimal for $\mathbb{L}^{\mathrm{m}}_{\alpha,T}(\lambda)$, if:

$$\mathbb{L}_{\alpha,T}(\sigma^{*}_{\alpha,\vartheta,\lambda},\lambda) \quad \leq \quad \mathbb{L}^{\mathrm{m}}_{\alpha,T}(\lambda) \; + \; \vartheta\,.$$

However, as we will see after in the Proposition 5.4.4, the existence of at least one mixed strategy $\sigma^{*}_{\alpha,\lambda}$ such that $\mathbb{L}_{\alpha,T}(\sigma^{*}_{\alpha,\lambda},\lambda) = \mathbb{L}^{\mathrm{m}}_{\alpha,T}(\lambda)$, is guaranteed.

In the following, we show the monotonicity of the expected value in the constraint. This is important, since taking a nondecreasing sequence of the dual variables, the expected value will descend monotonically to the constraint value and in such a way, we will construct an optimal solution.

Lemma 5.4.1

Suppose that we can obtain $\overline{\pi}^*_{\alpha,\lambda} \in \arg\min_{\overline{\pi}} \mathbb{L}_{\alpha,T}(\overline{\pi},\lambda)$ for each $\lambda \geq 0$. Then, $\lambda \mapsto \mathbb{E}_{s_0}^{\overline{\pi}^*_{\alpha,\lambda}}[\mathrm{TS}^{\mathcal{G}}]$ is nonincreasing.

Proof. See Proof B.2.3 in Appendix B.2. ∎

We introduce now the **pure strategy dual problem** related to the P-PS problem (5.13), and the **mixed strategy dual problem** related to the P-MS problem (5.19) resp. by:

Pure and Mixed Strategy Dual Problems

$$\begin{array}{lll} [\textbf{D-PS}]_\nu & \sup\limits_{\lambda\geq 0} & \mathbb{L}^{\mathrm{p}}_{\alpha,T}(\lambda) \\ [\textbf{D-MS}]_\nu & \sup\limits_{\lambda\geq 0} & \mathbb{L}^{\mathrm{m}}_{\alpha,T}(\lambda) \end{array} \tag{5.25}$$

When the context is clear we write D-PS and D-MS to refer resp. to the above problems without mentioning the threshold ν. In the following, the **optimal dual value** for pure and mixed strategies will be denoted resp. as:

$$\begin{array}{lll} \mathbb{L}^{\mathrm{p}*}_{\alpha,T} & := & \sup\limits_{\lambda\geq 0} \mathbb{L}^{\mathrm{p}}_{\alpha,T}(\lambda) \\ \mathbb{L}^{\mathrm{m}*}_{\alpha,T} & := & \sup\limits_{\lambda\geq 0} \mathbb{L}^{\mathrm{m}}_{\alpha,T}(\lambda) \end{array} \tag{5.26}$$

and if there exists $\lambda^* \geq 0$ optimal for the above problems, we write:

$$\begin{array}{lll} \mathbb{L}^{\mathrm{p}}_{\alpha,T}(\lambda^*) & = & \mathbb{L}^{\mathrm{p}*}_{\alpha,T} \\ \mathbb{L}^{\mathrm{m}}_{\alpha,T}(\lambda^*) & = & \mathbb{L}^{\mathrm{m}*}_{\alpha,T} \end{array}$$

We will show that effectively, the optimal dual value λ^* is the same for the two dual problems (5.25). It turns out that the dual problems are always concave, even when the initial problem is not convex [24]. Indeed, the set $\{\lambda \geq 0\}$ is convex and the minimum of the Lagrange functions are concave functions of λ. The latter holds because in eq. (5.24), each function is pointwise minimum of an affine function. Formally, we have the following.

Proposition 5.4.2

The Lagrange functions $\lambda \mapsto \mathbb{L}_{\alpha,T}(\overline{\pi},\lambda)$ and $\lambda \mapsto \mathbb{L}_{\alpha,T}(\sigma,\lambda)$ are linear on $\lambda \geq 0$. Moreover, the Lagrange dual functions $\lambda \mapsto \mathbb{L}^{\mathrm{p}}_{\alpha,T}(\lambda)$ and $\lambda \mapsto \mathbb{L}^{\mathrm{m}}_{\alpha,T}(\lambda)$ are concave.

Proof. See Proof B.2.4 in Appendix B.2. ∎

In this way, maximizing the Lagrange dual function over the values of $\lambda \geq 0$ is a more tractable problem than solving (5.23). An important result connecting primal and dual problems is the **weak duality theorem**, which states the following [23].

Lemma 5.4.3

Let $\mathbb{J}^{p*}_{\alpha,T}$ and $\mathbb{J}^{m*}_{\alpha,T}$ resp. the optimal values of the P-PS and P-MS problem, see eq. (5.14) and eq. (5.20). Then,

$$\mathbb{L}^{p*}_{\alpha,T} \leq \mathbb{J}^{p*}_{\alpha,T} \quad \text{and} \quad \mathbb{L}^{m*}_{\alpha,T} \leq \mathbb{J}^{m*}_{\alpha,T} \tag{5.27}$$

Proof. See Proof B.2.5 in Appendix B.2. ∎

The non-negative difference between the above values (5.27), for the pure or mixed problem, is known as **duality gap**; and if it is zero, we say that there is **strong duality**, while otherwise, we say that there is only a **weak duality**. Strong duality holds if the optimization problem is convex and a feasible strategy exists (i.e., it satisfies the constraint in expectation). In that case, the solution of the primal and dual problems are equivalent, in the sense that an optimal strategy can be constructed by the (optimal) Lagrange dual function. However, it is not the case for the D-PS and P-PS problems, but it holds for the problems D-MS and P-MS as shown in the following.

Proposition 5.4.4

There is strong duality for mixed strategy problems D-MS and P-MS, i.e.,

$$\mathbb{L}^{m*}_{\alpha,T} = \mathbb{J}^{m*}_{\alpha,T} . \tag{5.28}$$

Moreover, for $\lambda \geq 0$ fixed, there exists $\sigma^*_{\alpha,\lambda} \in \Delta[\underline{\Pi}_\gamma]$ such that:

$$\mathbb{L}_{\alpha,T}(\sigma^*_{\alpha,\lambda}, \lambda) = \mathbb{L}^{m}_{\alpha,T}(\lambda) . \tag{5.29}$$

Proof. See Proof B.2.6 in Appendix B.2. ∎

Thus, there is no duality gap for mixed strategies and the existence of at least one mixed strategy reaching the minimum in the Lagrange function $\mathbb{L}_{\alpha,T}(\sigma, \lambda)$ is guaranteed for $\lambda \geq 0$ fixed. On the other hand, the optimal dual values of the dual problems for pure and mixed strategies, see eq. (5.26), are the same as shown below.

Proposition 5.4.5

Consider the dual problems D-PS and D-MS of (5.25). Then,

$$\mathbb{L}^{\mathrm{p}*}_{\alpha,T} \quad = \quad \mathbb{L}^{\mathrm{m}*}_{\alpha,T} \tag{5.30}$$

Proof. See Proof B.2.8 in Appendix B.2. ∎

In the following, we consider a threshold $\nu' \leq \nu$ to show an equivalence between two problems. More precisely, to show how to find an optimal solution for each one. The first one is the mixed strategy primal problem (5.19), but considering the inequality in the constraint in expectation to be not strict with threshold ν', as the one for the $[\text{P-MS}]_{<\nu}$ problem with ν. Wherefore, such a problem is denoted by $[\text{P-MS}]_{\leq\nu'}$. The second problem is naturally understood as the underlying mixed strategy dual problem (5.25), but considering the threshold ν'. So, this is denoted as $[\text{D-MS}]_{\nu'}$. The following proposition shows how to find a solution for the problems $[\text{P-MS}]_{\leq\nu'}$ and $[\text{D-MS}]_{\nu'}$. Analysing such problems, we will proceed to show how to find a solution for the original $[\text{P-MS}]_{<\nu}$ problem. The feasibility of the $[\text{P-MS}]_{\leq\nu'}$ problem depends on the threshold ν', which will also be explained a little further.

Proposition 5.4.6

Let $\sigma^*_\alpha \in \Delta[\,\underline{\Pi}_\gamma]$, $\lambda^* \geq 0$ and $\nu' \leq \nu$ a threshold. The following statements are equivalent:

(i) σ^*_α is an optimal mixed strategy solution of the $[\text{P-MS}]_{\leq\nu'}$ problem and λ^* is an optimal dual solution of the underlying $[\text{D-MS}]_{\nu'}$ problem.

(ii) $\sigma^*_\alpha \in \displaystyle\arg\min_\sigma \mathbb{L}_{\alpha,T}(\sigma,\lambda^*)$, $\mathbb{E}^{\sigma^*_\alpha}_{s_0}[\,\mathrm{TS}^{\mathcal{G}}] \leq \nu'$, and

$$\lambda^*\Big(\mathbb{E}^{\sigma^*_\alpha}_{s_0}[\,\mathrm{TS}^{\mathcal{G}}] - \nu'\Big) \quad = \quad 0 \tag{5.31}$$

Proof. See Proof B.2.9 in Appendix B.2. ∎

We focus in the sufficient conditions of the previous proposition to find an optimal solution for the $[\text{P-MS}]_{\leq\nu'}$ problem. Recall that we aim to find a mixed strategy such that the constraint in expectation is satisfies with the strict inequality for the threshold ν. Before relating such problems, we make some comments.

First, note that if we find a strategy $\sigma^*_\alpha \in \arg\min_\sigma \mathbb{L}_{\alpha,T}(\sigma,\lambda^*)$, such that $\mathbb{E}^{\sigma^*_\alpha}_{s_0}[\,\mathrm{TS}^{\mathcal{G}}] \leq \nu'$ for $\lambda^* = 0$, then σ^*_α will be a solution for the $[\text{P-MS}]_{\leq\nu'}$ problem. However, this does not provide much information to effectively solve such a problem. More precisely, the following holds under the $[\text{P-MS}]_{\leq\nu'}$ context:

$$\begin{aligned}\mathbb{J}^{\mathrm{m}*}_{\alpha,T} &= \inf_{\sigma}\left\{\ \mathbb{J}_{\alpha,T}(\sigma)\ \mid\ \mathbb{E}^{\sigma}_{s_0}[\,\mathrm{TS}^{\mathcal{G}}]\ \leq\ \nu'\ \right\}\\ &\qquad\text{(by definition (5.20) of the mixed strategy optimal value)}\\ &= \inf_{\sigma}\left\{\ \mathbb{L}_{\alpha,T}(\sigma,0)\ \mid\ \mathbb{E}^{\sigma}_{s_0}[\,\mathrm{TS}^{\mathcal{G}}]\ \leq\ \nu'\ \right\}\\ &\qquad\text{(by definition (5.22) of the Lagrange function)}\end{aligned}$$

Thus, when considering a dual variable to be $\lambda^* = 0$, and by using the fact that a mixed strategy exists for such a dual variable by Proposition 5.4.4, i.e., a strategy $\sigma^*_{\alpha,0}$ such that $\mathbb{L}^{\mathrm{m}}_{\alpha,T}(0) = \mathbb{L}^{\mathrm{m}}_{\alpha,T}(\sigma^*_{\alpha,0}, 0)$, and satisfying the constraint in expectation, then $\sigma^*_{\alpha,0}$ satisfies the sufficient conditions of the Proposition 5.4.6 and thus, $\sigma^*_{\alpha,0}$ and $\lambda^* = 0$ are resp. a solution of the problems $[\text{P-MS}]_{\leq\nu'}$ and $[\text{D-MS}]_{\nu'}$. However, we notice that it is the same as solving directly the $[\text{P-MS}]_{\leq\nu'}$ problem as is shown above. That is why we will focus in strictly positive dual variables $\lambda^* > 0$.

Concerning the feasibility of the $[\text{P-MS}]_{\leq\nu'}$ problem, i.e., the existence of at least one strategy satisfying the not strict inequality in the constraint in expectation with the threshold $\nu' \leq \nu$, we have the following. First, we know there is at least one feasible strategy for the $[\text{P-MS}]_{<\nu}$ problem. More precisely, the strategy π_γ of the expected SSP-problem (5.6), which gives an expected SSP-value from the initial state s_0, denoted here by $\mathbb{SP}^{\pi_\gamma}_{s_0}$, less than ν. We can thus consider the threshold $\nu' = \nu - \varsigma$, where

$$0\ \leq\ \varsigma\ \leq\ \nu - \mathbb{SP}^{\pi_\gamma}_{s_0}\ . \tag{5.32}$$

In such a way, the problem $[\text{P-MS}]_{\leq\nu'}$ is feasible as well.

Based on these observations and on Proposition 5.4.6, a sufficient condition for optimality under mixed strategies to solve the $[\text{P-MS}]_{\leq\nu'}$ problem, can be written as follows.

Corollary 5.4.7

Let $\lambda^* > 0$ and $\nu' = \nu - \varsigma$ a threshold, where ς satisfies the eq. (5.32). Suppose that there is $\sigma^*_\alpha \in \arg\min_{\sigma}\ \mathbb{L}_{\alpha,T}(\sigma,\lambda^*)$ such that:

$$\mathbb{E}^{\sigma^*_\alpha}_{s_0}[\,\mathrm{TS}^{\mathcal{G}}]\ =\ \nu'\ . \tag{5.33}$$

Then, σ^*_α is an optimal mixed strategy solution of the $[\text{P-MS}]_{\leq\nu'}$ problem, λ^* is an optimal dual solution of the underlying $[\text{D-MS}]_{\nu'}$ problem, and the following holds:

$$\mathbb{L}^{\mathrm{m}*}_{\alpha,T} = \mathbb{L}^{\mathrm{m}}_{\alpha,T}(\lambda^*) = \mathbb{L}^{\mathrm{p}*}_{\alpha,T} = \mathbb{L}^{\mathrm{p}}_{\alpha,T}(\lambda^*) = \mathbb{J}^{\mathrm{m}*}_{\alpha,T} = \mathbb{J}_{\alpha,T}(\sigma^*_\alpha) \leq \mathbb{J}^{\mathrm{p}*}_{\alpha,T}$$

Proof. See Proof B.2.10 in Appendix B.2. ∎

The idea is then to find a mixed strategy σ_α^* satisfying the hypothesis of the Corollary 5.4.7 to solve the $[\text{P-MS}]_{\leq\nu'}$ problem. Note that in such a case, this strategy gives an expected value in the constraint equal to ν' as is show the Figure 5.3. More precisely, it holds:

$$\mathbb{E}_{s_0}^{\sigma_\alpha^*}[\,\mathrm{TS}^{\mathcal{G}}\,] \;=\; \nu' \;=\; \nu - \varsigma \;\leq\; \nu\,. \tag{5.34}$$

We analyze three cases concerning the fixed value of ς to define the $[\text{P-MS}]_{\leq\nu'}$ problem:

(1) If $\varsigma = \nu - \mathbb{SP}_{s_0}^{\pi_\gamma}$ above, the expectation is the expected SSP-value and then, the solution to the $[\text{P-MS}]_{\leq\nu'}$ problem will be the pure strategy of π_γ, which exists by Assumption 4.2.3. In such a case, this strategy satisfies the constraint in expectation for ν, but could not give the optimal value for the $[\text{P-MS}]_{<\nu}$ problem.

(2) If $0 < \varsigma < \nu - \mathbb{SP}_{s_0}^{\pi_\gamma}$, then the expected value under a mixed strategy that we focus (equal to $\nu' = \nu - \varsigma$), will be between the expected SSP-value and ν. Then, such a mixed strategy may be constructed with another (better) pure strategy than the one of the expected SSP-problem. In other words, if there is a pure strategy $\overline{\pi}_\alpha$ such that:

$$\mathbb{SP}_{s_0}^{\pi_\gamma} \;<\; \mathbb{E}_{s_0}^{\overline{\pi}_\alpha}[\,\mathrm{TS}^{\mathcal{G}}\,] \;<\; \nu\,,$$

then we can eventually use $\overline{\pi}_\alpha$ to construct a solution of the $[\text{P-MS}]_{\leq\nu'}$ problem, with ς satisfying the following:

$$0 \;<\; \varsigma \;\leq\; \nu - \mathbb{E}_{s_0}^{\overline{\pi}_\alpha}[\,\mathrm{TS}^{\mathcal{G}}\,]\,.$$

Note in particular that if ς is fixed to be $\varsigma = \nu - \mathbb{E}_{s_0}^{\overline{\pi}_\alpha}[\,\mathrm{TS}^{\mathcal{G}}]$, the solution of the $[\text{P-MS}]_{\leq\nu'}$ problem could be the pure strategy $\overline{\pi}_\alpha$, provided that $\overline{\pi}_\alpha \in \arg\min_\sigma \mathbb{L}_{\alpha,T}(\sigma,\lambda^*)$ for a dual variable $\lambda^* > 0$.

(3) If $\varsigma = 0$, then the $[\text{P-MS}]_{\leq\nu'}$ problem is the same as the $[\text{P-MS}]_{\leq\nu}$ problem. By using the Corollary 5.4.7, we could have a mixed strategy σ_α^* such that $\mathbb{E}_{s_0}^{\sigma_\alpha^*}[\,\mathrm{TS}^{\mathcal{G}}] = \nu$ and then, this strategy does not solve the original $[\text{P-MS}]_{<\nu}$ problem, since the inequality in the constraint under σ_α^* is not strict. However, as we will see in next, an optimal strategy can be constructed as a combination of at most two pure strategies: one satisfying the strict inequality in the constraint under ν and another one that does not. Combining them, we will have a mixed strategy σ_α^* such that $\mathbb{E}_{s_0}^{\sigma_\alpha^*}[\,\mathrm{TS}^{\mathcal{G}}] = \nu$. We can thus "perturb" the solution, defining another mixed strategy satisfying the strict inequality in the constraint from such a strategy and to be near to the optimum, more precisely, to define a κ-optimal strategy, for $\kappa > 0$ small enough. We focus in this approach and we will quantify the perturbation-value in what follows.

Proposition 5.4.8

There exist two pure strategies $\overline{\pi}'_\alpha,\ \overline{\pi}''_\alpha \in \arg\min_{\overline{\pi}} \mathbb{L}_{\alpha,T}(\overline{\pi}, \lambda^*)$ for a dual variable $\lambda^* > 0$, defining a mixed strategy $\sigma^*_\alpha \in \Delta[\,\{\overline{\pi}'_\alpha,\ \overline{\pi}''_\alpha\}\,]$ to be a solution of the $[\text{P-MS}]_{\leq\nu}$ problem.

Proof. See Proof B.2.11 in Appendix B.2. ∎

Based on the previous proposition and the Corollary 5.4.7, we claim that the two pure strategies satisfy $\mathbb{E}^{\overline{\pi}''_\alpha}_{s_0}[\,\mathrm{TS}^{\mathcal{G}}\,] < \nu \leq \mathbb{E}^{\overline{\pi}'_\alpha}_{s_0}[\,\mathrm{TS}^{\mathcal{G}}\,]$ (see Proof B.2.11 for details). So that an optimal mixed strategy to solve the $[\text{P-MS}]_{\leq\nu}$ problem is of the form $\sigma^*_\alpha \in \Delta[\,\{\overline{\pi}'_\alpha,\ \overline{\pi}''_\alpha\}\,] \subseteq \Delta[\,\underline{\Pi}_\gamma\,]$, which combines convexly over $\overline{\pi}'_\alpha$ and $\overline{\pi}''_\alpha$, i.e., $\sigma^*_\alpha(\overline{\pi}'_\alpha) + \sigma^*_\alpha(\overline{\pi}''_\alpha) = 1$, and also, by the Corollary 5.4.7, be such that $\mathbb{E}^{\sigma^*_\alpha}_{s_0}[\,\mathrm{TS}^{\mathcal{G}}\,] = \nu$. Based on this and the eq. (5.18), it is easy to see that σ^*_α is to be equal to:

$$
\begin{aligned}
\sigma^*_\alpha(\overline{\pi}'_\alpha) &= \frac{\nu - \mathbb{E}^{\overline{\pi}''_\alpha}_{s_0}[\,\mathrm{TS}^{\mathcal{G}}\,]}{\mathbb{E}^{\overline{\pi}'_\alpha}_{s_0}[\,\mathrm{TS}^{\mathcal{G}}\,] - \mathbb{E}^{\overline{\pi}''_\alpha}_{s_0}[\,\mathrm{TS}^{\mathcal{G}}\,]} \\
\sigma^*_\alpha(\overline{\pi}''_\alpha) &= \frac{\mathbb{E}^{\overline{\pi}'_\alpha}_{s_0}[\,\mathrm{TS}^{\mathcal{G}}\,] - \nu}{\mathbb{E}^{\overline{\pi}'_\alpha}_{s_0}[\,\mathrm{TS}^{\mathcal{G}}\,] - \mathbb{E}^{\overline{\pi}''_\alpha}_{s_0}[\,\mathrm{TS}^{\mathcal{G}}\,]}
\end{aligned}
\tag{5.35}
$$

Note that σ^*_α gives an expected value in the constraint to be equal to ν. However, we want a strategy giving the strict inequality in the constraint to solve the $[\text{P-MS}]_{<\nu}$ problem. To do that, we perturb the mixed strategy σ^*_α as it is shown below.

Considering the expected value under the pure strategy $\overline{\pi}''_\alpha$, which is strictly less than ν, we define ς to be:

$$
0 < \varsigma < \nu - \mathbb{E}^{\overline{\pi}''_\alpha}_{s_0}[\,\mathrm{TS}^{\mathcal{G}}\,] \,.
$$

Fixing one on these ς, based on the mixed strategy σ^*_α we define $\sigma^\star_{\alpha,\varsigma} \in \Delta[\,\{\overline{\pi}'_\alpha,\ \overline{\pi}''_\alpha\}\,]$, as the strategy:

$$
\begin{aligned}
\sigma^\star_{\alpha,\varsigma}(\overline{\pi}'_\alpha) &:= \sigma^*_\alpha(\overline{\pi}'_\alpha) - \frac{\varsigma}{\mathbb{E}^{\overline{\pi}'_\alpha}_{s_0}[\,\mathrm{TS}^{\mathcal{G}}\,] - \mathbb{E}^{\overline{\pi}''_\alpha}_{s_0}[\,\mathrm{TS}^{\mathcal{G}}\,]} \\
\sigma^\star_{\alpha,\varsigma}(\overline{\pi}''_\alpha) &:= \sigma^*_\alpha(\overline{\pi}''_\alpha) + \frac{\varsigma}{\mathbb{E}^{\overline{\pi}'_\alpha}_{s_0}[\,\mathrm{TS}^{\mathcal{G}}\,] - \mathbb{E}^{\overline{\pi}''_\alpha}_{s_0}[\,\mathrm{TS}^{\mathcal{G}}\,]}
\end{aligned}
\tag{5.36}
$$

It is straightforward to show that $\sigma^\star_{\alpha,\varsigma}$ satisfying the expected value to be $\mathbb{E}^{\sigma^\star_{\alpha,\varsigma}}_{s_0}[\,\mathrm{TS}^{\mathcal{G}}\,] = \nu - \varsigma$. The following also holds.

Proposition 5.4.9

For any $\kappa > 0$, the mixed strategy $\sigma^\star_{\alpha,\varsigma}$ defined in (5.36) is κ-optimal for the $[\text{P-MS}]_{<\nu}$ problem, i.e.,

$$
\mathbb{J}_{\alpha,T}(\sigma^\star_{\alpha,\varsigma}) \leq \mathbb{J}^{\mathrm{m}*}_{\alpha,T} + \kappa \,.
$$

Proof. See Proof B.2.12 in Appendix B.2. ∎

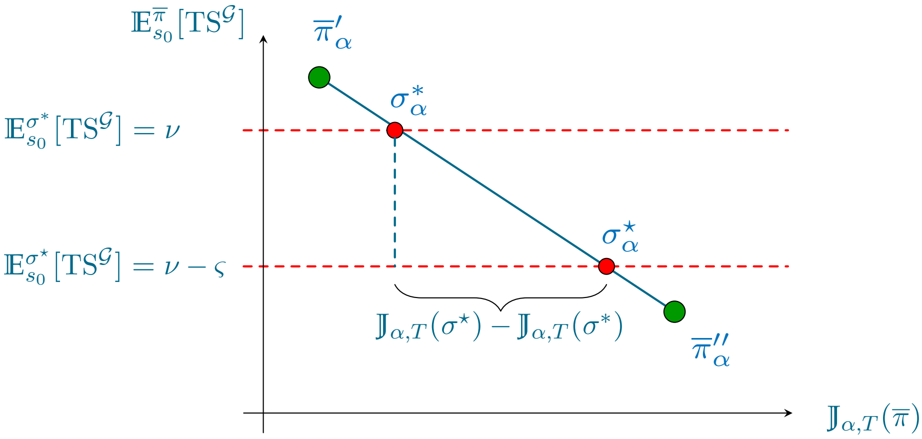


Figure 5.3: Expected values of the strategies σ^*_α and $\sigma^\star_\alpha$. The strategy σ^*_α is built by combining the pure strategies π'_α and π'_α. The other strategy $\sigma^\star_\alpha$ is built by considering σ^*_α but perturbed by a constant ς.

In summary, we have by Proposition 5.4.4 that for mixed strategies, solving the dual problem is equivalent to solving the primal problem, i.e., the D-MS and P-MS problems are equivalents. In addition, by the Proposition 5.4.5, solving the D-MS problem is the same as solving the D-PS problem of pure strategies, i.e., the primal problem for pure strategies and the problem for mixed strategies have the same optimal dual solution λ^*. Thus, to find an optimal mixed strategy strategy solution for the P-MS problem, we find the dual variable λ^* from the dual problem under feasible pure strategies. Using the Corollary 5.4.7 by restricting such a dual variable to be positive, we find two pure strategies, one satisfying the constraint in the expectation and another one that does not. In such a way, we combine them to construct the optimal mixed strategy, which gives the equality in constraint. Since we need a strategy giving the strict inequality in the constraint, we slightly perturbed such a mixed strategy to have another one that effectively satisfies the strict inequality and to be κ-optimal, for $\kappa > 0$ small enough.

5.4.2 *ALGORITHMIC APPROACH*

In this section, we provide a way to compute a pure strategy reaching the minimum in the Lagrange function for a fixed dual variable, i.e., how to solve (5.24) for pure strategies. Thus, we can built a mixed strategy in an iterative way to converge to the optimum shown in Proposition 5.4.8. More precisely, this will help us to compute the strategies:

$$\overline{\pi}^*_{\alpha,\underline{\lambda}_n} \in \arg\min_{\overline{\pi}} \ \mathbb{L}_{\alpha,T}(\overline{\pi},\underline{\lambda}_n) \qquad \text{and} \qquad \overline{\pi}^*_{\alpha,\overline{\lambda}_n} \in \arg\min_{\overline{\pi}} \ \mathbb{L}_{\alpha,T}(\overline{\pi},\overline{\lambda}_n)$$

for fixed dual variables, $\underline{\lambda}_n$ and $\overline{\lambda}_n$, at each iteration $n \in \mathbb{N}$, to construct the mixed strategy of the Proposition 5.4.8, where

$$\underline{\lambda}_n \nearrow \lambda^* \qquad \text{and} \qquad \overline{\lambda}_n \searrow \lambda^* \qquad \text{as} \qquad n \to +\infty \ ,$$

are constructed, e.g., by the so-called bisection method, to approach the optimal dual variable for feasible pure strategies:

$$\lambda^* \;=\; \sup\Big\{\, \lambda \geq 0 \;\mid\; \mathbb{E}_{s_0}^{\overline{\pi}^*_{\alpha,\lambda}}[\,\mathrm{TS}^{\mathcal{G}}] \;\leq\; \nu \;,\; \overline{\pi}^*_{\alpha,\lambda} \in \arg\min_{\overline{\pi}} \; \mathbb{L}_{\alpha,T}(\overline{\pi},\lambda) \Big\} \;.$$

See Proof B.2.11 for details.

Proposition 5.4.10

For $\lambda \geq 0$ fixed, a pure strategy $\overline{\pi}^*_{\alpha,\lambda} \in \arg\min_{\overline{\pi}} \; \mathbb{L}_{\alpha,T}(\overline{\pi},\lambda)$ can be computed by the classical Bellman backward recursion.

Proof. See Proof B.2.13 in Appendix B.2. ∎

We can thus built a mixed strategy at each iteration $n \in \mathbb{N}$, as is shown below.

Proposition 5.4.11

Let $n \in \mathbb{N}$. For $0 < \underline{\lambda}_n \leq \overline{\lambda}_n$ fixed, let $\sigma^*_{\alpha,n} \in \Delta[\{\overline{\pi}^*_{\alpha,\underline{\lambda}_n}, \overline{\pi}^*_{\alpha,\overline{\lambda}_n}\}]$ a mixed strategy combining convexly between the pure strategies:

$$\overline{\pi}^*_{\alpha,\underline{\lambda}_n} \in \arg\min_{\overline{\pi}} \; \mathbb{L}_{\alpha,T}(\overline{\pi},\underline{\lambda}_n) \quad \text{and} \quad \overline{\pi}^*_{\alpha,\overline{\lambda}_n} \in \arg\min_{\overline{\pi}} \; \mathbb{L}_{\alpha,T}(\overline{\pi},\overline{\lambda}_n) \;,$$

and defined as:

$$\begin{aligned}
\sigma^*_{\alpha,n}(\overline{\pi}^*_{\alpha,\underline{\lambda}_n}) \;&=\; \frac{\nu \;-\; \mathbb{E}_{s_0}^{\overline{\pi}^*_{\alpha,\overline{\lambda}_n}}[\mathrm{TS}^{\mathcal{G}}]}{\mathbb{E}_{s_0}^{\overline{\pi}^*_{\alpha,\underline{\lambda}_n}}[\mathrm{TS}^{\mathcal{G}}] \;-\; \mathbb{E}_{s_0}^{\overline{\pi}^*_{\alpha,\overline{\lambda}_n}}[\mathrm{TS}^{\mathcal{G}}]} \\
\sigma^*_{\alpha,n}(\overline{\pi}^*_{\alpha,\overline{\lambda}_n}) \;&=\; \frac{\mathbb{E}_{s_0}^{\overline{\pi}^*_{\alpha,\underline{\lambda}_n}}[\mathrm{TS}^{\mathcal{G}}] \;-\; \nu}{\mathbb{E}_{s_0}^{\overline{\pi}^*_{\alpha,\underline{\lambda}_n}}[\,\mathrm{TS}^{\mathcal{G}}] \;-\; \mathbb{E}_{s_0}^{\overline{\pi}^*_{\alpha,\overline{\lambda}_n}}[\,\mathrm{TS}^{\mathcal{G}}]}
\end{aligned} \tag{5.37}$$

Then, $\mathbb{J}_{\alpha,T}(\sigma^*_{\alpha,n}) \;\longrightarrow\; \mathbb{J}^{\mathrm{m}*}_{\alpha,T}$ as $n \to +\infty$. Moreover, there exists a constant $E \in \mathbb{R}^+$, such that the number of iterations needed $n \in \mathbb{N}$ to achieve a given tolerance $\epsilon > 0$, is such that:

$$n \;\geq\; \log_2\left(E\,\frac{(\overline{\lambda}_0 \,-\, \underline{\lambda}_0)}{\epsilon}\right)$$

Proof. See Proof B.2.14 in Appendix B.2. ∎

The way in which the mixed strategy $\sigma^*_{\alpha,n}$ is computed, it is shown in the Algorithm 1.

Input : $\alpha \in \{0,1\}$, $T \in \mathbb{N}$, $\lambda^+ > 0$, $\epsilon > 0$, π_γ.
Output : Mixed strategy $\sigma^*_{\alpha,n}$.

$(\underline{\lambda}_0, \overline{\lambda}_0) \leftarrow (0, \lambda^+)$
$\overline{\pi}_{\underline{\lambda}_0} \leftarrow \overline{\pi}^*_{\delta,\alpha} \in \arg\min_{\overline{\pi}} \mathbb{J}_{\alpha,T}(\overline{\pi})$
if $\mathbb{E}^{\overline{\pi}_{\underline{\lambda}_0}}_{s_0}[\mathrm{TS}^{\mathcal{G}}] \leq \nu$ **then**
  return $\overline{\pi}_{\underline{\lambda}_0}$
$\overline{\pi}_{\overline{\lambda}_0} \leftarrow \overline{\pi}^*_{\alpha,\overline{\lambda}_0} \in \arg\min_{\overline{\pi}} \mathbb{L}_{\alpha,T}(\overline{\pi}, \overline{\lambda}_0)$
while $\mathbb{E}^{\overline{\pi}_{\overline{\lambda}_0}}_{s_0}[\mathrm{TS}^{\mathcal{G}}] \geq \nu$ **do**
  if $\mathbb{E}^{\overline{\pi}_{\overline{\lambda}_0}}_{s_0}[\mathrm{TS}^{\mathcal{G}}] = \nu$ **then**
    return $\overline{\pi}_{\overline{\lambda}_0}$
  $\overline{\lambda}_0 \leftarrow \lambda^+ > \overline{\lambda}_0$
  $\overline{\pi}_{\overline{\lambda}_0} \leftarrow \overline{\pi}^*_{\alpha,\overline{\lambda}_0} \in \arg\min_{\overline{\pi}} \mathbb{L}_{\alpha,T}(\overline{\pi}, \overline{\lambda}_0)$
$E_0 \leftarrow \frac{\epsilon}{(\overline{\lambda}_0 - \underline{\lambda}_0)}$
$n \leftarrow 0$
while $\log_2\left(2E_n \frac{(\overline{\lambda}_0 - \underline{\lambda}_0)}{\epsilon}\right) > n$ **do**
  $n \leftarrow n+1$
  $\lambda_n \leftarrow \frac{\overline{\lambda}_{n-1} + \underline{\lambda}_{n-1}}{2}$
  $\overline{\pi}_{\lambda_n} \leftarrow \overline{\pi}^*_{\alpha,\lambda_n} \in \arg\min_{\overline{\pi}} \mathbb{L}_{\alpha,T}(\overline{\pi}, \lambda_n)$
  case $\mathbb{E}^{\overline{\pi}_{\lambda_n}}_{s_0}[\mathrm{TS}^{\mathcal{G}}] = \nu$ **do**
    return $\overline{\pi}_{\lambda_n}$
  case $\mathbb{E}^{\overline{\pi}_{\lambda_n}}_{s_0}[\mathrm{TS}^{\mathcal{G}}] > \nu$ **do**
    $(\underline{\lambda}_n, \overline{\lambda}_n) \leftarrow (\lambda_n, \overline{\lambda}_{n-1})$
    $(\overline{\pi}_{\underline{\lambda}_n}, \overline{\pi}_{\overline{\lambda}_n}) \leftarrow (\overline{\pi}_{\lambda_n}, \overline{\pi}_{\overline{\lambda}_{n-1}})$
  case $\mathbb{E}^{\overline{\pi}_{\lambda_n}}_{s_0}[\mathrm{TS}^{\mathcal{G}}] < \nu$ **do**
    $(\underline{\lambda}_n, \overline{\lambda}_n) \leftarrow (\underline{\lambda}_{n-1}, \lambda_n)$
    $(\overline{\pi}_{\underline{\lambda}_n}, \overline{\pi}_{\overline{\lambda}_n}) \leftarrow (\overline{\pi}_{\underline{\lambda}_{n-1}}, \overline{\pi}_{\lambda_n})$
  $E_n \leftarrow \frac{\left(\mathbb{E}^{\overline{\pi}_{\underline{\lambda}_n}}_{s_0}[\mathrm{TS}^{\mathcal{G}}] - \nu\right)\left(\nu - \mathbb{E}^{\overline{\pi}_{\overline{\lambda}_n}}_{s_0}[\mathrm{TS}^{\mathcal{G}}]\right)}{\mathbb{E}^{\overline{\pi}_{\underline{\lambda}_n}}_{s_0}[\mathrm{TS}^{\mathcal{G}}] - \mathbb{E}^{\overline{\pi}_{\overline{\lambda}_n}}_{s_0}[\mathrm{TS}^{\mathcal{G}}]}$
$\left(\sigma^*_{\alpha,n}(\overline{\pi}_{\underline{\lambda}_n}), \sigma^*_{\alpha,n}(\overline{\pi}_{\overline{\lambda}_n})\right) \leftarrow \left(\frac{E_n}{\mathbb{E}^{\overline{\pi}_{\underline{\lambda}_n}}_{s_0}[\mathrm{TS}^{\mathcal{G}}] - \nu}, \frac{E_n}{\nu - \mathbb{E}^{\overline{\pi}_{\overline{\lambda}_n}}_{s_0}[\mathrm{TS}^{\mathcal{G}}]}\right)$
return $\sigma^*_{\alpha,n}$

Algorithm 1: Algorithmic approach to compute the mixed strategy $\sigma^*_{\alpha,n}$. We initialize with the problem under $\alpha \in \{0,1\}$, a depth T of the unfolding $\mathcal{M}_T$ (which is implicit here), an initial dual variable λ^+, an accuracy level wanted ϵ for stopping criteria in terms of convergence, and a pure strategy π_γ optimal for the SSP problem.

5.4.3 *FROM MIXED TO RANDOMIZED SOLUTION*

To finish this Chapter 5, we define here the randomized counterpart of the mixed strategy $\sigma^\star_\alpha$ found above. This can be done by using the Kuhn's Theorem 2.6.11 mentioned in the Section 2.6.3.

In our context, the unfolded-MDP can be seen as a game in extensive form with two players: the scheduling controller and the "Nature". The latter is a player who has no strategic interests in the outcome (or paths) in the system and plays randomly. Thus, the probability transition between state can be seen as a fixed randomized strategy of the Nature. Now, noticing that the strategy $\sigma^\star_\alpha$ is a convex combination of the pure strategies π'_α and π''_α, let x_t a fixed state in the unfolded-MDP. An equivalent randomized strategy for $\sigma^\star_\alpha$ can be defined for each action a_t available at x_t, from eq. (2.26), as:

$$\delta(x_t)(a_t) \quad = \quad \sigma^\star_\alpha(\pi'_\alpha)\,\mathbb{1}_{\pi'_\alpha(x_t)}(a_t) \;+\; \sigma^\star_\alpha(\pi''_\alpha)\,\mathbb{1}_{\pi''_\alpha(x_t)}(a_t)\;,$$

where $\mathbb{1}_{\pi(\omega_t)}(a_t) = 1$ if $\pi(\omega_t) = a_t$, and $\mathbb{1}_{\pi(\omega_t)}(a_t) = 0$ otherwise. Of course, Kuhn's theorem is more general since a randomized strategy can be defined over sets of information. It is straightforward to show that this strategy satisfies the requirements of the optimization problems defined in Chapter 4.

5.5 DISCUSSION

In this chapter we have investigated the existence and synthesis of a solution for an optimization problem under mixed strategies. Particularly, in an optimal mixed strategy solving a problem that comes from Chapter 4, in which its objective function is a probability and the constraint is an expectation upper bounded by a fixed threshold. Obtaining a mixed strategy here, it allows us to define a randomized strategy solution for the general multi-constrained problem of Chapter 4. An optimal mixed strategy is obtained by solving a dual optimization problem. This is built by combining up to two pure strategies, namely: one which satisfies the constraint in the expectation and another that does not. This mixed strategy gives the equality in the constraint in expectation (with the upper bound threshold). Since we focus in a strategy with the strict inequality in the constraint, we perturb such a mixed strategy to have another one that effectively satisfies the strict inequality and to be close to the optimum. With this strategy in mind, we use a theorem from game theory (namely, the Kuhn's theorem, that can be applied in this context), to define a randomized strategy solution of the initial problem.

Part III

DECENTRALIZED MODELING

This part presents the decentralized approach of the power consumption scheduling problem. The scheduling strategies for consumers are controlled and built by several decision-makers, each of them representing a consumer.

6

DECENTRALIZED POWER CONSUMPTION SCHEDULING PROBLEM

Abstract:

In this chapter, we focus on the decentralized modeling of the power consumption scheduling problem studied in Chapter 3. The consumption strategies are computed in this chapter by several decision-makers, each of them representing a consumer. The key idea to build the strategies is to use a technique so-called the sequential best response dynamics for each approach proposed in Chapter 3. At each iteration, a decision-maker updates its own strategy based on some information about the total consumption that is based on a deterministic or stochastic forecast of the noncontrollable part and the fixed consumption strategies of the others. To conduct the corresponding analysis, a common cost function has to be minimized, which represents the impact of the consumption operations on the distribution network.

LIST OF ABBREVIATIONS AND SYMBOLS

DN	Distribution Network	HS	Hot-Spot (temperature)
EVs	Electric Vehicle(s)	PaC	Plug-and-Charge
SNR	Signal-to-Noise Ratio	AAF	Accelerated Aging Factor
MDP	Markov Decision Process	SSP	Stochastic Shortest Path
ECP	Electrical Consumption Payment	PoD	Price of Decentralization
		$°C$	Degree Celsius
min	minutes	h	hour
V	Volt	kV	Kilovolt
kW	Kilowatt	kWh	Kilowatt-hour
kVA	Kilovolt-Ampere	km	Kilometres
dB	decibel	¢	cents of Australian dollar

$\mathcal{A}_i$	finite action space in $\mathcal{M}_i$
C	cost function of the power consumption problem
$C_{x_1}^{\boldsymbol{\pi}}$	cost function under a fixed $\boldsymbol{\pi}$ and x_1
C_i	cost function, instantaneous energy of consumer i
$\boldsymbol{C}_i$	cost function between transitions in $\mathcal{M}_{i,T}$
Δ_t	time-step duration of each time-slot t
$\mathcal{D}(\mathcal{S})$	set of probability distributions over a finite set $\mathcal{S}$
ε_ℓ	risk in probability of exceeding $\ell^{\max}$
ε_x	risk in probability of exceeding $x^{\max}$
e_i	energy demand of consumer i
$\mathcal{E}_i$	aggregate energy space of e_i
$\mathbb{E}_{x_1}^{\boldsymbol{\pi}_i}$	expectation operator under a fixed $\boldsymbol{\pi}_i$ and x_1
f	evolution law of the system state
$\mathcal{G}_i$	set of goal states in an MDP model of i
i	consumer
I	number of consumers
$\mathcal{I}$	set of consumers
$\mathrm{id}_{\mathcal{A}}$	identity function on a set $\mathcal{A}$
$\ell_i^{\min}$	minimum power of consumer i
$\ell_i^{\max}$	maximal power of consumer i
$\ell_{i,t}$	controllable load of consumer i at t
$L_{i,t}$	function representing $\ell_{i,t}$
$\boldsymbol{\ell}_i$	controllable load profile of i of length T
$\boldsymbol{L}_i$	function representing $\boldsymbol{\ell}_i$
$\ell_{0,t}$	real noncontrollable load at t
$\boldsymbol{\ell}_0$	real noncontrollable load profile of length T

$\widetilde{\ell}_{0,t}$	deterministic forecast of $\ell_{0,t}$ at t
$\widetilde{\boldsymbol{\ell}}_0$	deterministic forecast profile of length T
$\widetilde{L}_{0,t}$	stochastic forecast of $\ell_{0,t}$ at t
$\widetilde{\boldsymbol{L}}_0$	stochastic forecast profile of length T
$\ell^{\max}$	maximal power of the DN-transformer
ℓ_t	total load consumption at t
$\widetilde{\ell}_{-i,t}$	total load consumption except i under $\widetilde{\ell}_{0,t}$ at t
$\widetilde{L}_{-i,t}$	function representing $\widetilde{\ell}_{-i,t}$ at t
$\widetilde{\boldsymbol{L}}_{-i}$	profile of length T where each element is $\widetilde{L}_{-i,t}$
$\widetilde{L}_t$	function representing the total load under $\widetilde{\boldsymbol{L}}_0$ at t
$\mathcal{L}_i$	total load consumption space from the viewpoint of i
m	round or iteration of the BRD
$\mathcal{M}_i$	MDP model of i
$\mathcal{M}_{i,T}$	unfolding of $\mathcal{M}_i$ with depth T
$\mathcal{N}(\mu, \sigma^2)$	Gaussian (normal) distribution with mean μ and variance σ^2
$\omega_{i,t}$	history or path of length t of the system available for i
P_i	transition probability between states in $\mathcal{M}_i$
$P_{i,T}$	transition probability between states in $\mathcal{M}_{i,T}$
P_0	probability distribution of $\widetilde{L}_{0,t}$
P_{-i}	probability distribution of $\widetilde{L}_{-i,t}$
proj_j	projection function on the j-component of a sequence
$\boldsymbol{\pi}$	scheduling strategy vector of length I
$\boldsymbol{\pi}_{-i}$	scheduling strategy vector except the strategy of i
$\boldsymbol{\pi}_i$	scheduling strategy profile of i
$\boldsymbol{\pi}_i^{\mathrm{RP}}$	a $\boldsymbol{\pi}_i$ built by rectangular profile method
$\boldsymbol{\pi}_i^{\mathrm{DC}}$	a $\boldsymbol{\pi}_i$ built by dynamic charging method
$\boldsymbol{\pi}_i^{\mathrm{VF}}$	a $\boldsymbol{\pi}_i$ built by valley-filling method
s_t	state in $\mathcal{S}_{i,T}$ at t
S_t	state process in $\mathcal{M}_{i,T}$ representing s_t
$\mathcal{S}_{i,T}$	augmented space of states in $\mathcal{M}_{i,T}$
t	time-slot
t_i^{start}	time at which the consumption starts for i
t_i^{stop}	time at which the consumption ends for i
T	finite horizon time
$\mathcal{T}$	set of time-slots
$x^{\max}$	upper bound of the system state values
x_t	system state at t
$\widetilde{x}_t$	system state under $\widetilde{\ell}_{0,t}$ at t
$\widetilde{X}_t$	function representing the system state under $\widetilde{\boldsymbol{L}}_0$ at t
$\mathcal{X}$	finite set of system states

6.1 MOTIVATION AND CONTRIBUTIONS

Motivated by Chapter 3, we aim to solve the power consumption scheduling problem in a decentralized way. Such a problem consists in a scenario in which several controllable consumption entities (also called controllable electric devices or simple consumers here) have a certain energy demand and want to have this demand to be fulfilled before a set deadline. The simple instance of such a scenario used for numerical purposes in this chapter, is the case of a pool of Electric Vehicles (EVs) which have to recharge their battery to a given state of charge within a given time window set by the EV owner. Thus, each consumer has to choose at any time-step the consumption power so that the accumulated energy reaches a desired level.

While in Chapter 3 the power consumption scheduling strategies of consumers were assumed to be centralized (i.e., a single entity has controlled and built their strategies), we assume here that the operations must be decentralized. First, each consumer is free to make its own decision in terms of choosing its consumption power. Second, it has to be decentralized information-wise, i.e., the scheduling algorithm or procedure (when it is implemented by a machine, which is the most common scenario), only relies on local or scalable information. In this way, each consumer is treated as a decision-maker in this chapter, and each one needs a certain information about the total consumption, which is based on a deterministic or stochastic forecast of the noncontrollable part and the fixed consumption of the others. A typical scenario considered here is that a day-head decision has to be made and some knowledge (imperfect deterministic/stochastic forecast on the noncontrollable loads) is available to schedule. To design appropriate decentralized strategies in each approach, the key idea is to use an iterative technique so-called sequential Best Response Dynamics (BRD), which can be seen as a generalization of well-known iterative techniques (see [79] for example). This technique acts as a coordination mechanism among the decision-makers to ensure that the scheduled strategies are consistent with the problem.

6.1.1 *STRUCTURE*

The main contributions of this chapter explain our work reported in [M4] and in a forthcoming paper [M5]. This and the structure of this chapter can be summarized as follows.

(i) In Section 6.2, we recapitulate the general mathematical formulation of the power consumption scheduling problem studied in the Chapter 3, but in a decentralized point of view. The objective is to minimize the impact of total load consumption on the Distribution Network (DN), where the scheduling strategies of consumers are chosen by several decision-makers (each of them representing a consumer). Such an impact is taken into account by an objec-

tive function of interest that can be deterministic or stochastic depending of the noncontrollable load consumption forecast.

(ii) In Section 6.3, a deterministic forecast of the noncontrollable consumption is assumed for the consumption scheduling procedure. In this case, the strategies of consumers are reduced to be parameter vectors in which each component is a consumption power. Unlike the Section 3.3, the three sorts of strategies are built here based on sequential BRD. These can be summarized as follows.

First, the consumption strategies are imposed to be non-interruptible and correspond to rectangular windows. Therefore, each strategy boils down to choosing the time instant at which the consumption operation should start.

Second, each power consumption does not need to be binary anymore but can be take arbitrary values. In this case, the dynamic structure of the scheduling problem is taken into account explicitly and the state of the system can be thus controlled. For instance, it is possible to satisfy constraints over the system state that is not very suitable for the previous method (it is less flexible since the consumption is uninterrupted).

Third, a valley-filling algorithm replaces the scheduling problem under consideration (see [106] for example), which is based only on the minimization of the total consumption and not on any other measure of impact over the DN.

(iii) In Section 6.4, the key idea is to take into account forecast errors. For this, a stochastic forecast of the noncontrollable consumption is accounted for, leading to the use of Markov decision processes (see [96] for example) to schedule the controllable loads. The resulting strategies can therefore be adapted to different noncontrollable consumption events, i.e., more robust to forecast noises.

(iv) Numerical results are shown in Section 6.5, where the objective function of the scheduling problem models the degradation of the DN. First, in terms of the DN-transformer lifetime, as it is shown in Section 6.5.1, and second, in terms of the total electricity consumption payment, as it is shown in Section 6.5.2.

6.2 PROBLEM FORMULATION

We give here a brief summary of the formulation of the problem presented in Section 3.2, but turning to the decentralized case. The formulation of the problem can be explained as follows.

LOAD CONSUMPTION

We start by considering a Distribution Network (DN) comprising one transformer (referred as DN-transformer here), in which two groups of electric devices are connected: a set $\mathcal{I} = \{1, ..., I\}$, $I \in \mathbb{N}$, of controllable electric devices (also called consumers), e.g., Electric Vehicles (EVs), dishwashers, water-heaters, etc.; and a set of other electric devices that are assumed to induce a load consumption which is independent of the controllable electric devices and therefore referred to as the noncontrollable load consumption (e.g., heating, lighting, cooking, etc.). Time is assumed to be slotted and indexed by $t \in \mathcal{T} = \{1, ..., T\}$, $T \in \mathbb{N}$. For example, if the whole time window under consideration is from 5 pm to 8 am (of the next day), there are thirty time-slots ($T = 30$) whose duration is 30 min, on which a consumer may be active or not. The extent to which a consumer $i \in \mathcal{I}$ is active on time-slot $t \in \mathcal{T}$, it is measured by the **controllable load** consumption that it generates, which is denoted by $\ell_{i,t} \in \mathbb{R}_0^+$. The corresponding **noncontrollable load** consumption is denoted by $\ell_{0,t} \in \mathbb{R}_0^+$ and its profile over time set $\mathcal{T}$ is expressed by:

$$\boldsymbol{\ell}_0 \quad = \quad (\ell_{0,t})_{t \in \mathcal{T}} \ . \tag{3.1}$$

The **total load** consumption generated on the DN-transformer at time t can be then expressed by:

$$\ell_t \quad = \quad \ell_{0,t} + \sum_{i \in \mathcal{I}} \ell_{i,t} \ . \tag{3.3}$$

A natural constraint on the DN-transformer is due to its maximal admissible power, e.g., the maximal power of a typical DN-transformer in a urban district is 90 kW. The total load consumption has to be then lower than the **maximal power** denoted here by $\ell^{\max} \in \mathbb{R}^+$, i.e.,

$$\ell_t \quad \leq \quad \ell^{\max} \ . \tag{3.4}$$

The main difference with Chapter 3 is that here, there is not a centralized scheduler choosing the power loads of consumers. In this chapter, each consumer $i \in \mathcal{I}$ is understood as an individual decision-maker who updates its own **controllable load profile**:

$$\boldsymbol{\ell}_i \quad := \quad (\ell_{i,t})_{t \in \mathcal{T}} \ , \tag{6.1}$$

so that its **energy demand**, denoted by $e_i \in \mathbb{R}^+$ (e.g., it is typically $e_i = 24$ kWh for an EV), to complete a corresponding task has to be satisfied before time-horizon $t = T$. Mathematically, it is required that:

$$\Delta_t \sum_{t \in \mathcal{T}} \ell_{i,t} \;\geq\; e_i \;, \tag{3.5}$$

where Δ_t is the duration of a time-slot, e.g., if each t represents $30\,\mathrm{min}$, then $\Delta_t = 0.5\,\mathrm{h}$. In addition, it is required that the power load of each consumer i is at least a **minimum power**, denoted by $\ell_i^{\min} \in \mathbb{R}_0^+$, and that it does not exceed a **maximal power**, denoted by $\ell_i^{\max} \in \mathbb{R}^+$, at which i can be consuming, i.e.,

The minimum power refers when, e.g., i is switched off but is designed to draw some load in standby mode.

$$\ell_i^{\min} \;\leq\; \ell_{i,t} \;\leq\; \ell_i^{\max} \,. \tag{3.6}$$

Figure 6.1 illustrates a typical scenario that is encompassed by the considered model. This figure represents a set of consumers (each of them is represented by an EV) which are connected to a DN-transformer. The household is understood as the noncontrollable part of the total load consumption. Each EV is a decision-maker who chooses separately its own controllable load profile, see eq. (6.1), i.e., in a decentralized manner. For instance, four sorts of consumption (charging models) of an EV i are represented in Figure 3.2, namely: rectangular, continuous and discrete charging. In the first case, the controllable load $\ell_{i,t}$ can only take two values: either $\ell_{i,t} = \ell_i^{\min}$ or $\ell_{i,t} = \ell_i^{\max}$, and also when $\ell_{i,t} = \ell_i^{\max}$, the consumption operation is uninterrupted. This method is assumed in Section 6.3.1. In the second case, $\ell_{i,t}$ takes any arbitrary value between $\ell_i^{\min}$ and $\ell_i^{\max}$, see eq. (3.6). This method of consumption is assumed in Section 6.3.2 and Section 6.3.3. The last case is a discretization of the type of previous consumption, which is assumed in Section 6.4.1.

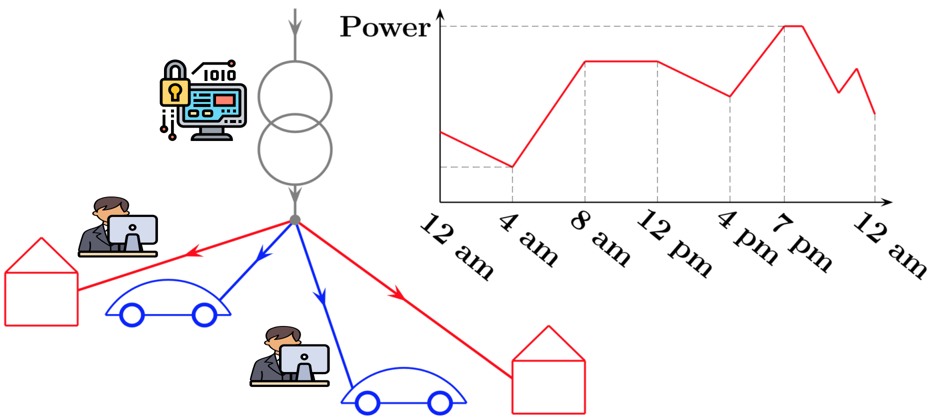


Figure 6.1: A typical scenario which is captured by the model analyzed in this chapter. Each EV controls its charging power profile to reach its demanded state of charge, for instance for its next trip. The charging consumption is based on some coordination mechanism among the EVs and some knowledge about the day-ahead (aggregated) noncontrollable part of the total consumption.

SYSTEM STATE

For the system of interest (namely, the DN), the state is denoted by x_t (e.g., this can represent the total electricity bill, the DN-transformer temperature, etc.). In this chapter, it is required to remain the state of the system upper bounded by a given threshold, which is by $x^{\max} \in \mathbb{R}^+$, i.e.,

$$x_t \quad \leq \quad x^{\max} \; . \tag{3.7}$$

A general (nonlinear for instance) model assumed in this chapter for the **evolution law of the system state** is expressed by:

$$x_{t+1} \quad = \quad f(x_t,\, \ell_{1,t}, ..., \, \ell_{I,t},\, \ell_{0,t}) \tag{6.2}$$

for each $t = 1, ..., T$; where $x_1 \in \mathbb{R}^+$ is a given initial condition of the system state. The main difference with the evolution law of eq. (3.8), is that here we express explicitly the dependence of f over the controllable loads of consumers, which are chosen separately by them. A practical assumption in this chapter to make an effective calculation of controllable loads and to ensure convergence of algorithms, is the Assumption 2.5.1, which states that f is a function depending of the total load consumption (3.3).

SCHEDULING STRATEGIES

The way in which each controllable load profile of each consumer is chosen, see eq. (6.1), it is according to a function called power consumption scheduling strategy (also called decision rule). As we said earlier, in this chapter there are I decision-makers, which control their own load profile (6.1) and then, the scheduling strategies are built separately by the individual decision-makers over time. Once the scheduling strategies are fixed, a cost is incurred at each time t and we can therefore compare the effectiveness of each strategy. The information that each decision maker $i \in \mathcal{I}$ takes into account to schedule a power consumption strategy is the history (also called path) of the system, represented by the visited states and the controllable loads chosen previously. More precisely, assume the following composed **history** or **path** of the system up to t is available for the decision-maker i:

$$\omega_{i,t} \quad := \quad (x_1,\, \ell_{i,1},\, x_2,\, \ell_{i,2}, \ldots,\, x_{t-1},\, \ell_{i,t-1},\, x_t) \; . \tag{6.3}$$

A **scheduling strategy** to select the controllable loads of consumer i at time t is defined by the following function:

$$\pi_{i,t}(\omega_{i,t}) \quad = \quad \ell_{i,t} \; . \tag{6.4}$$

We denote by $\boldsymbol{\pi}_i := (\pi_{i,t})_{t\in\mathcal{T}}$ the profile of the latter. Note that this definition is according to the one of pure strategies seen in Section 2.6.3. Recall that, in this context, $\boldsymbol{\pi}_i$ is **memoryless** or **Markov** if for each $t \in \mathcal{T}$,

$$\pi_{i,t}(\omega_{i,t}) \quad = \quad \pi_{i,t}(\omega'_{i,t'}) \; ,$$

where $\omega_{i,t} = (x_1, \ell_{i,1}, ..., \ell_{i,t-1}, x_t)$ and $\omega'_{i,t} = (x_1, \ell'_{i,1}, ..., \ell'_{i,t-1}, x'_t)$ are any paths of the system such that $x_t = x'_t$, that is to say that the only relevant information that a Markov strategy needs is contained in the current state of the system.

In this chapter, we are interested in a decentralized version of the (centralized) scheduling problem (3.11) defined in Chapter 3, wherein the strategies are scheduled by a single entity, also-called centralized operator. Here, this is not assumed anymore. Before introducing the decentralized problem in a deterministic or stochastic context and to show how the strategies are built, we introduce a general point of view to understand the decentralization among the decision-makers.

Let $\boldsymbol{\pi} = (\boldsymbol{\pi}_i)_{i\in\mathcal{I}}$ a scheduling strategy vector. For any $i \in \mathcal{I}$, we can represent such a vector as $\boldsymbol{\pi} = (\boldsymbol{\pi}_i, \boldsymbol{\pi}_{-i})$, where $\boldsymbol{\pi}_{-i}$ denotes the i-reduced strategy vector, i.e., the vector of all strategies except the one of i:

$$\boldsymbol{\pi}_{-i} \quad = \quad (\boldsymbol{\pi}_1, ..., \boldsymbol{\pi}_{i-1}, \boldsymbol{\pi}_{i+1}, ..., \boldsymbol{\pi}_I) \; . \tag{6.5}$$

To schedule a strategy $\boldsymbol{\pi}_i$ under information (6.3), the decision-maker i needs to know the system state values, which depend on all controllable loads of consumers and the noncontrollable part of the total load consumption, see eq. (6.2). First, suppose that $\boldsymbol{\pi}_{-i}$ is in a certain way fixed. To represent $\boldsymbol{\pi}$ under such an assumption, we write:

$$\boldsymbol{\pi} \quad = \quad (\boldsymbol{\pi}_i \mid \boldsymbol{\pi}_{-i}) \; . \tag{6.6}$$

That is, we refer to what we are looking to schedule $\boldsymbol{\pi}_i$ under the fixed strategies of the other decision-makers, to know effectively the value of the system state. In this chapter, we aim to build strategies without revealing the strategies of the decision-makers. Before going into this detail, we introduce the problem under consideration in the following.

SCHEDULING PROBLEM

Continuing with the general representation of the scheduling strategy vector (6.6), a decision-maker $i \in \mathcal{I}$ selects its own power consumption scheduling strategy from the following arg min set:

Decentralized Power Consumption Scheduling Problem

$$\begin{aligned} \underset{\boldsymbol{\pi}_i}{\arg\min} \quad & \sum_{t\in\mathcal{T}} C^{\boldsymbol{\pi}}_{x_1}(x_t, \ell_{1,t}, ..., \ell_{i,t}, ..., \ell_{I,t}, \ell_{0,t}) \\ \text{s.t.} \quad & (3.3),\ (3.4),\ (3.5),\ (3.6),\ (3.7),\ (6.2),\ (6.6) \; . \end{aligned} \tag{6.7}$$

So, we are looking at solving such a problem for each $i \in \mathcal{I}$, but at first glance, it needs some coordination mechanism through the decision-makers to ensure that the scheduling strategies are consistent with the problem, in particular with the system states values, see for instance [75]. To handle this problem and make an effective calculation of the strategies, a forecast on the real noncontrollable load consumption (3.1) is assumed to be available, which can be deterministic or stochastic, see Section 2.4 for details. In this way, the strategies are scheduled in an **offline** way, defining the controllable loads that are executed on the time under consideration. Once the strategies are determined, these can be effectively run **online**. Also, as we have said before, the evolution law of the system state in eq. (6.2), is assumed to be dependent of the total load consumption (3.3), see Assumption 2.5.1. Thus, based the latter and an available forecast, this allows naturally to build a way to coordinate information among the decision-makers to determine their "best response behaviors" by considering their "opponents" are part of the noncontrollable load consumption. This will be explained a little further in the following. Before that, we present in the next section, the scheduling problem (6.7) under a deterministic forecast together with a general version of the "best responses" of the decision-makers.

6.3 SOLUTION METHODOLOGY IN THE DETERMINISTIC CASE

In this section, a deterministic forecast of the real noncontrollable load consumption (3.1) is assumed to schedule the controllable loads. Following the description in Section 2.4.1, we represent a **deterministic forecast** by:

$$\widetilde{\boldsymbol{\ell}}_0 \quad = \quad (\widetilde{\ell}_{0,t})_{t\in\mathcal{T}} \ . \tag{3.12}$$

In such a case, the decentralized problem of power consumption scheduling to solve for each $i \in \mathcal{I}$, can be rewritten as:

Decentralized Deterministic Problem

$$\begin{aligned} \arg\min_{\boldsymbol{\pi}_i} \quad & \sum_{t\in\mathcal{T}} C^{\boldsymbol{\pi}}_{x_1}(\widetilde{x}_t,\, \ell_{1,t}, ...,\, \ell_{i,t}, ...,\, \ell_{I,t}\, ;\, \widetilde{\ell}_{0,t}) && (6.8)\\ \text{s.t.} \quad & (3.3),\ (3.4),\ (3.5),\ (3.6),\ (3.7),\ (3.12),\ (6.2),\ (6.6)\ . \end{aligned}$$

Here, each noncontrollable load consumption is a simple parameter, i.e., a value of the deterministic forecast. Note that the system state and the total load consumption are also affected by the forecast, because these depend on the noncontrollable loads, see resp. eq. (6.2) and eq. (3.3). Thus, choosing the controllable loads, the system state is completely determined under the dynamic (6.2). A scheduling strategy for each consumer can therefore be built under the deterministic information of the system state and based on some coordination mechanism among the decision-makers. The algorithms proposed in this chapter, are all based on an iterative technique from game theory [79, 84]. First, we will introduce a general version of such a technique under the context of the decentralized deterministic problem (6.8). Second, we show the type of information to share iteratively among the decision-makers, without revealing the information about the strategies.

THE SEQUENTIAL BEST RESPONSE DYNAMICS

Here, we provide three sorts of strategies, all based on the Best Response Dynamics (BRD) algorithm, which can be seen as a generalization of well-known iterative techniques, often resulting in local optima, see for instance [33, 79, 81]. In its most used form, the BRD operates sequentially such that the consumers update their strategies in a round-robin manner. Within round $m \in \mathbb{N}$, a scheduling strategy selected by decision-maker $i \in \mathcal{I}$, is built by choosing:

Deterministic Problem - General Best Response Dynamics

$$\boldsymbol{\pi}_i^{(m)} \in \arg\min_{\boldsymbol{\pi}_i} \sum_{t\in\mathcal{T}} C_{x_1}^{\boldsymbol{\pi}^{(m)}}(\widetilde{x}_t, \ell_{1,t}, \dots, \ell_{i,t}, \dots, \ell_{I,t}\,;\, \widetilde{\ell}_{0,t}) \tag{6.9}$$
$$\text{s.t.} \quad (3.3), (3.4), (3.5), (3.6), (3.7), (3.12), (6.2), (6.6)\,.$$

Note that the cost function in the latter problem follows by the strategy vector $\boldsymbol{\pi}$ at iteration m to schedule the strategy of i, i.e., selecting $\boldsymbol{\pi}_i$ above, $\boldsymbol{\pi}$ from eq. (6.6) is written in this context as:

$$\boldsymbol{\pi}^{(m)} \quad = \quad \left(\boldsymbol{\pi}_i \mid \boldsymbol{\pi}_{-i}^{(m)}\right)\,,$$

where as before, $\boldsymbol{\pi}_{-i}^{(m)}$ represents the i-reduced strategy vector, see eq. (6.5), but here it is constructed sequentially. At iteration m, this is of the form:

$$\boldsymbol{\pi}_{-i}^{(m)} \quad = \quad \left(\boldsymbol{\pi}_1^{(m)}, \dots, \boldsymbol{\pi}_{i-1}^{(m)}, \boldsymbol{\pi}_{i+1}^{(m-1)}, \dots, \boldsymbol{\pi}_I^{(m-1)}\right)\,.$$

The complete procedure is translated in the pseudo-code of Algorithm 2.

To update the strategies m times from the Algorithm 2, $m\,I$ iterations are required. The order in which decision-makers update their strategies does not matter to obtain convergence, see for instance [23]. A simple rule seems to be to update the strategies at each iteration in a sequential order (line 8-line 11). On the other hand, if the $\arg\min$ set is not a singleton (line 9), the scheduling strategy can randomly draw among the minimum points without affecting the performance. A variation of Algorithm 2 can be obtained by updating the scheduling strategies simultaneously. The main reason why we have not considered the parallel version is that it is known that there is no general analytical result for guaranteeing convergence [79]. When converging, the parallel implementation is faster but since the strategies are computed offline here, convergence time may be seen as a secondary feature. The way in which this algorithm is used, is according to some information exchanged between the decision-makers in each round, as it is shown below.

Input : $\epsilon > 0$, x_1, T, I, $\widetilde{\boldsymbol{\ell}}_0$, $\boldsymbol{\pi}^{\text{init}}$

Output: Solution of the problem (6.8), built by BRD

$\boldsymbol{\pi}^{(0)} \leftarrow \boldsymbol{\pi}^{\text{init}}$

$C_0 \leftarrow \sum_{t\in\mathcal{T}} C_{x_1}^{\boldsymbol{\pi}^{(0)}}(\widetilde{x}_t,\, \ell_{1,t}, ..., \ell_{i,t}, ..., \ell_{I,t}\,;\, \widetilde{\ell}_{0,t})$

$m \leftarrow 0$

while $\| C_m - C_{m-1} \|^2 > \epsilon$ **or** $m = 0$ **do**

  $m \leftarrow m+1$

  $\boldsymbol{\pi}^{(m)} \leftarrow \left(\boldsymbol{\pi}_1^{(m-1)} \mid \boldsymbol{\pi}_{-1}^{(m-1)}\right)$

  $i \leftarrow 1$

  while $i \leq I$ **do**

    $\boldsymbol{\pi}_i^{(m)} \in \begin{cases} \arg\min_{\boldsymbol{\pi}_i} \sum_{t\in\mathcal{T}} C_{x_1}^{\boldsymbol{\pi}^{(m)}}(\widetilde{x}_t, \ell_{1,t}, ..., \ell_{i,t}, ..., \ell_{I,t}; \widetilde{\ell}_{0,t}) \\ \quad \text{s.t. } (3.3), (3.4), (3.5), (3.6), (3.7), (6.2) \end{cases}$

    $i \leftarrow i+1$

    $\boldsymbol{\pi}^{(m)} \leftarrow \left(\boldsymbol{\pi}_i^{(m)} \mid \boldsymbol{\pi}_{-i}^{(m)}\right)$

  $C_m \leftarrow \sum_{t\in\mathcal{T}} C_{x_1}^{\boldsymbol{\pi}^{(m)}}(\widetilde{x}_t,\, \ell_{1,t}, ..., \ell_{i,t}, ..., \ell_{I,t}\,;\, \widetilde{\ell}_{0,t})$

return $\boldsymbol{\pi}^{(m)}$

Algorithm 2: The decentralized power consumption scheduling algorithm based on BRD. Here, we initialize with a strategy $\boldsymbol{\pi}^{\text{init}}$, an initial condition x_1 of the system state, an accuracy level wanted ϵ for stopping criteria in terms of convergence, a finite horizon time T, a number of consumers I, and a deterministic forecast of the noncontrollable consumption $\widetilde{\boldsymbol{\ell}}_0$.

SEQUENTIAL INFORMATION BASED ON THE DETERMINISTIC TOTAL LOAD CONSUMPTION

We focus here on how to share some information among the decision-makers, without revealing their strategies when the BRD algorithm is implemented. Suppose that for $i \in \mathcal{I}$ and iteration $m \in \mathbb{N}$,

$$\ell_{i,t}^{(m)} = \pi_{i,t}^{(m)}(\omega_{i,t}) \tag{6.10}$$

denotes the controllable load of consumer i at time t, computed by its scheduling strategy at iteration m. Suppose now that the decision-makers $j = 1, ..., i-1$ have update their strategies (iteration m) and all the others $j = i, ..., I$ not yet at iteration $m-1$ (if $m = 1$, the iteration zero for the latter decision-makers represents an initialized scheduling strategy before to apply the BRD). The total load consumption (3.3) due to the noncontrollable part (here, the forecast (3.12) to schedule strategies) and the controllable part (6.10) of consumers, can be written as:

$$\widetilde{\ell}_{i,t}^{(m)} := \widetilde{\ell}_{0,t} + \sum_{j=1}^{i-1} \ell_{j,t}^{(m)} + \sum_{j=i}^{I} \ell_{j,t}^{(m-1)}\,, \tag{6.11}$$

and then, under the same context, the constraint (3.4) is written as:

$$\widetilde{\ell}_{i,t}^{(m)} \quad \leq \quad \ell^{\max} \,. \tag{6.12}$$

The idea is then to use the latter for each decision-maker as a "non-controllable" load consumption at each iteration. More precisely, each controllable load of i at iteration m, is computed "in response" to the load consumption of his "opponents":

$$\widetilde{\ell}_{-i,t}^{(m)} \quad := \quad \widetilde{\ell}_{i,t}^{(m)} \; - \; \ell_{i,t}^{(m-1)} \,. \tag{6.13}$$

Therefore, the only sufficient information that the decision-maker i receives at m, is the profile:

$$\left(\widetilde{\ell}_{i,t}^{(m)}\right)_{t\in\mathcal{T}} \,. \tag{6.14}$$

Based on all above, a scheduling strategy selected by the decision-maker $i \in \mathcal{I}$ within round $m \in \mathbb{N}$, can be built by solving (6.9) now as the following problem:

Deterministic Problem - Best Response Dynamics

$$\boldsymbol{\pi}_i^{(m)} \in \arg\min_{\boldsymbol{\pi}_i} \sum_{t\in\mathcal{T}} C_{x_1}^{\boldsymbol{\pi}_i}\left(\widetilde{x}_t,\, \ell_{i,t}\,;\, \widetilde{\ell}_{-i,t}^{(m)}\right) \tag{6.15}$$
$$\text{s.t.} \quad (3.5),\, (3.6),\, (3.7),\, (3.12),\, (6.2),\, (6.11),\, (6.12),\, (6.13)\,.$$

Following (6.10), the solution of the latter problem, can also be written in this chapter as:

$$\boldsymbol{\pi}_i^{(m)} \quad = \quad \left(\ell_{i,1}^{(m)},\, \ldots,\, \ell_{i,T}^{(m)}\right) \,,$$

since the scheduling strategy of a decision-maker defines (offline) a sequence of controllable loads. Under the context of the problem (6.15), the Algorithm Algorithm 2 can be adapted as it is shown in Algorithm 3. In the following, three decentralized power consumption scheduling strategies are differentiated.

6.3.1 *ITERATIVE RECTANGULAR PROFILES*

Agreeing with Section 3.3.1, the strategies of decision-makers are imposed here to be rectangular load profiles. That is, each controllable load $\ell_{i,t}$ can take only two values according to the constraint (3.6): either $\ell_{i,t} = \ell_i^{\min}$ or $\ell_{i,t} = \ell_i^{\max}$. Also, when such a controllable load takes the value $\ell_i^{\max}$, this is imposed to be uninterrupted, i.e., i consumes $\ell_i^{\max}$ until the energy demand (3.5) is fulfilled. Recall that the motivations to use power consumption scheduling strategies of this form are [112]: first, each strategy is easy to implement since each strategy boils down to a simple decision for each scheduler i, namely: the time $t_i^{\text{start}} \in \mathcal{T}$ at which the consumption starts. Second, rectangular load profiles are believed to perform quite well in terms of noise robustness on the noncontrollable load forecast, see for instance [19, 20].

Input : $\epsilon > 0$, x_1, T, I, $\widetilde{\boldsymbol{\ell}}_0$, $\boldsymbol{\ell}^{\text{init}}$
Output : Solution of the problem (6.15)

$\widetilde{\boldsymbol{\ell}}^{(1)} \leftarrow \widetilde{\boldsymbol{\ell}}_0$
for $i \leftarrow 1$ **to** I **do**
  $\boldsymbol{\ell}_i^{(0)} \leftarrow \boldsymbol{\ell}_i^{\text{init}}$
  $\widetilde{\boldsymbol{\ell}}^{(1)} \leftarrow \widetilde{\boldsymbol{\ell}}^{(1)} + \boldsymbol{\ell}_i^{(0)}$
$C_0 \leftarrow \sum_{t \in \mathcal{T}} C_{x_1}\left(\widetilde{x}_t, \widetilde{\boldsymbol{\ell}}_t^{(1)}\right)$
$m \leftarrow 0$
while $\| C_m - C_{m-1} \|^2 > \epsilon$ **or** $m = 0$ **do**
  $m \leftarrow m+1$
  for $i \leftarrow 1$ **to** I **do**
    $\widetilde{\boldsymbol{\ell}}_{-i}^{(m)} \leftarrow \widetilde{\boldsymbol{\ell}}^{(m)} - \boldsymbol{\ell}_i^{(m-1)}$
    $\boldsymbol{\pi}_i^{(m)} \in$
    $\begin{cases} \arg\min_{\boldsymbol{\pi}_i} \sum_{t \in \mathcal{T}} C_{x_1}^{\boldsymbol{\pi}_i}\left(\widetilde{x}_t,\, \ell_{i,t}\,;\, \widetilde{\ell}_{-i,t}^{(m)}\right) \\ \quad \text{s.t. } (3.3), (3.4), (3.5), (3.6), (3.7), (6.2), (6.12) \end{cases}$
    $\boldsymbol{\ell}_i^{(m)} \leftarrow \boldsymbol{\pi}_i^{(m)}$
    $\widetilde{\boldsymbol{\ell}}^{(m)} \leftarrow \widetilde{\boldsymbol{\ell}}_{-i}^{(m)} + \boldsymbol{\ell}_i^{(m)}$
  $C_m \leftarrow \sum_{t \in \mathcal{T}} C_{x_1}\left(\widetilde{x}_t, \widetilde{\boldsymbol{\ell}}_t^{(m)}\right)$
return $\boldsymbol{\pi}^{(m)}$

Algorithm 3: The decentralized power consumption scheduling algorithm based on BRD and the shared information of the total load consumption (6.14). Here, we initialize with a strategy or profile of controllable loads $\boldsymbol{\ell}^{\text{init}}$, an initial condition x_1 of the system state, an accuracy level wanted ϵ for stopping criteria in terms of convergence, a finite horizon time T, a number of consumers I, and a deterministic forecast of the noncontrollable load consumption $\widetilde{\boldsymbol{\ell}}_0$.

Mathematically, rectangular profiles can be written for each decision-maker $i \in \mathcal{I}$ as follows:

$$\boldsymbol{\ell}_i \in \left\{ (\ell_{i,t})_{t \in \mathcal{T}} \in \{\ell_i^{\min}, \ell_i^{\max}\}^T \;\middle|\; \begin{array}{l} \forall t \in \{t_i^{\text{start}}, \ldots, t_i^{\text{stop}}\} \subseteq \mathcal{T},\ \ell_{i,t} = \ell_i^{\max} \\ \forall t \in \mathcal{T} \setminus \{t_i^{\text{start}}, \ldots, t_i^{\text{stop}}\}\ ,\ \ell_{i,t} = \ell_i^{\min} \end{array} \right\} \tag{3.14}$$

In practice, from the constraint of energy demand (3.5), each t_i^{start} is limited to being:

$$t_i^{\text{start}} \;\leq\; T - \frac{e_i}{\ell_i^{\max}}\ , \tag{3.15}$$

and t_i^{stop} can be chosen being the minimum stopping time such that:

$$(t_i^{\text{stop}} - t_i^{\text{start}})\, \ell_i^{\max} \;\geq\; e_i\ , \tag{3.16}$$

where in eq. (3.15) and eq. (3.16) we do not take into account the load due to the minimum power $\ell_i^{\min}$, since it only refers when, e.g., the consumption of i is switched off but is designed to draw some load in standby mode.

In this case, choosing an optimal scheduling strategy amounts to choosing an optimal consumption start time t_i^{start} separately for each decision-maker $i \in \mathcal{I}$, which is chosen "in response" to the consumption (6.13). This redefines the scheduling problem under the BRD (6.15) as follows:

Deterministic Problem - BRD - Rectangular Profiles

$$\boldsymbol{\pi}_i^{(m)} \in \underset{\boldsymbol{\pi}_i^{\text{RP}}}{\arg\min} \sum_{t \in \mathcal{T}} C_{x_1}^{\boldsymbol{\pi}_i^{\text{RP}}} \left(\widetilde{x}_t,\, \ell_{i,t}\,;\, \widetilde{\ell}_{-i,t}^{(m)} \right) \tag{6.16}$$

s.t. (3.7), (3.12), (3.14), (3.15), (3.16), (6.2), (6.11), (6.12), (6.13).

This problem redefines the line 11 of the BRD Algorithm 3 for the rectangular profiles context. On the other hand, since the state of the system is completely determined under the assumption of the shared information of the consumption (6.11), this problem (6.16) is reduced to a simple optimization problem to computing the individual start time of load consumptions. However, rectangular scheduling strategies of this form are not well suited in presence of saturation constraints, such as (3.7) and (6.12), e.g., when the maximal temperature or the maximal power of the DN-transformer could be reached, see for example [19][M1]. A rectangular profile is less flexible since it is uninterrupted while it is consuming. In addition, the initial strategy $\boldsymbol{\ell}^{\text{init}}$ in Algorithm 3 has to satisfy such constraints to be suitable (i.e., feasible for the problem), but it is not easy to find such an initial strategy if I and/or T are large to initialize the Algorithm 3. For example, if $\boldsymbol{\ell}^{\text{init}}$ is chosen to be a plug-and-charge profile (i.e., each i consumes $\ell_i^{\max}$ as soon as it is plugin to the grid) which is a particular case of rectangular profile, the constraint of the system state can be quickly not satisfied for I large. A suitable scheduling method that can easily integrate this constraints is shown in the next Section.

6.3.2 *ITERATIVE DYNAMICAL CHARGING*

In contrast with the previous section, each controllable load consumption does not need to be binary anymore but can be take continuous values. Thus, the consumption profile is no longer rectangular and can be arbitrary (under the constraint of the minimum and maximum power consumption). Thus, each scheduling strategy does not boil down to a single scalar anymore, i.e., the consumption start instant. The motivation for this is to have a better performance for the consumers but also to be able to control the system state. In the previous section, the dynamical system was controlled in a one-shot manner. Here, the state evolution law is taken into account explicitly and the state can be controlled. For instance, it is possible to guarantee that that the upper bound (3.7) on the system state is not violated.

Agreeing with Section 3.3.2, we show the modelling framework that allows us to integrate dynamical aspects. This analysis is conducted under the assumption of the sufficient information of the consumption (6.14) that each decision-maker $i \in \mathcal{I}$ receives at iteration $m \in \mathbb{N}$ of the BRD to schedule its strategy. Indeed, the profile (6.14) is a simple vector of parameters, which allows to know explicitely the value of the system state. Here, we also handle the dynamic law of the states (6.2). This observation allows us to convert the scheduling problem (6.15) under BRD, into a standard optimization problem [30]. Following the definition of the functions $(g_t)_{t\in\mathcal{T}}$ in eq. (3.18), we let $g_1(x_1) = x_1$ as the initial condition, and

$$\begin{aligned} & g_{t+1}\Big(x_1, \ell_{i,1}, ..., \ell_{i,t}\,;\, \widetilde{\ell}^{(m)}_{-i,1}, ..., \widetilde{\ell}^{(m)}_{-i,t}\Big) \\ & \qquad := \quad f\Big(g_t\Big(x_1, \ell_{i,1}, ..., \ell_{i,t-1}\,;\, \widetilde{\ell}^{(m)}_{-i,1}, ..., \widetilde{\ell}^{(m)}_{-i,t-1}\Big),\, \ell_{i,t};\, \widetilde{\ell}^{(m)}_{-i,t}\Big) \end{aligned} \tag{6.17}$$

So that, we will have $g_{t+1}\Big(x_1, \ell_{i,1}, ..., \ell_{i,t}\,;\, \widetilde{\ell}^{(m)}_{-i,1}, ..., \widetilde{\ell}^{(m)}_{-i,t}\Big) = \widetilde{x}_{t+1}$. In this way, the constraint (3.7) can be rewritten as:

$$g_t\Big(x_1, \ell_{i,1}, ..., \ell_{i,t-1}\,;\, \widetilde{\ell}^{(m)}_{-i,1}, ..., \widetilde{\ell}^{(m)}_{-i,t-1}\Big) \;\leq\; x^{\max}\,. \tag{6.18}$$

The problem of power consumption scheduling(6.15) under BRD, is expressed now as a standard optimization problem (which is solved iteratively for decision-maker $i \in \mathcal{I}$ by BRD), as follows:

Deterministic Problem-BRD-Dynamical Charging (6.19)

$$\boldsymbol{\pi}_i^{(m)} \in \arg\min_{\boldsymbol{\pi}_i^{\text{DC}}} \sum_{t\in\mathcal{T}} C_{x_1}^{\boldsymbol{\pi}_i^{\text{DC}}}\Big(g_t\Big(x_1, \ell_{i,1}, ..., \ell_{i,t-1};\widetilde{\ell}^{(m)}_{-i,1}, ..., \widetilde{\ell}^{(m)}_{-i,t-1}\Big), \ell_{i,t}; \widetilde{\ell}^{(m)}_{-i,t}\Big)$$

$$\text{s.t.}\;\; (3.5),\, (3.6),\, (3.12),\, (6.11),\, (6.12),\, (6.13),\, (6.17),\, (6.18)\,.$$

This problem redefines the line 11 of the BRD Algorithm 3 for the dynamical charging context. Under the following Proposition 6.3.1, convexity on the latter problem can be guaranteed. The sufficient conditions of such a proposition are assumed to hold here. More precisely, it holds for the models considered in the numerical Section 6.5. Thus, an element of the arg min set can be obtained by solving the corresponding convex optimization problem by using known numerical techniques, precisely in line 11 of the BRD Algorithm 3 under this context. Since the composition of convex functions is also convex, and the inequality constraints of the problem above define a convex and compact set, we simply quote the following proposition.

Proposition 6.3.1

The problem for dynamical charging BRD-based (6.19), is a convex optimization problem if C_{x_1}, $g_1, ..., g_t$, and f are convex.

Compared to a centralized approach shown in Section 3.3.2, the complexity of this problem (6.19) is reduced, since the sequential BRD is linear in the number of rounds needed for convergence (say N, which typically equals 3 or 4) and the number decision-makers I. Therefore, for a numerical routine whose complexity is cubic in the problem dimension, the complexity for the centralized implementation is of the order of $I^3\,T^3$, whereas it is of the order of $N\,I\,T^3$ with the decentralized implementation. Observe also that, in terms of information, all the model parameters of the system, need to be known by each decision-maker for this decentralized implementation to be run. If this turns out to be a critical aspect in terms of identification in practice (e.g., if the DN does not want to reveal physical parameters about its DN-transformer), other techniques which only exploit directly measurable quantities such as the sum of load consumption could be used. This is one of the purposes of the following scheme.

6.3.3 *ITERATIVE VALLEY-FILLING ALGORITHM*

Agreeing with Section 3.3.3, the valley-filling or water-filling algorithm is a quite well-known technique to allocate a given additional energy demand (here, the one induced by the controllable loads of consumers) over time, given a primary demand profile (here, the noncontrollable part of the total load consumption). The main idea is to consume when the primary demand is sufficiently low. Contrary to the valler-filling algorithm presented in Section 3.3.3, where the strategy defines the (aggregated) sum of loads in a centralized way, the implementation here is seen as its iterative version based on BRD, proposed in [19][M4]. For instance, valley-filling has been used in [106] to design a scheduling algorithm, but the iterative implementation is not explored. An iterative version is proposed in [56] which relies on a parallel implementation, where the controllable load profiles are updated simultaneously at each iteration, and convergence is obtained by adding a “penalty” or “stabilizing term” to the cost function. Note that one of the drawbacks in the latter approach is that the weight assigned to the added term has to be tuned properly. Here, the sequential version based on BRD does not have this drawback and can be seen as the power system counterpart of the iterative water-filling algorithm used in communications problems [114].

Following the definition of the centralized problem for the valley-filling method (3.22), its decentralized counterpart (BRD-based) at iteration $m \in \mathbb{N}$, can be written for the decision-maker $i \in \mathcal{I}$ as follows:

Deterministic Problem - BRD - Valley-Filling

$$\boldsymbol{\pi}_i^{(m)} \in \arg\min_{\boldsymbol{\pi}_i^{\mathrm{VF}}} \sum_{t\in\mathcal{T}} \Phi^{\boldsymbol{\pi}_i^{\mathrm{VF}}}\left(\ell_{i,t} + \widetilde{\ell}_{-i,t}^{(m)}\right) \quad \text{s.t.} \quad (3.5),\ (3.6),\ (3.12),\ (6.11),\ (6.13)\ . \tag{6.20}$$

In this problem, Φ is any strictly convex function, see [56] for instance. This problem redefines the line 11 of the BRD Algorithm 3 for the valley-filling context. A solution to the latter problem can be found by using the Lagrange multipliers method [24, 30], which defines the optimal load consumption at time t as:

$$\ell_{i,t}^{(m)} = \min\left\{\ell_i^{\max},\ \max\ \left\{\ell_i^{\min},\ \mu_i - \widetilde{\ell}_{-i,t}^{(m)}\right\}\right\},$$

where μ_i is a threshold chosen to satisfy the constraint (3.5) of the energy demand. Compared to the iterative dynamical charging scheme of Section 6.3.2, an important practical advantage of the decentralized valley-filling is that it relies only on the measure of the total load consumption. However, this solution is based on continuous load levels, which may not be met in some real applications. Additionally, just as the problem of noise robustness for high-order modulations in digital communications, this scheme may also be sensitive to uncertainties on the knowledge of the noncontrollable load demand, e.g., the noise of the forecast (3.12).

To conclude this section, we provide a result which guarantees the convergence of the three proposed decentralized strategies for the deterministic case.

BRD CONVERGENCE UNDER THE PROPOSED STRATEGIES

Convergence is ensured thanks to the exact potential property of the associated charging game (see [18][M4] for more details on the definition of this game). This can be summarized in the following proposition from [19, 79].

Proposition 6.3.2

The iterative schemes BRD-based: rectangular profiles, dynamical charging and valley-filling algorithm; always converge.

This result can be proved by identifying each of the described decentralized strategies as the sequential BRD of a certain "auxiliary" strategic-form game. The key observation to be made is that since a common cost function is considered for the I decision-makers and the individual strategies are reduced to be vectors in $\mathbb{R}^T$, the corresponding problem can be formulated as an exact potential strategic-form game [79]. The important consequence of this is that the convergence of dynamics such as the sequential BRD is guaranteed due to the "finite improvement path" property [89]. Note that although Proposition 6.3.2 provides the convergence for the described power scheduling strategies, the efficiency of the point(s) of convergence is not ensured a priori. Typically, this efficiency can be measured relatively to the one obtained by the centralized problem, that is, the solution built by the central operator which controls all the strategies of consumers under a perfect knowledge on the noncontrollable load consumption. On the other hand, in the game-theory literature, this is called “Price of Anarchy”; finding some bounds providing the maximal loss (“anarchy”) induced by a decentralized implementation is not an easy task in general [79]. In the case of the EVs charging problem, [20] presents some special cases for which explicit bounds are available (even with a zero loss in the asymptotic case of an infinite number of electric vehicles). In the setting of this chapter, this question will be addressed numerically in the Section 6.5 to measure the performance of the proposed strategies.

6.4 SOLUTION METHODOLOGY IN THE STOCHASTIC CASE

In the previous sections, the effect of the forecasting noise on the noncontrollable loads has been ignored. Indeed, the power consumption scheduling algorithms have been designed by assuming a forecast completely deterministic, using it as a simple parameter, representing a single scenario of the noncontrollable loads. If the algorithm is robust against noise, this approach may be suitable. However, the algorithms proposed in the previous sections may not be robust when there is enough noise in the forecast, as shown in [19][M1]. This motivates a modeling which accounts for the effect of the forecasting noise a priori, i.e., based on a stochastic forecast (statistics). The principal motivation of using it is that provide a way to model the noncontrollable load consumption by generating several scenarios, so that a deterministic forecast can be seen as one of these. This sort of forecast model can then be understood as several deterministic forecasts with associated probabilities of occurrence for which large databases exist in order to extract precise statistics. The model using for, is based on Markov decision processes [96]. The principal difference with Section 3.4.1 is that here, we provide an approximation of the Best Response Dynamic [79] to build suitable power consumption scheduling strategies. Since the distribution network experiences an increased amount of variable loads consumption depending on the controllable part, introducing this kind of (adaptable) scheduling strategies allow consumers to reduce their impacts on the distribution network [M1]. Otherwise, it could produce new load variations and possibly causing transformer overloading, power losses, or moreover increasing the transformer ageing [19].

In agreement with Section 2.4.2, we represent a **stochastic forecast** of the (real) noncontrollable load consumption profile $\boldsymbol{\ell}_0$ by:

$$\widetilde{\boldsymbol{L}}_0 = (\widetilde{L}_{0,t})_{t\in\mathcal{T}} , \tag{3.23}$$

which is a finite collection of i.i.d. random variables defined by[1] describing a probability distribution P_0. Suppose now that, for $i \in \mathcal{I}$, $\boldsymbol{L}_i$ is the sequence taking controllable load profiles $\boldsymbol{\ell}_i$, see (6.1). We can then write:

$$\boldsymbol{L}_i \quad = \quad (L_{i,t})_{t\in\mathcal{T}} \ ,$$

where $L_{i,t}$ takes a controllable load value $\ell_{i,t}$ at time t. In this way, the total load consumption (3.3) at t, can be expressed now by:

$$\widetilde{L}_t \quad = \quad \widetilde{L}_{0,t} \ + \ \sum_{i\in\mathcal{I}} L_{i,t} \ . \tag{3.24}$$

Contrary to Section 3.4.1, we are interested here in a decentralized version of the (centralized) scheduling problem (3.30). This also can be understood as looking for the stochastic counterpart of the deterministic problem (6.8), where the noncontrollable load consumption profile (3.23) boils down to a simple vector of parameters and the forecast noise is not taken into account. In this section, the forecast errors are accounted for, under the sequence (3.23).

Based on the stochastic forecast, the induced system state process $(\widetilde{X}_t)_{t\in\mathcal{T}}$, see Definition 2.6.2 for details, has of course a stochastic behavior and then, from the probability distribution P_0, we can get an explicit representation of the transition probabilities between the system states. Consider that the system is at state $\widetilde{x}_t$ at time $t \in \mathcal{T}$, and that each decision-maker $i \in \mathcal{I}$ has chosen its controllable load $\ell_{i,t}$. For the moment, this is supposed to be done "locally", i.e., each decision-maker $i = 1, ..., I$ chooses its controllable load $\ell_{i,t}$ in a state $\widetilde{x}_t$ (according to its scheduling strategy for instance). The probability that the state of the system is $\widetilde{x}_{t+1}$ at time $t+1$ (following the evolution law f of eq. (6.2)), can be computed as:

$$\begin{aligned} P\Big[\, \widetilde{X}_{t+1} = \widetilde{x}_{t+1} \ \Big| \ \widetilde{X}_t = \widetilde{x}_t,\, L_{1,t} = \ell_{1,t},\, ...,\, L_{I,t} = \ell_{I,t} \Big] \\ = \ P_0\Big[\, \widetilde{L}_{0,t} \in \mathcal{L}^f_{0,t} \ \Big| \ \widetilde{X}_t = \widetilde{x}_t\, ,\, L_{1,t} = \ell_{1,t}\, , ...,\, L_{I,t} = \ell_{I,t} \Big] \ , \end{aligned} \tag{6.21}$$

where

$$\mathcal{L}^f_{0,t} \quad := \quad \Big\{\, \widetilde{\ell}_{0,t} \ \mid \ \widetilde{x}_{t+1} \ = \ f(\widetilde{x}_t\, ,\, \ell_{1,t}\, , ...,\, \ell_{I,t}\, ;\, \widetilde{\ell}_{0,t}) \,\Big\} \ .$$

Note that the deterministic case shown in Section 6.3, can be see as a particular case of the present one. Indeed, when $\ell_{1,t}, ..., \ell_{I,t}$ are selected, we can make $\mathcal{L}^f_{0,t}$ to be a singleton (a set with exactly one element) at each t to represent the deterministic forecast (3.12), thus the next state of $\widetilde{x}_t$ is fully determined by making the transition probability to be equal to one for a (next) state according to the evolution law f, and zero for all other (candidate) next states.

As we have seen in Section 2.6.5, some constraints under the stochastic forecast assumption have been discussed. Here, we accept that an

1 Either the original probability measure of a probability space or describing a probability distribution P_0 can be used to compute probabilities of events involving $\widetilde{\boldsymbol{L}}_0$. Here, we use w.l.o.g. P_0 without needing to refer to a common underlying probability space.

event remains under a certain risk of probability that is coded by a given threshold. Under the same arguments of the constraint (3.26) on the total load consumption, and the one on (3.27) of the system state, we write resp. these constraints by:

$$\prod_{t\in\mathcal{T}} P_0\Big[\,\widetilde{L}_t \;\; \leq \ell^{\max}\,\Big|\;\; L_{1,t} \;=\; \ell_{1,t} \;\;, \ldots , \;\; L_{I,t} \;=\; \ell_{I,t} \;\Big] \quad \geq 1-\varepsilon_\ell \tag{6.22}$$

$$\prod_{t\in\mathcal{T}} P\Big[\widetilde{X}_{t+1} \leq x^{\max}\,\Big|\,\widetilde{X}_t = \widetilde{x}_t, L_{1,t} = \ell_{1,t}, \ldots, L_{I,t} = \ell_{I,t}\Big] \quad \geq 1-\varepsilon_x \tag{6.23}$$

where $(\varepsilon_\ell, \varepsilon_x) \in [0,1]^2$ represents the risk in probability of exceeding resp. the upper bounds $\ell^{\max}$ and $x^{\max}$.

To be consistent with the notation, we write the energy demand constraint (3.5) for consumer i as:

$$\Delta_t \sum_{t\in\mathcal{T}} L_{i,t} \;\; \geq \;\; e_i \;, \tag{3.28}$$

with Δ_t is the duration of a time-slot, e.g., if each t represents $30\,\text{min}$, then $\Delta_t = 0.5\,\text{h}$. We also express the constraint (3.6) of not exceeding the maximal power $\ell_i^{\max} \in \mathbb{R}^+$ and to be at least a minimum power $\ell_i^{\min} \in \mathbb{R}_0^+$ at which i can be charging, by:

$$\ell_i^{\min} \;\; \leq \;\; L_{i,t} \;\; \leq \;\; \ell_i^{\max} \;. \tag{3.29}$$

Based on all practical considerations, the decentralized problem of power consumption scheduling (6.7) to be solved for each consumer, can be introduced. A decision-maker $i \in \mathcal{I}$ selects its own power consumption scheduling strategy from the following $\arg\min$ set:

Decentralized Stochastic Problem

$$\underset{\pi_i}{\arg\min} \; \sum_{t\in\mathcal{T}} \mathbb{E}^{\pi}_{x_1}\Big[C^{\pi}(\widetilde{X}_t,\, L_{1,t}, \ldots,\, L_{i,t}, \ldots,\, L_{I,t}\,;\, \widetilde{L}_{0,t})\Big] \tag{6.24}$$

s.t. (3.23), (3.24), (3.28), (3.29), (6.6), (6.21), (6.22), (6.23).

For be this problem computationally tractable, a discretization over the consumption is considered. This problem can thus be modeled using finite Markov Decision Processes [96]. The main difference with the problem (3.30) in Section 3.4 is that here, we are looking for a decentralized scheduling strategy, i.e., each decision-maker i builds its own strategy $\boldsymbol{\pi}_i$ under the fixed strategies of the other decision-makers (see eq. (6.6)) and observing the value of the system state. Agreeing with the iterative solution proposed in Section 6.3 (when a deterministic forecast is available), we also want here to build the strategies of the consumers based on a dynamic of best-response. That is why we present in the next section an approximation of such a dynamic, that was implemented to find a solution of the problem (6.15), but here it will be used on the stochastic context of the problem (6.24).

6.4.1 *ITERATIVE MARKOV DECISION PROCESSES - BASED*

A suitable model to adapt the scheduling problem (6.24) in a discrete manner, is based on multi-weighted Markov Decision Process [M3], see Definition 2.6.9 for more details. Before we explain in detail such a model, we start with an ordinary Markov Decision Process (MDP) model:

$$\mathcal{M} \; = \; \Big(\mathcal{X}, \mathcal{A}, P\Big)\,, \tag{3.31}$$

where $\mathcal{X}$ is a finite set of states, $\mathcal{A}$ is a finite space of actions, and P a transition probability between states. Based on this structure (3.31), the problem of power consumption scheduling can (a priori) be modeled by taking $\mathcal{X}$ as the set of the system states, $\mathcal{A}$ as the space of controllable loads of consumers, and P as the transition probability induced by the probability distribution P_0 of the stochastic forecast, see eq. (6.21). However, it remains to discuss one thing: the way in which the actions are chosen separately by the decision-makers. Moreover, how the scheduling strategy can be found and built in a decentralized manner (recall that in this chapter, each consumer is free to make its own decision in terms of choosing its power consumption strategy). Some models have been proposed when these are chosen separately. Multi-agent MDP (MMDP) is a first instance, which generalize an MDP describing sequential decisions in which there are several agents (decision-makers), where each of them chooses an individual action at every time and jointly optimize a common cost function. However in such a model, the cooperative coordination between decision-makers is to communicate at every time-step, making the complexity increase exponentially in the number of decision-makers [97]. Moreover, it is generally necessary to consider a joint information space with a centralized approach to be optimal, rendering this method impractical in most cases and outside of what we are looking for: to solve sequentially the decentralized scheduling problem (6.24), without revealing the strategies of the consumers among them.

In absence of a central controller choosing a joint scheduling strategy, it is required some coordination mechanism among the decision-makers. Observability is also a subject that can be taken into account. If the decision-makers choose their decisions based on an incomplete observation of the system states, we get into Partially Observable MDP class (POMDP). The complexity of a POMDP (to find an optimal strategy in a finite horizon time problem) with one decision-maker (or a central controller) is PSPACE-COMPLETE[2] even if the state dynamic is stationary and the time-horizon T is smaller than the size of the system state space [93]. When several decision-makers are considered in a POMDP, decentralized executions raise severe difficulties during coordination even for a joint full observability particular case, namely decentralized MDP, which is based on the assumption that the system

2 The PSPACE class refers to all decision problems that can be solved in polynomial space (amount of memory space necessary for the computation). The hardest problems in PSPACE are the PSPACE-COMPLETE problems, since a solution to any one such problem could easily be used to solve any other problem of the PSPACE class.

state is completely determined from the current (joint) observation of the decision-makers. However, solving a decentralized MDP (to find an optimal strategy in a finite horizon time problem) is NEXPTIME-COMPLETE[3] to find an optimal scheduling strategy even for $I \geq 3$ decision-makers [22].

Using any of the previous models to build an optimal strategy for each consumer is a challenging problem, causing it to be very difficult when there are many state in the MDP model and the coordination among the decision-makers is made at every time-step. Moreover, the power consumption scheduling problem considered in this chapter, has constraints (energy demand, upper bound over the system state, etc.), which are not easy to implement in these types of models. Here, an iterative technique among the decision-makers based on the sequential BRD, is performed using "iterative" MDPs.

SEQUENTIAL INFORMATION ON STOCHASTIC TOTAL LOAD CONSUMPTION

We focus here on introducing the scheduling problem (6.24) but from a BRD point of view. As we have seen in Section 6.3, the idea is that for each $i \in \mathcal{I}$, the load consumption of the other consumers $j \in \mathcal{I} \setminus \{i\}$ is part of the noncontrollable load consumption. More precisely, the loads $L_{j,t}$ are added to the stochastic forecast (3.23) resp. at each t, since by Assumption 2.5.1, the law of the system state depends on the total load consumption, here expressed as in eq. (3.24). Suppose for the moment that $\widetilde{L}_{-i,t}$ represents the total load except the load of i. Roughly speaking (it will be explained a further below), this can be written as:

$$\widetilde{L}_{-i,t} \quad = \quad \widetilde{L}_t \; - \; L_{i,t} \; . \tag{6.25}$$

Recall that the information to be shared among the decision-makers at each iteration in Section 6.3, is effectively $\widetilde{L}_{-i,t}$ but in a deterministic sense, see eq. (6.13). The idea is then to build here, a random variable representing the aggregated load consumption of decision-makers except i, such as the one of (6.25), so that a sequential algorithm can be implemented (BRD). Note that in Section 6.3, there is a single value of the consumption except i to be shared at each time, so the subtraction (6.13) is simple to perform, here it is not the case.

Without going into the details (for the moment) of the variable $\widetilde{L}_{-i,t}$, we denote by P_{-i} its underlying probability distribution. We can compute now the probability (6.21) from the point of view of the decision-maker i as follows. Suppose that the action chosen by i is $\ell_{i,t}$ at the state $\widetilde{x}_t$, then the probability to go to $\widetilde{x}_{t+1}$ is computed by:

$$\begin{aligned} P[\, \widetilde{X}_{t+1} = \widetilde{x}_{t+1} \;\mid\; \widetilde{X}_t \; = \widetilde{x}_t,\, L_{i,t} = \ell_{i,t} \,] \\ = \;\; P_{-i}[\, \widetilde{L}_{-i,t} \in \mathcal{L}^f_{-i,t} \;\mid\; \widetilde{X}_t = \widetilde{x}_t \,,\, L_{i,t} = \ell_{i,t} \,] \;, \end{aligned} \tag{6.26}$$

3 That is, the set of all decision problems that can be solved in nondeterministic exponential time. The problems in the NEXPTIME-COMPLETE class are the hardest problems in NEXPTIME. A solution to any one such problem could easily be used to solve any other problem of the NEXPTIME class.

where now,

$$\mathcal{L}^{f}_{-i,t} \;:=\; \{\, \widetilde{\ell}_{-i,t} \mid \widetilde{x}_{t+1} \;=\; f(\widetilde{x}_t\,,\, \ell_{i,t}\,;\, \widetilde{\ell}_{-i,t})\,\}\,.$$

Additionally, the constraints (6.22) and (6.23) are relaxed and imposed to be now:

$$\prod_{t\in\mathcal{T}} P_{-i}[\,\widetilde{L}_t \;\le \ell^{\max} \mid\; L_{i,t} \;=\; \ell_{i,t}\,] \qquad \ge\; 1-\varepsilon_\ell \tag{6.27}$$

$$\prod_{t\in\mathcal{T}} P[\widetilde{X}_{t+1} \le x^{\max} \mid \widetilde{X}_t = \widetilde{x}_t,\, L_{i,t} = \ell_{i,t}\,] \ge\; 1-\varepsilon_x \tag{6.28}$$

In the following, we construct an iterative technique to develop a decentralize methodology BRD-based among the decision-makers to build their scheduling strategies. A decision-maker $i \in \mathcal{I}$ chooses its strategy at iteration $m \in \mathbb{N}$ from the following arg min set:

Decentralized Stochastic Problem - Approximate BRD

$$\pi_i^{(m)} \in \; \arg\min_{\pi_i} \; \sum_{t\in\mathcal{T}} \mathbb{E}^{\pi_i}_{x_1}\Big[\, C^{\pi_i}\Big(\, \widetilde{X}_t,\, L_{i,t}\,;\, \widetilde{L}^{(m)}_{-i,t}\Big)\Big] \tag{6.29}$$

$$\text{s.t.} \quad (3.23),\; (3.28),\; (3.29),\; (6.26),\; (6.27),\; (6.28)\,.$$

Note that the latter scheduling problem can be seen as the stochastic counterpart of the problem (6.15). This problem (6.29) will be modeled by a doubly-weighted MDP, under the a priori definition of $\widetilde{L}^{(m)}_{-i,t}$ at each iteration m as it is shown in the following.

MDP MODELING

First, we write $\mathcal{M}_i$ to refer to the MDP (3.31) that the decision-maker i must resolve to build its own strategy. We focus in the construction of such a model $\mathcal{M}_i$ to solve the scheduling (6.29). For that, consider that a second component is defined on the state space $\mathcal{X}$ of the MDP $\mathcal{M}_i$ to keep the information of the total load consumption at each time-step. More precisely, let $\mathcal{L}_i$ be a (finite) set representing the possible values of the total load (from the point of view of i):

$$\widetilde{L}_{-i,t} \;+\; L_{i,t}\;,$$

and define thus the aggregate space of states as $\mathcal{X} \times \mathcal{L}_i$. Second, since there is a constraint on the energy demand for the decision-maker i, see eq. (3.28), a cost function on state transitions can be defined as the energy due by the controllable loads (actions). More precisely a cost is incurred every time an action is used, that is:

$$C_i(L_{i,t}) \;\;=\;\; \Delta_t\, L_{i,t}\;,$$

where Δ_t is a constant (e.g., if each time-step represents a slot of $30\,\mathrm{min}$ then $\Delta_t = 0.5\,\mathrm{h}$). Thus, a scheduling strategy for i must satisfy the energy demand constraint (3.28) on the accumulated sum of such costs. In addition, another cost function over the state transitions can be defined as the cost C of the scheduling problem (6.29) to minimize. As we will see in the following, finding an optimal scheduling strategy is reduced to solve the so-called Stochastic Shortest Path (SSP) problem [25], see Definition 2.6.23 for details. Thus globally, an MDP with two cost functions can be used to model the scheduling problem under consideration and to find a solution. This model is called doubly-weighted MDP.

While SSP problem for a single-weighted MDPs is well-know to be solved in POLYNOMIAL-TIME [14, 25], multi-weighted MDPs subject to constraints imposed on several objectives (that is the case here, e.g., the energy demand for a consumer), take POLYNOMIAL-TIME in the size of the MDP model and EXPONENTIAL-TIME in the size of the requirements [50]. To reduce the requirement of the energy demand and find an optimal scheduling strategy for i, we build the unfolding of the doubly-weighted MDP by adding recursively the information of the accumulated cost function C_i on the states at each time to reduce the doubly-weighted MDP into a simply-weighted MDP, see Definition 2.6.10 for details. Thus, the new constraint of the energy demand is reduced to be a stochastic reachability objective in such a simply-weighted MDP, see Definition 2.6.16. Contrary to Section 3.4.1, here we do not need any allocation problem, because an optimal scheduling strategy will be found for each consumer i individually by an iterative (BRD) technique between MDPs, and aggregated energy is not relevant here.

The doubly-weighted MDP model to use in our case is of the form:

$$\mathcal{M}_i \;=\; \Big(\mathcal{X} \times \mathcal{L}_i,\, (x_1, 0),\, \mathcal{A}_i,\, P_i,\, C,\, C_i\Big)\,, \tag{6.30}$$

where we are fixed the initial state to be $(x_1, 0)$. We define in the following the unfolding of the latter MDP model, to simplify the constraint of the energy demand to a quantitative reachability objective into an MDP with only one cost function, see Definition 2.6.10.

UNFOLDING THE MDP

The unfolding of $\mathcal{M}_i$ with depth T used here is the following structure:

$$\mathcal{M}_{i,T} \;=\; \Big(\mathcal{S}_{i,T},\, s_1,\, \mathcal{A}_i,\, P_{i,T},\, \boldsymbol{C}_i\Big)\,, \tag{6.31}$$

where the space of states is:

$$\mathcal{S}_{i,T} \;:=\; \mathcal{X} \times \mathcal{L}_i \times \mathcal{E}_i \times \mathcal{T}\,, \tag{6.32}$$

with $\mathcal{E}_i$ representing the space of the accumulated energy of the consumer i, i.e.,

$$\mathcal{E}_i \;=\; [0,\, T\,\ell_i^{\max}]\,.$$

The given initial state is $s_1 = (x_1, 0, 0, 1)$, the set of actions is defined as the set controllable loads of i:

$$\mathcal{A}_i \;=\; \bigcup_{t\in\mathcal{T}} \Big\{ \ell_{i,t} \;\Big|\; \ell_i^{\min} \;\le\; \ell_{i,t} \;\le\; \ell_i^{\max} \Big\}\,. \tag{6.33}$$

The transition probability between states is $P_{i,T} : \mathcal{S}_{i,T} \times \mathcal{A}_i \to \mathcal{D}(\mathcal{S}_{i,T})$ and defined as:

$$P_{i,T}(s_t, \ell_{i,t})(s_{t+1}) \;=\; P_i(\mathrm{proj}_1(s_t), \ell_{i,t})(\mathrm{proj}_1(s_{t+1}))$$

$$\begin{array}{rrcl} \text{if} & (i)\ \mathrm{proj}_1(s_{t+1}) & = & f(\mathrm{proj}_1(s_t),\, \ell_{i,t}\,;\, \widetilde{\ell}_{-i,t})\,, \\ & (ii)\ \mathrm{proj}_2(s_{t+1}) & = & \widetilde{\ell}_{-i,t} + \ell_{i,t}\,, \\ & (iii)\ \mathrm{proj}_3(s_{t+1}) & = & \min\{\, e_i,\, \mathrm{proj}_3(s_t) \,+\, C_i(\ell_{i,t})\,\}, \\ & (iv)\ \mathrm{proj}_4(s_{t+1}) & \leq & T, \end{array}$$

and $P_{i,T}(s_t, \ell_{i,t})(s_{t+1}) = 0$ otherwise. Note that (i) above is implicitly included in the definition of P_i. Finally, the cost function $\boldsymbol{C}_i$ is the same as C. Additionally, from the Definition 2.6.2, we denote by

$$\widetilde{S}_t \;=\; (\widetilde{X}_t,\, \widetilde{L}_t,\, E_t,\, T_t)$$

the state process in the unfolding (6.31).

Since we are interested in reaching the accumulated energy demand e_i of each consumer i, we can naturally define a **set of goal states** $\mathcal{G}_i$ by:

$$\mathcal{G}_i \;:=\; \Big\{ s_t \in \mathcal{S}_{i,T} \;\mid\; \mathrm{proj}_3(s_t) = e_i \;\text{ and }\; \mathrm{proj}_4(s_t) = T \Big\}\,. \tag{6.34}$$

This set is of our interest and we want to find a scheduling strategy for which $\mathcal{G}_i$ is reached, that is to say that the energy demand e_i is achieved. Thus, the problem of interest is reduced to the expected SSP problem, see Definition 2.6.23 for details. Note that there is a natural one-to-one correspondence between the "histories or paths" in the doubly-weighted MDP (6.30) and the unfolding (6.31), and therefore, strategies can equivalently be seen in both MDPs[4].

The only thing that we need is the definition of the variable $\widetilde{L}_{-i,t}$, which represents the aggregated load consumption "exogenous" to the consumer i, see eq. (6.25). Such a variable will be constructed by means of an iterative BRD approach. More precisely, this will evolve in each round $m \in \mathbb{N}$ of an iterative technique based on BRD, by solving sequentially each (unfolding) MDP model of consumers. This can be understood as a multiple-sequential learning method, where each consumer i learns the environment in each MDP to determine its "best response" considering that the other consumers are part of such an environment, i.e., to be part of the noncontrollable load of from the point of view of i, allowing often results in local optima [33]. We detail this dynamics in the following.

ITERATIVE MDP METHODOLOGY

As we said earlier, since decentralized models (e.g., MMDP, POMDP, etc.) distributing explicit state information at each time-step among the decision-makers are computationally costly, we present here an approximated version of BRD between the (unfolding of the) MDPs of consumers, without updating the information of the controllable loads (i.e., the actions of decision-makers) at each state of the model. Here,

4 A formal proof is made in Chapter 4.

we update for each $i \in \mathcal{I}$ and each round $m \in \mathbb{N}$ the amount of the variable $\widetilde{L}_{-i,t}$, which can be seen as an update of a matrix representing the total loads consumption starting from the stochastic forecast (3.23) and so, adding sequentially the load consumption of consumers. To develop the dynamic that we propose, we define first a correspondence between two MDPs already resolved[5]. For the following, we denote by $\mathcal{M}_{i,T}^{(m)}$ the unfolded-MDP of consumer i when the round is m.

Definition 6.4.1: Correspondence between two MDPs

Let $\mathcal{M}_{i,T}^{(m)}$ and $\mathcal{M}_{j,T}^{(n)}$ two unfolded-MDPs already resolved in the rounds $m, n \in \mathbb{N}_0$ resp. for the consumers $i, j \in \mathcal{I}$. Suppose that we extract the values of the total load consumption $\widetilde{\ell}_t^{(n)} = \mathrm{proj}_2(s_t)$ from the states s_t in $\mathcal{M}_{j,T}^{(n)}$ for each $t \in \mathcal{T}$. Fixing one $\widetilde{\ell}_t^{(n)}$, we define by going through the states from the initial state in $\mathcal{M}_{i,T}^{(m)}$, the probability that $\widetilde{L}_t^{(m)}$ is the value $\widetilde{\ell}_t^{(n)}$ by:

$$\mathbb{P}_{s_1}^{\boldsymbol{\pi}_i^{(m)}}\left[\widetilde{L}_t^{(m)} = \widetilde{\ell}_t^{(n)}\right] \quad = \quad \mathbb{P}_{s_1}^{\boldsymbol{\pi}_i^{(m)}}\left[\widetilde{S}_t^{(m)} \in \left[\,\widetilde{\ell}_t^{(n)}\,\right]_i^m\right], \qquad (6.35)$$

where $\left[\,\widetilde{\ell}_t^{(n)}\,\right]_i^m$ is the equivalence class of states in $\mathcal{M}_{i,T}^{(m)}$, defined by:

$$\left[\,\widetilde{\ell}_t^{(n)}\,\right]_i^m \quad = \quad \left\{ s_t \in \mathcal{S}_{i,T}^{(m)} \;\middle|\; \mathrm{proj}_2(s_t) \in \underset{\widetilde{\ell}_t \in \mathcal{L}_i^{(m)}}{\arg\min} \left| \widetilde{\ell}_t^{(n)} - \widetilde{\ell}_t \right| \right\}.$$

In the latter definition, the probability $P_{i,T}^{\boldsymbol{\pi}_i^{(m)}}$ is the one in the model $\mathcal{M}_{i,T}^{(m)}$ when we fix the optimal strategy $\boldsymbol{\pi}_i^{(m)}$ computed by the SSP problem on $\mathcal{M}_{i,T}^{(m)}$. See Section 2.6.4 for details about the induced probability by fixing strategies. Moreover, the probability (6.35) can be computed explicitly from the one shown in the eq. (2.17) when we fix the strategy $\boldsymbol{\pi}_i^{(m)}$, i.e., by means of:

$$\mathbb{P}_{s_1}^{\boldsymbol{\pi}_i^{(m)}}\left[\widetilde{L}_t^{(m)} = \widetilde{\ell}_t^{(n)}\right] \;=\; \sum_{s_1,\dots,s_{t-1}} \; \sum_{s_t \in \left[\,\widetilde{\ell}_t^{(n)}\,\right]_i^m} \; \prod_{\tau=1}^{t} P_{i,T}^{\boldsymbol{\pi}_i^{(m)}}(s_{\tau-1})(s_\tau)\,, \qquad (6.36)$$

where the summation is over all tuples $s_1, \dots, s_t$ in $\mathcal{M}_{i,T}^{(m)}$, such that the state s_t belongs to the equivalence class $\left[\,\widetilde{\ell}_t^{(n)}\,\right]_i^m$. See Figure 6.2 for a graphical representation.

Suppose now that the MDPs will be resolved sequentially for each consumer $i = 1, \dots, I$. We define the following profile:

$$\widetilde{\boldsymbol{L}}_{-i}^{(m)} \quad = \quad \left(\widetilde{L}_{-i,t}^{(m)}\right)_{t \in \mathcal{T}} \qquad (6.37)$$

of the aggregated total load consumption except i at round m, which plays the role of the stochastic environment in the model $\mathcal{M}_{i,T}^{(m)}$, and

5 We say that an MDP is resolved if we have already computed an optimal scheduling strategy for it (taking an optimal action at each state of the MDP), i.e., we have resolve the nondeterministic choice in the MDP in an optimal way.

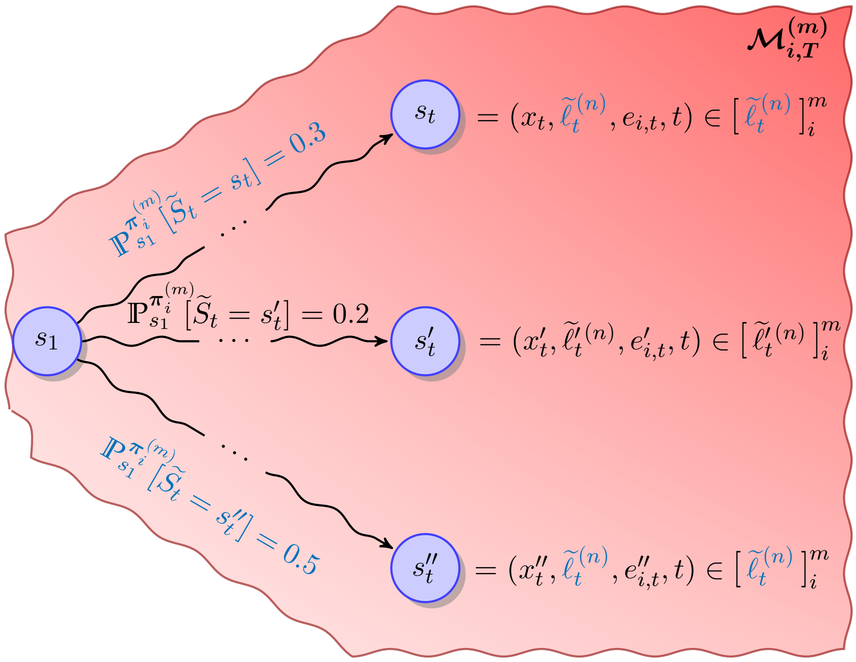


Figure 6.2: Unfolded-MDP $\mathcal{M}_{i,T}^{(m)}$ of the decision-maker i at iteration m. The probability (6.36) for $\widetilde{\ell}_t^{(n)}$ and $\widetilde{\ell}_t'^{(n)}$ are resp. $\mathbb{P}_{s_1}^{\pi_i^{(m)}}[\widetilde{L}_t^{(m)} = \widetilde{\ell}_t^{(n)}] = 0.8$, and $\mathbb{P}_{s_1}^{\pi_i^{(m)}}[\widetilde{L}_t^{(m)} = \widetilde{\ell}_t'^{(n)}] = 0.2$.

defines the probability distribution $P_{i,T}^{(m)}$ (more precisely, it defines $P_{-i}^{(m)}$ in eq. (6.26)). We proceed to explain the sequential construction in function of a consumer $i \in \mathcal{I}$ and a round $m \in \mathbb{N}_0$. First, a round $m = 0$ is considered to initialize the scheduling strategies of the decision-makers.

Round $m = 0$ for $i = 1$: in the beginning, $\widetilde{\boldsymbol{L}}_{-i}^{(m)}$ in (6.37) is exactly the same as the stochastic forecast $\widetilde{\boldsymbol{L}}_0$ of the noncontrollable load consumption (3.23), because there is no other load on the DN to schedule an optimal strategy for i. Thus, the transition probability between the states in the (not yet resolved) unfolded-MDP $\mathcal{M}_{i,T}^{(m)}$ is given by the known probability distribution of such a forecast $\widetilde{\boldsymbol{L}}_0$. Once such a model is built, we can define the set $\mathcal{G}_i^{(m)}$ of states that satisfy the energy constraint for i, see eq. (6.34). Solving the expected SSP problem with the goal set $\mathcal{G}_i^{(m)}$, we can compute an optimal strategy for i, and the model for the next decision-maker can be built.

Round $m = 0$ for $i > 1$: suppose that $\mathcal{M}_{i-1,T}^{(m)}$ has already been resolved. To build $\mathcal{M}_{i,T}^{(m)}$ and then to find a scheduling strategy for i, first we need the information of the aggregated load consumption $\widetilde{\boldsymbol{L}}_{-i}^{(m)}$ of (6.37) due to the noncontrollable loads (the stochastic forecast) and the load consumption of the consumers $1, \ldots, i-1$. Second, after the construction of the model $\mathcal{M}_{i,T}^{(m)}$ is done, we can define a goal set $\mathcal{G}_i^{(m)}$ of states that satisfy the energy constraint for i, see eq. (6.34) for details, to find an optimal strategy for him.

To build $\mathcal{M}_{i,T}^{(m)}$, the necessary information of $\widetilde{\boldsymbol{L}}_{-i}^{(m)}$ is entirely determined by the previously resolved model $\mathcal{M}_{i-1,T}^{(m)}$, since the aggregated load except i is contained at each time t in the random variable $\widetilde{L}_t^{(m)}$ within $\mathcal{M}_{i-1,T}^{(m)}$. More precisely, for a fixed t, each possible value of $\widetilde{L}_t^{(m)}$ is extracted by going through the states of $\mathcal{M}_{i-1,T}^{(m)}$ from the initial state, and identifying the values of the total load consumption $\widetilde{\ell}_t^{(m)}$ and so building equivalence classes of the form $\big[\,\widetilde{\ell}_t^{(m)}\big]_{i-1}^{m}$ in the same model $\mathcal{M}_{i-1,T}^{(m)}$ to define the respective probability distribution of $\widetilde{L}_t^{(m)}$. In other words, we use the Definition 6.4.1 to compute probabilities of the form $\mathbb{P}_{s_1}^{\boldsymbol{\pi}_{i-1}^{(m)}}\big[\,\widetilde{L}_t^{(m)} = \widetilde{\ell}_t^{(m)}\,\big]$ from the model $\mathcal{M}_{i-1,T}^{(m)}$. Thus, each possible value of the "exogenous" variable (from the point of view of i) $\widetilde{L}_{-i,t}^{(m)}$ at t in (6.37) is $\widetilde{\ell}_t^{(m)}$ from the (already resolved) $\mathcal{M}_{i-1,T}^{(m)}$, since $\widetilde{L}_{-i,t}^{(m)}$ is the variable $\widetilde{L}_t^{(m)}$ in the model $\mathcal{M}_{i-1,T}^{(m)}$. The probability distribution of $\widetilde{L}_t^{(m)}$ can be obtained by the correspondence of $\mathcal{M}_{i-1,T}^{(m)}$ with himself (see Definition 6.4.1 with $n = m$ and the consumers $i-1$ and $j = i-1$). This defines thus the stochastic environment for the transition probability $P_{i,T}^{(m)}$ between the states in the not yet resolved model $\mathcal{M}_{i,T}^{(m)}$ for the consumer i. Once that the latter model is built, an optimal strategy for i can be computed by defining the expected SSP problem under the set $\mathcal{G}_i^{(m)}$ in (6.34).

When a strategy for the consumer $i = I$ is already computed, we proceed with the sequential technique based on an approximate BRD for the rounds $m > 0$ as follows.

Round $m > 0$ for $i \geq 1$: as we have said above, all the information to built the model $\mathcal{M}_{i,T}^{(m)}$ is contained in the previous resolved model[6] $\mathcal{M}_{i-1}^{(m)}$. The main difference with the previous case is that each total load consumption expressed by the variable $\widetilde{L}_t^{(m)}$ in the model $\mathcal{M}_{i-1,T}^{(m)}$, it contains the controllable loads of the consumer i made in the previous round $m-1$. The idea is to do something like in eq. (6.13), but here we should do a priori:

$$\widetilde{L}_{-i,t}^{(m)} \;=\; \widetilde{L}_t^{(m)} \;-\; L_{i,t}^{(m-1)} \;, \tag{6.38}$$

since $\widetilde{L}_{-i,t}^{(m)}$ will act as the stochastic environment for the model $\mathcal{M}_{i,T}^{(m)}$. However, in the previous equation, the elements are different:

(i) $\widetilde{L}_t^{(m)}$ is the variable of total load consumption in the model $\mathcal{M}_{i-1,T}^{(m)}$ of the consumer $i-1$ at round m.

(ii) $L_{i,t}^{(m-1)}$ is the variable representing the controllable loads in the model $\mathcal{M}_{i,T}^{(m-1)}$ of the consumer i at round $m-1$.

6 Well understood that if $i = 1$, the information is in $\mathcal{M}_{I,T}^{(m)}$.

The problem is the following: suppose that we take a value of $\widetilde{L}_t^{(m)}$, let say $\widetilde{\ell}_t^{(m)}$, then:

$$\begin{gathered}\textit{Which controllable load } \ell_{i,t}^{(m-1)} \textit{ could we use to do} \\ \textit{the subtraction } \widetilde{\ell}_t^{(m)} - \ell_{i,t}^{(m-1)} \textit{ ?}\end{gathered} \tag{6.39}$$

Of course, there could be many choices when we are trying to identify $\widetilde{\ell}_t^{(m)}$ in the states of the model $\mathcal{M}_{i,T}^{(m-1)}$. Indeed, since the states in the latter model are of the form $s_t \in \mathcal{S}_{i,T}^{(m-1)}$, see eq. (6.32), then it is possible that there are several states s_t having the second component to be $\widetilde{\ell}_t^{(m)}$, i.e., there could be many states $s_t \in \big[\,\widetilde{\ell}_t^{(m)}\,\big]_i^{m-1}$. The question is now:

$$\begin{gathered}\textit{Which action was selected by the decision-maker } i \textit{ at} \\ \textit{time } t-1 \textit{ to get to the states } s_t \in \big[\,\widetilde{\ell}_t^{(m)}\,\big]_i^{m-1} \textit{ ?}\end{gathered}$$

Of course, there could be several. Even if $\Big|\big[\,\widetilde{\ell}_t^{(m)}\,\big]_i^{m-1}\Big| = 1$, there could not be a single choice. Indeed, since the second component of a state $s_t \in \mathcal{S}_{i,T}^{(m-1)}$ are of the form $\mathrm{proj}_2(s_t) = \widetilde{\ell}_{-i,t-1}^{(m-1)} + \ell_{i,t-1}^{(m-1)}$, we could have several combinations to get the sum:

$$\widetilde{\ell}_{-i,t-1}^{(m-1)} \;+\; \ell_{i,t-1}^{(m-1)} \;=\; \widetilde{\ell}_t^{(m)}\;,$$

because this depends on the "exogenous" load $\widetilde{\ell}_{-i,t-1}^{(m-1)}$ and the action $\ell_{i,t-1}^{(m-1)}$ selected by i. Moreover, if $\ell_{i,t-1}^{(m-1)}$ is unique,

$$\textit{From which state } s_{t-1} \textit{ was the action } \ell_{i,t-1}^{(m-1)} \textit{ selected ?}$$

Again, there could be many states $s_{t-1} \in \mathcal{S}_{i,T}^{(m-1)}$ having a positive probability $P_{i,T}^{(m-1)}\big(s_{t-1}, \ell_{i,t-1}^{(m-1)}\big)(s_t) > 0$ for $s_t \in \big[\,\widetilde{\ell}_t^{(m)}\,\big]_i^{m-1}$.

For all these reasons, we propose a method to make an effective calculation of the eq. (6.38). More precisely, we answer the question (6.39). For that, we will define in the following an "expected action" (controllable load) from the model $\mathcal{M}_{i,T}^{(m-1)}$ to do the subtraction of loads under consideration. Let $\widetilde{\ell}_t^{(m)}$ a fixed value of the variable $\widetilde{L}_t^{(m)}$ in the model $\mathcal{M}_{i-1,T}^{(m)}$. Recall that in the model $\mathcal{M}_{i,T}^{(m-1)}$ we have computed an optimal scheduling strategy $\boldsymbol{\pi}_i^{(m-1)}$ from the SSP problem. We define from $\mathcal{M}_{i,T}^{(m-1)}$ the set:

$$\mathcal{S}_{i,t-1}^{(m-1)}\Big(\big[\,\widetilde{\ell}_t^{(m)}\,\big]_i^{m-1}\Big) \;:=\; \bigcup_{s_t \in \big[\widetilde{\ell}_t^{(m)}\big]_i^{m-1}} \Big\{ s_{t-1} \in \mathcal{S}_{i,T}^{(m-1)} \mid P_{i,T}^{\boldsymbol{\pi}_i^{(m-1)}}(s_{t-1})(s_t) > 0 \Big\}$$

of the precedent states of each $s_t \in \big[\,\widetilde{\ell}_t^{(m)}\,\big]_i^{m-1}$ by applying $\boldsymbol{\pi}_i^{(m-1)}$, see Figure 6.3 for a graphical representation.

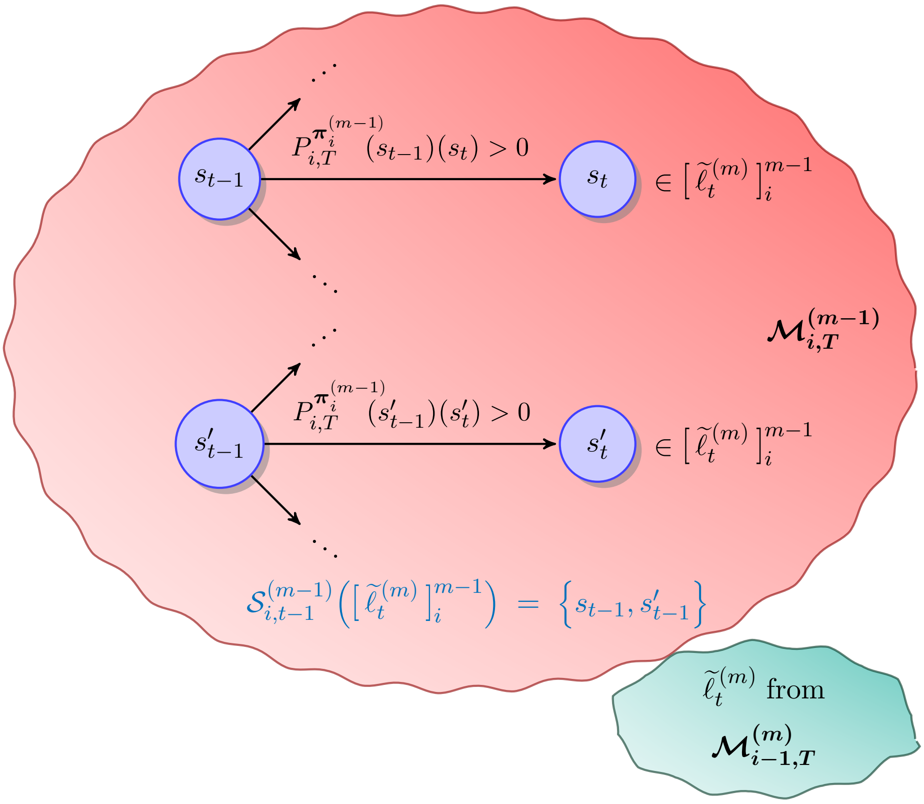


Figure 6.3: In the unfolded-MDP $\mathcal{M}_{i,T}^{(m-1)}$ of the decision-maker i at iteration $m-1$, we build the equivalence class of each $\widetilde{\ell}_t^{(m)}$ extracted from the unfolded-MDP $\mathcal{M}_{i-1,T}^{(m)}$ of the decision-maker $i-1$ at iteration m, and we identify the previous states to go to the states of $[\widetilde{\ell}_t^{(m)}]_i^{m-1}$.

Thus, we define the expectation on the controllable load (action) selected by i at iteration $m-1$ at time t, $L_{i,t}^{(m-1)}$, according to the scheduling strategy $\boldsymbol{\pi}_i^{(m-1)}$ applied over all the states belonging to the set $\mathcal{S}_{i,t-1}^{(m-1)}\Big(\big[\, \widetilde{\ell}_t^{(m)} \,\big]_i^{m-1}\Big)$, by:

$$\mathbb{E}_{s_1}^{\boldsymbol{\pi}_i^{(m-1)}}\Big[\, L_{i,t-1}^{(m-1)} \;\Big|\; \widetilde{S}_{t-1} \in \mathcal{S}_{i,t-1}^{(m-1)}\Big(\big[\, \widetilde{\ell}_t^{(m)} \,\big]_i^{m-1}\Big) \Big]$$

$$= \frac{\displaystyle\sum_{s_1,\dots,s_{t-2}} \; \sum_{s_{t-1}\in\mathcal{S}_{i,t-1}^{(m-1)}([\widetilde{\ell}_t^{(m)}]_i^{m-1})} \; \sum_{s_t\in[\widetilde{\ell}_t^{(m)}]_i^{m-1}} \pi_{i,t-1}^{(m-1)}(s_{t-1}) \prod_{\tau=1}^{t} P_{i,T}^{\boldsymbol{\pi}_i^{(m)}}(s_{\tau-1})(s_\tau)}{\mathbb{P}_{s_1}^{\boldsymbol{\pi}_i^{(m-1)}}\big[\, \widetilde{L}_t^{(m-1)} = \widetilde{\ell}_t^{(m)} \,\big]}$$

where $\pi_{i,t-1}^{(m-1)}(s_{t-1}) = \ell_{i,t-1}^{(m-1)}$ is the action selected at the state s_{t-1}. This is a possible value of the random variable $L_{i,t-1}^{(m-1)}$ according to the strategy $\boldsymbol{\pi}_i^{(m-1)}$. Note that the denominator represents a normalization, which can be computed by means of the Definition (6.4.1) by considering resp. the models $\mathcal{M}_{i,T}^{(m-1)}$ and $\mathcal{M}_{i-1,T}^{(m)}$. See Figure 6.4 for a graphical representation, where the expectation above gives:

$$\mathbb{E}_{s_1}^{\boldsymbol{\pi}_i^{(m-1)}}\left[L_{i,t-1}^{(m-1)} \mid \widetilde{S}_{t-1} \in \{s_{t-1}, s'_{t-1}\}\right]$$
$$= \frac{0.2 \times 2 \times (0.3+0.5) + 0.1 \times 3 \times 0.6}{0.2 \times (0.3+0.5) + 0.1 \times 0.6} = 2.27$$

Note that such an expectation can be understood as the value

$$\text{“}\mathbb{E}_{s_1}^{\boldsymbol{\pi}_i^{(m-1)}}\left[\pi_{i,t}^{(m-1)}\Big(\{s_{t-1}, s'_{t-1}\}\Big)\right].\text{”} .$$

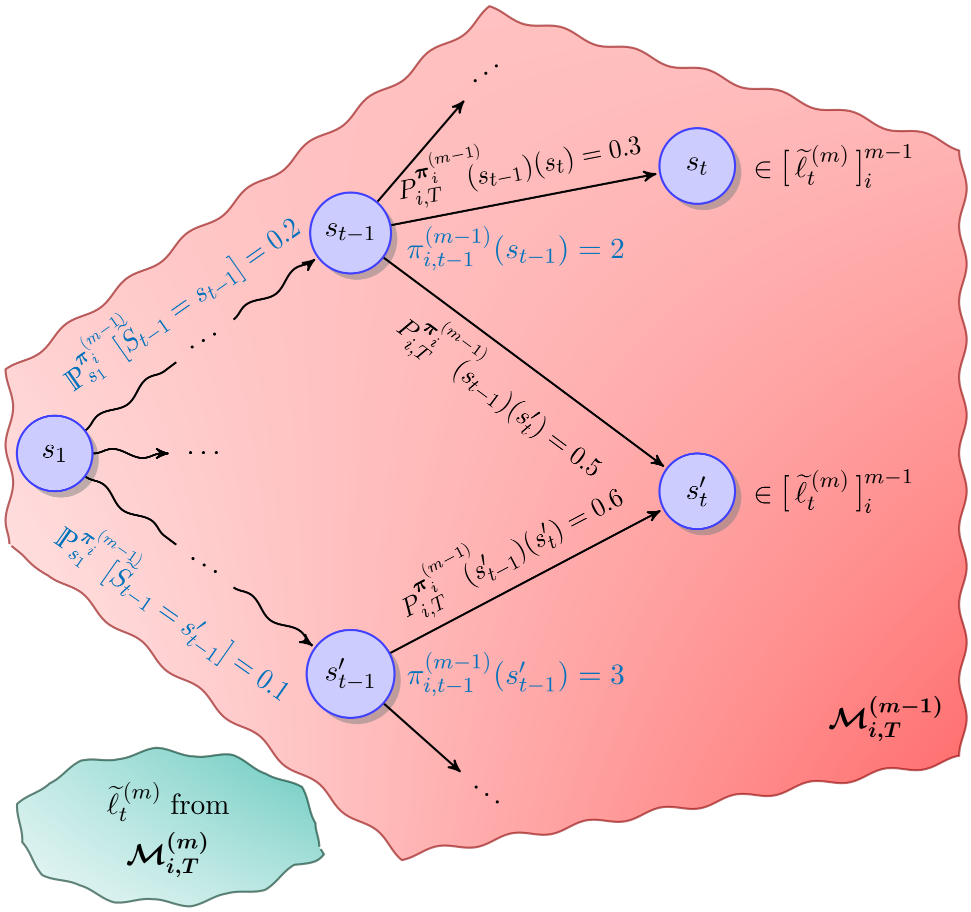


Figure 6.4: Expected action (controllable load) selected by the decision-maker i at iteration $m-1$ at time $t-1$, in the unfolded-MDP $\mathcal{M}_{i,T}^{(m-1)}$.

Finally, the aggregated total load consumption at t except i, i.e., $\widetilde{L}_{-i,t}^{(m)}$, to build the model $\mathcal{M}_{i,T}^{(m)}$, can be defined by means the following equation:

$$\widetilde{L}_{-i,t}^{(m)} = \widetilde{L}_t^{(m)} - \mathbb{E}_{s_1}^{\boldsymbol{\pi}_i^{(m-1)}}\left[L_{i,t-1}^{(m-1)} \mid \widetilde{S}_{t-1}^{(m-1)} \in \mathcal{S}_{i,t-1}^{(m-1)}\Big(\big[\widetilde{L}_t^{(m)}\big]_i^{m-1}\Big)\right].$$

6.5 NUMERICAL APPLICATION

The numerical applications are based on the same conditions of those in Section 3.5, we briefly explain the scenarios and the value of parameters.

First, this section is separated in two main parts. The first one is based on a technical approach [70], and the second one on an economical approach [8]. Both are analyzed in the noisy deterministic and stochastic forecast cases. Before introducing the numerical application, we introduce three other techniques which we will use to compare the performance of the proposed strategies to solve the decentralized power consumption scheduling problem presented in this chapter.

OTHER METHODS TO COMPARE THE PERFORMANCES

To compare the performance of the proposed scheduling strategies based on a sequential best response dynamic (rectangular profiles, dynamical charging, valley filling and Markov decision process), three other methods to schedule power consumption strategies are considered in this section, all are based on the deterministic forecast case:

(i) **Plug-and-Charge (PaC)**: this method is one of the best known to schedule strategies, which is obtained by assuming that the consumption of the controllable electric devices starts as soon as they plugin to the grid, minimizing the time needed to reach the cumulative energy demand.

(ii) ***Gan et al.*** [56]: this algorithm aims at minimizing a cost which results of two terms: a valley-filling solution and a penalty to stabilize the parallel implementation-based algorithm (the I charging vectors $(\boldsymbol{\ell}_i)_{i\in\mathcal{I}}$ are update simultaneously over the algorithm iterations). Convergence to this algorithm is obtained by adding a penalty (or stabilizing) term to the cost. The weight put on the penalty term here is 0.5 according to [18, 56]. If this weight is not tuned properly, the implementation of *Gan et al.* algorithm may lead to significant performance losses [113].

(iii) ***Shinwari et al.*** [106]: this method considers that the energy demand of each consumer, as the one in (3.5), is spread by filling the "holes" of the noncontrollable load consumption as a valley-filling method. However with this algorithm, for each consumer i, a proportion of the energy demand is allocated to a given time-slot t proportionally to:

$$\frac{-\ell_{0,t} + \max_{t\in\mathcal{T}} \ell_{0,t}}{\sum_{t\in\mathcal{T}} -\ell_{0,t} + \max_{t\in\mathcal{T}} \ell_{0,t}},$$

and the remainder is uniformly allocated.

SYSTEM STATE

From the evolution law of the system state (6.2), we consider a function parametrized by two values to represent the two approaches jointly (one technical and one economical). In this section, the system state evolves with the function f defined with a single period time-lag in the load consumption, according to:

$$x_{t+1} = f(x_t, \ell_{1,t}, ..., \ell_{I,t}, \ell_{0,t}, \ell_{1,t-1}, ..., \ell_{I,t-1}, \ell_{0,t-1}) ,$$

where $x_1 \in \mathbb{R}^+$ is the given initial condition for the system state, $(\ell_{i,t-1})_{i\in\mathcal{I}}$ is the zero vector in $\mathbb{R}^I$ when $t = 1$ (do not confuse with the real noncontrollable load profile (3.1)), and $\ell_{0,0}$ can be understood as the noncontrollable load consumption of base at $t = 0$. Now, for two parameters $p, q \in \mathbb{N}$, a general representation of the function f showing the two approaches that we are interested, has the form:

$$\begin{aligned} & f(x_t, \ell_{1,t}, ..., \ell_{I,t}, \ell_{0,t}, \ell_{1,t-1}, ..., \ell_{I,t-1}, \ell_{0,t-1}) \\ & \quad = \alpha x_t + \beta\Big(\ell_{0,t} + \sum_{i\in\mathcal{I}} \ell_{i,t}\Big)^p + \gamma\Big(\ell_{0,t-1} + \sum_{i\in\mathcal{I}} \ell_{i,t-1}\Big)^q + z_t , \end{aligned} \tag{6.40}$$

where $\alpha, \beta, \gamma \in \mathbb{R}$ are constants of the (technical or economical) model, and $z_t \in \mathbb{R}_0^+$ is completely deterministic making always x_t positive. In particular, the parameters p and q are of the form $p = q = 1$ for the economical model, and $p = q = 2$ for the technical one.

The simulations are performed over the chosen time (slot) corresponds to $30\,\mathrm{min}$, so that $\Delta_t = 0.5\,\mathrm{h}$, and the DN-transformer is a $20\,\mathrm{kV}/410\,\mathrm{V}$ transformer whose apparent power is $100\,\mathrm{kVA}$ and nominal (active) power is $90\,\mathrm{kW}$, which approximately corresponds to a district of 30 households.

CONTROLLABLE LOAD CONSUMPTION

The controllable load operations occur within the time window from $5\,\mathrm{pm}$ to $7\,\mathrm{am}$ of the next day, i.e., $T = 30$. During the rest of the day (from $7\,\mathrm{am}$ to $5\,\mathrm{pm}$ of the next day), the total load consumption on the DN-transformer only consists of the noncontrollable loads. The controlled devices (consumers) are considered here be a set of Electric Vehicles (EVs), and each $i \in \mathcal{I}$ represents one of them. Unless specified otherwise, the minimum and maximum controllable load induced by one EV i are resp.

$$\ell_i^{\min} = 0\,\mathrm{kW} \qquad \text{and} \qquad \ell_i^{\max} = 3\,\mathrm{kW} . \tag{6.41}$$

This is a standard assumption for home charging without any additional connectors to plugin [19, 56]. The arrival and depart time of the EVs are fixed for the simulations and are chosen randomly to be the closest integers of realizations of Gaussian random variables $\mathcal{N}(4, 1.5)$ and $\mathcal{N}(28, 0.75)$ resp. Additionally, the total energy demand of each EV is obtained from the distance to be covered for each of them in the next trip[7]:

$$e_i = \lambda_i \frac{24\,\mathrm{kWh}}{150\,\mathrm{km}} (29.4 + 8)\,\mathrm{km}\,,$$

where $24\,\mathrm{kWh}$ is the capacity of a Renault Zoe battery, $150\,\mathrm{km}$ is the corresponding (average) distance covered, $(29.4 + 8)\,\mathrm{km}$ is the average daily distance covered ($29.4\,\mathrm{km}$ to commute and $8\,\mathrm{km}$ for another trip), and λ_i is the taken margin by EV-users to be confident not running out of energy when driving. The latter is generated randomly once between $\{1.5,\ 2,\ 2.5,\ 3\}$.

NONCONTROLLABLE LOAD CONSUMPTION

To represent the real noncontrollable loads profile $(\ell_{0,t})_{t\in\mathcal{T}}$, four scenarios are used. First, based on historical data taken from the Ausgrid Australian DN-operator for Sydney [3], we choose randomly 30 households representing a district. Second, we use a subtractive clustering method [40] with an influence range of 0.9 to generate four representative clusters, where each one stands for a scenario of the real noncontrollable load profile, e.g., representing a season of the year. Figure 3.5 shows a graphical representation of the four scenarios.

One important component is how to assess the impact of not being able to forecast the noncontrollable load consumption perfectly, i.e., the values of the profile $(\ell^k_{0,t})_{t\in\mathcal{T}}$ in each scenario $k = 1, ..., 4$. Here, we consider that the forecast can be either deterministic or stochastic, see Section 2.4 for details. When the forecast is deterministic (resp. stochastic), it turns into a noise vector (resp. a random variable), one for each scenario of the noncontrollable loads. In both cases, the noise of the forecast is assumed to be a zero-mean additive white Gaussian noise with known variance σ_k^2. Thus, the models (2.4) and (2.6) are used for the numerical purposes. When the stochastic approach is assumed, a discretization is considered (this will be explained a little further). Considering k fixed, the variance is obtained on the time under consideration $\mathcal{T}$ by **Signal-to-Noise Ratio** (SNR) expressed in decibel (dB), which allows one to measure to what extent the noncontrollable load consumption of the scenario k can be forecasted:

$$\mathrm{SNR} := 10\log_{10}\left(\frac{1}{T\sigma_k^2}\sum_{t\in\mathcal{T}}(\ell^k_{0,t})^2\right).$$

For example, fixing a $\mathrm{SNR} = 7\,\mathrm{dB}$, we compute σ_k^2 for the scenario k and then, we obtain

(i) a noisy vector of length T for the deterministic forecast wherein each component is a single value from $\mathcal{N}(0, \sigma_k^2)$, and

(ii) a random variable distributing $\mathcal{N}(0, \sigma_k^2)$ to construct a stochastic forecast.

7 In this way, the energy demand represents approx. the 40%–80% of capacity of a RENAULT Zoe or Fluence EV, similarly that those in [82].

For computational aspects, a discretization over is made on the latter, wherein a normalized histogram is used to obtain relative frequencies. A graphical representation is shown in the Figure 3.6.

DYNAMICAL CHARGING SETTINGS

To solve the problem for dynamical charging BRD-based (6.19) for each decision-maker $i \in \mathcal{I}$, we need an explicit representation of the functions $(g_t)_{t\in\mathcal{T}}$ in (6.17). In this case, it is straightforward to show that $(g_t)_{t\in\mathcal{T}}$ under the dynamic law (6.40) for the parameters $p = q = 1$ (economical case) and $p = q = 2$ (technical case), are of the form: $g_1(x_1) = x_1$ and

$$\begin{aligned} &g_{t+1}\Big(x_1, \ell_{i,1}, \ldots, \ell_{i,t}\,;\, \widetilde{\ell}^{(m)}_{-i,1}, \ldots, \widetilde{\ell}^{(m)}_{-i,t}\Big) \\ &\quad = \alpha^t x_1 + \sum_{\tau=1}^{t} \alpha^{t-\tau}\Big(\beta\Big(\widetilde{\ell}^{(m)}_{-i,\tau} + \ell_{i,\tau}\Big)^p + \gamma\Big(\widetilde{\ell}^{(m)}_{-i,\tau-1} + \ell_{i,\tau-1}\Big)^q + z_\tau\Big)\,. \end{aligned}$$

Under realistic values of α, β and γ (that we gives in next for the two cases analyzed here), these functions are convex. In particular when $p = q = 2$, the convexity is guaranteed if $\alpha\,\beta + \gamma \geq 0$, which is the case here.

MARKOV DECISION PROCESS SETTINGS

Concerning the MDP used to built the scheduling strategy when a stochastic forecast is considered, we focus on $\mathcal{M}_{i,t}$ for each $i \in \mathcal{I}$ defined in (6.31). Wherein, the energy demand of each EV (3.28) defines a qualitative objective, more precisely it defines a set of goal states $\mathcal{G}_i$ to be reached, see eq. (6.34). On the other hand, the space of actions $\mathcal{A}_i$ of each EV i defined in eq. (6.33), is discretized according to the minimum and maximum controllable load induced by EVs (6.41) with a fixed parameter $\Delta_{\mathcal{A}} = \ell_i^{\max}$, so that this stands as the set $\mathcal{A}_i = \{0,\ \ell_i^{\max}\}$. The discretization over the stochastic forecast is made with $\Delta_b = 0.05$ to adjust the bin width for the (discrete) probability distribution, wherein normalized histograms are considered to obtain relative frequencies. See Figure 3.6 for a graphical representation of the discretization on the stochastic forecast. Under this practical considerations, the space of states in each $\mathcal{M}_{i,T}$ is discrete as well.

To solve the MDPs under consideration, we use PRISM tool [4, 78], which is a probabilistic model checker, having direct support for MDPs and incorporates the majority of the techniques from [53] to quantify properties specified on MDPs. With this practical tool,, we only need to specify the dynamic (6.40) and the parameters of the model to generate each unfolded-MDP $\mathcal{M}^{(m)}_{i,T}$. In addition, we use MATLAB and PRISM iteratively to coordinate the information of the consumption among the decision-makers i at each iteration m. So that, PRISM solves the MDPs and MATLAB builds and distributes the sequential information for the BRD-approach.

PRICE OF DECENTRALIZATION

In the game-theory literature (see e.g., [79, 84]), to compute maximal losses induced by a decentralized implementation (typically by "Price of Anarchy" based on Nash equilibrium) is not an easy task in general [46, 51, 90]. Here, we present a new notion to interpret the loss of efficiency due to decentralized decision-making for the simulations. Intuitively, we propose to call it the **Price of Decentralization** (PoD), and it is defined as follows.

Consider that $\boldsymbol{\ell}_0$ is the real noncontrollable load consumption profile (3.1). We define PoD to be the ratio between the optimal value of the objective function under an optimal scheduling strategy (computed by solving its respective deterministic or stochastic power consumption problem), and the optimal value of the objective function under a "perfect forecast", i.e., knowing perfectly $\boldsymbol{\ell}_0$ without noise. The latter assumption means that the problem is completely deterministic and since there is no forecast noise on the noncontrollable load consumption, the centralized dynamical consumption strategy introduced in Section 3.3.2 is optimal for the two cases studied here. Denoting the respective optimal scheduling strategy by $\boldsymbol{\pi}^*$, the PoD can be written as:

$$\mathrm{PoD}(\boldsymbol{\pi}) := \frac{\sum_{t\in\mathcal{T}} C_{x_1}^{\boldsymbol{\pi}}(x_t,\, \ell_{1,t}, \dots, \ell_{I,t},\, \ell_{0,t})}{\sum_{t\in\mathcal{T}} C_{x_1}^{\boldsymbol{\pi}^*}\Big(g_t(x_1, \boldsymbol{\ell}_1, \dots, \boldsymbol{\ell}_{t-1}\,;\, \ell_{0,1}, ..., \ell_{0,t-1}),\, \ell_t\,;\, \ell_{0,t}\Big)}\,, \quad (6.42)$$

where, $\boldsymbol{\pi}$ can be the strategy $\boldsymbol{\pi}^{\mathrm{RP}}$, $\boldsymbol{\pi}^{\mathrm{DC}}$, $\boldsymbol{\pi}^{\mathrm{VF}}$ or $\boldsymbol{\pi}^{\mathrm{MDP}}$, computed resp. in Section 6.3.1, Section 6.3.2, Section 6.3.3, and Section 6.4.1. Note that the performance of the strategies are computed "a posteriori", i.e., forecast errors on data are simulated and the strategies are scheduled over a noisy (deterministic or stochastic) scenario of the environment to observe after their performances over a "perfect forecast", i.e., over the real profile $\boldsymbol{\ell}_0$, which is a simple sequence of values.

6.5.1 DN-TRANSFORMER LIFETIME

The goal of this section is to quantify the performance of the different scheduling strategies on the DN-transformer aging. For this, the dynamic law (6.40) represents the Hot-Spot (HS) temperature of the DN-transformer, whose nominal temperature is assumed to be $x_1 = 98\,°\mathrm{C}$, and the shut-down HS temperature is $x^{\max} = 150\,°\mathrm{C}$. The corresponding values of the parameters in (6.40) are as in [70]: $p = q = 2$, $\alpha = 0.83$, $\beta = 30.91\,°\mathrm{C.kW}^{-2}$, $\gamma = -19.09\,°\mathrm{C.kW}^{-2}$ and,

$$z_t \;=\; 0.17\,(8.47 + x_t^{\mathrm{amb}})\,,$$

where x_t^{amb} denotes the ambient temperature at time t. Data corresponding to the latter temperature is obtained from the Australian Bureau of Meteorology for the New South Wales territory [5].

To model the DN-transformer lifetime (in years), we consider that when a scheduling strategy $\boldsymbol{\pi}_i$ for each EV $i = 1, \ldots, I$ is fixed, then writing $\boldsymbol{\pi} = (\boldsymbol{\pi}_i)_{i \in \mathcal{I}}$, the lifetime is given by:

$$\text{Lifetime} \quad := \quad T \frac{\text{Lifetime}_0}{\sum_{t \in \mathcal{T}} C_{x_1}^{\boldsymbol{\pi}}(x_t)} ,$$

where the noncontrollable loads consumption is normalized so that without the consumption of the EVs, the DN-transformer lifetime is $\text{Lifetime}_0 = 40\,\text{years}$. In addition, the cost function to be used by each decision-maker i in the scheduling problem (deterministic or stochastic), is considered as the instantaneous Accelerated Ageing Factor (AAF), which measures the speed of degradation relatively to the given nominal transformer temperature. A well admitted model for AAF is [68]:

$$C_{x_1}^{\boldsymbol{\pi}_i}(x_t) \quad = \quad \exp(a x_t + b) ,$$

where $a = 0.12\,°\text{C}^{-1}$ and $b = -11.32$.

PERFORMANCE OF THE SCHEDULING STRATEGIES

The Figure 6.5 shows the performance of the different strategies proposed in this work, with addition of the method of PaC and the charging strategies of *Gan et al.* [56] and of *Shinwari et al.* [106], in function of the lifetime of the DN-transformer (mean over scenarios) against the number of EVs, where the forecast of the noncontrollable load consumption is assumed to have a noise based on $\text{SNR} = 15\,\text{dB}$ in each scenario. The black top dashed line corresponds to the case under the real noncontrollable consumption and without EVs. The performances are similar to those of Figure 3.7 for the centralized strategies counterpart. In particular, the decentralized version of rectangular profiles has little better performance here, about 5 years in terms of lifetime for $I = 20$ EVs. This can be explained because of rectangular profiles in the centralized case, places the rectangular profiles in a one-shot manner for the consumers, making the method less flexible that in the decentralized case, because here it "learns" at each iteration to find a best response until convergence. On the other hand, in the same case for $I = 20$ EVs, the strategy built by the iterative MDP-based approach has an insignificant loss of performance (about 2 years in terms of lifetime), that can be explained because of the discretization of the MDPs at each iteration. However it is still the same for all other amounts I of EVs.

As before, the PaC strategy is seen to be non-acceptable, and have the same performance in Figure 6.5 as in the centralized case of Figure 3.7. This is evident since the method is the same (the consumption starts as soon as an EV is plugged in to the grid). The methods of of *Gan et al.* and *Shinwari et al.* are not better than the proposed methods here, as it is shown in Figure 6.5. First, the strategy built by the method of *Gan et al.*, which is valley-filling-based, has almost the same performance that other proposed methods in terms of DN-transformer lifetime up to $I = 10$ EVs. However, this is not better than the others after $I = 12$ EVs. This can be explained because this method has a "penalty" to stabilize the parallel implementation, which is a very sensible parameter affecting the efficiency [113]. Second, the method of *Shinwari et al.* has the worst performance compared to the other methods (with exception of PaC). This one is also valley-filling-based but the energy demanded is allocated in a different way, impacting considerably its performance in terms of DN-transformer lifetime.

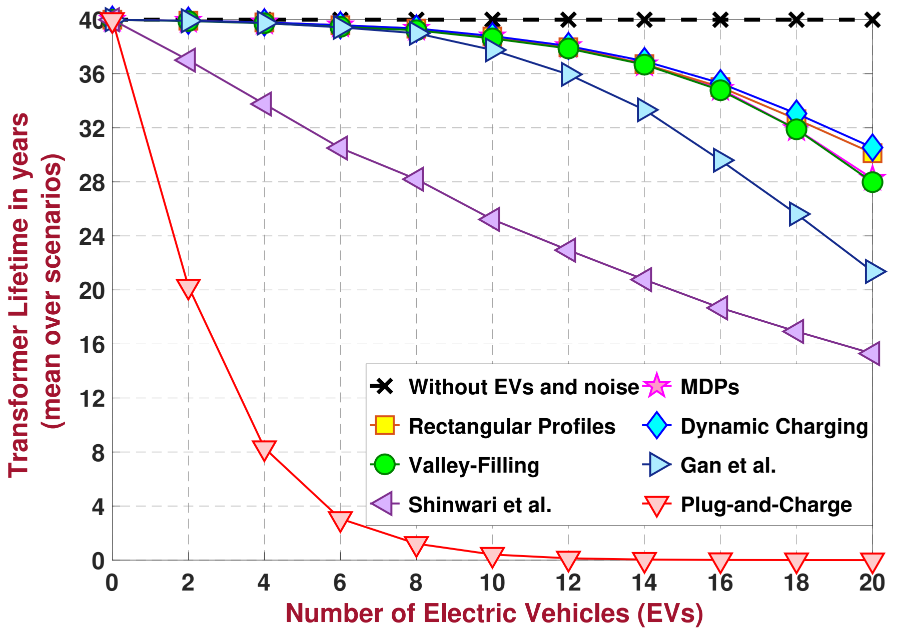


Figure 6.5: DN-transformer lifetime (mean over the scenarios) against the number of EVs (I), based on a forecast of the noncontrollable consumption with a noise (in each scenario) based on SNR = 15 dB (decentralized case).

Figure 6.6 represents the DN-transformer lifetime against different values of forecast noise for $I = 15$ EVs. For small noise values (i.e., when the SNR is high), all methods proposed in this work have good performances. The only strategy which is robust, it is the one built by the iterative MDP-based approach. Its performance in terms of DN-transformer lifetime remains proportional to the performances of the centralized case shown in Figure 3.9 (which is based on $I = 10$ EVs), and it is almost insensitive to forecast noises compared to the other methods. Again, the methods (BRD-based) rectangular profiles, dynamical charging and valley-filling lose performance considerably up to a noise based on SNR = 7 dB. The strategy built by the method of *Gan et al.* has a performance close to valley-filling, since this is a parallelization valley-filling-based. The method of *Shinwari et al.*. looks more stable but it has the worst performance. The particular efficiency over each scenario confirms all these observations in Figure 6.7. Note the behavior of the methods for SNR values below 7 dB. The only strategy that gives an stable efficiency is the one built by the iterative MDP method. The others methods are considerably "chaotic". The presence of oscillations on certain curves could be explained by the fact that the four scenarios of the noncontrollable consumption used in this work, they are in themselves not to be "smooth curves" (they are stochastic), compared to the inputs of others works (see e.g., [20, 56, 82]). By the fact that the scenarios are stochastic, adding a high forecast noise, in some cases this "helps" to the performance of some algorithms. This is shown for example for SNR values below 7 dB in Figure 6.6 and Figure 6.7.

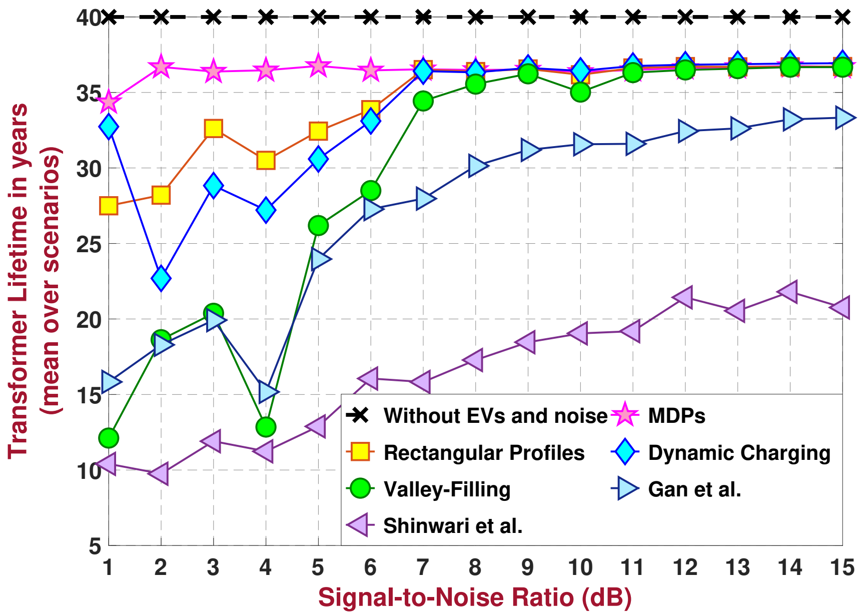


Figure 6.6: DN-transformer lifetime (mean over scenarios) against forecast noises of the noncontrollable consumption, for $I = 15$ EVs (decentralized case).

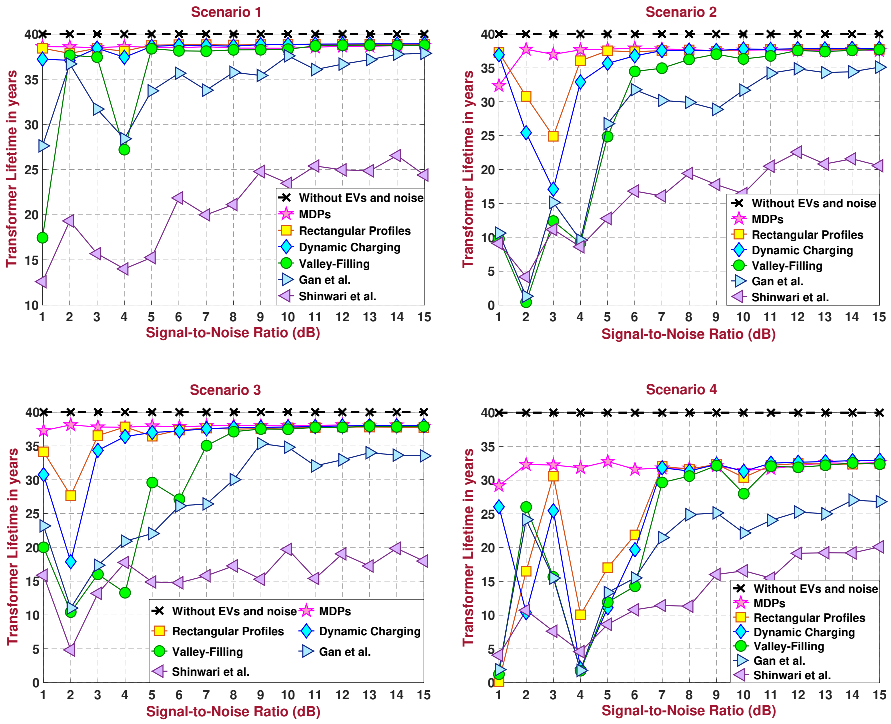


Figure 6.7: DN-transformer lifetime (in each scenario) against forecast noises on the noncontrollable consumption, for $I = 15$ EVs (decentralized case).

We turn now to the verification of the strategy built by the iterative MDPs approach. The probability constraints (6.27) and (6.28) are computed a posteriori, i.e., after that the strategy has been built, we compute the probabilities of satisfying the constraints in the MDPs. In this case, we have observed that the constraint of exceeding the upper bound of the maximal power is satisfied always with probability one. This is due because the value of the maximal power in the simulations ($\ell^{\max} = 90\,\text{kW}$) was sufficient for all energy demands of EVs. However, the probability of exceeding the maximum HS temperature of the DN-transformer is only satisfied with probability one from a forecast noise based on SNR $= 7\,\text{dB}$ and up, as it is shown in Figure 6.8 which is based on the assumption of $I = 15$ EVs. This is due mainly because the variance of the stochastic forecast is very high when SNR is approaching zero (i.e., when the forecast noise is high). This is confirmed by the probability values of the dashed line in Figure 6.8, which represents the the values without any EV, i.e., even when there is no EVs, the forecast model of the noncontrollable consumption makes the maximum temperature of the DN-transformer is reached with some probability. This result is in line with the values shown in Figure 3.13.

To conclude this section, the convergence of the iterative MDPs is only analyzed numerically for stopping criteria. This is shown in Figure 6.9, in which the expected value for $I = 4$ EVs in plotted in function of the iterations between the MDPs. Since we have not used formal methods to analyze and formalize a convergence criterion, this problem remains open.

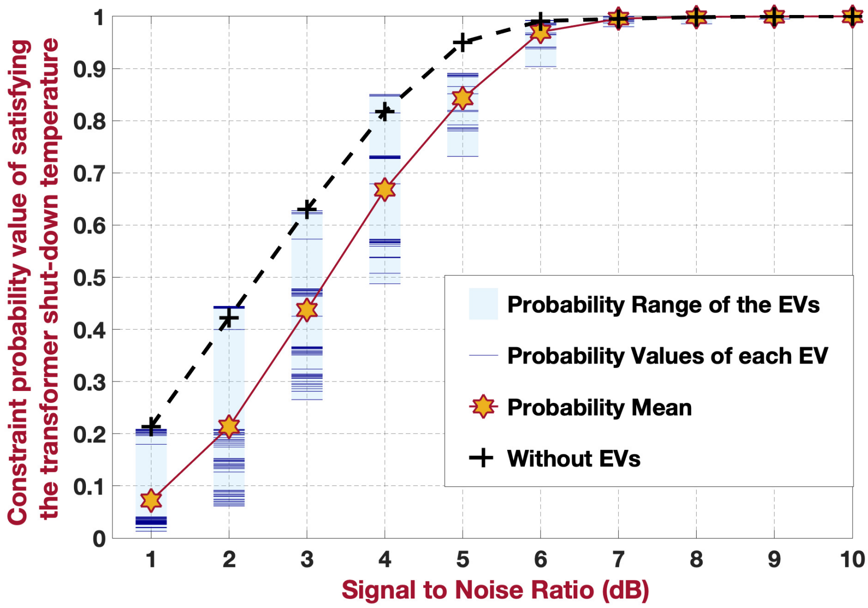


Figure 6.8: Probability value of satisfying (6.28) against forecast noises of the noncontrollable consumption, for $I = 15$ EVs (decentralized case).

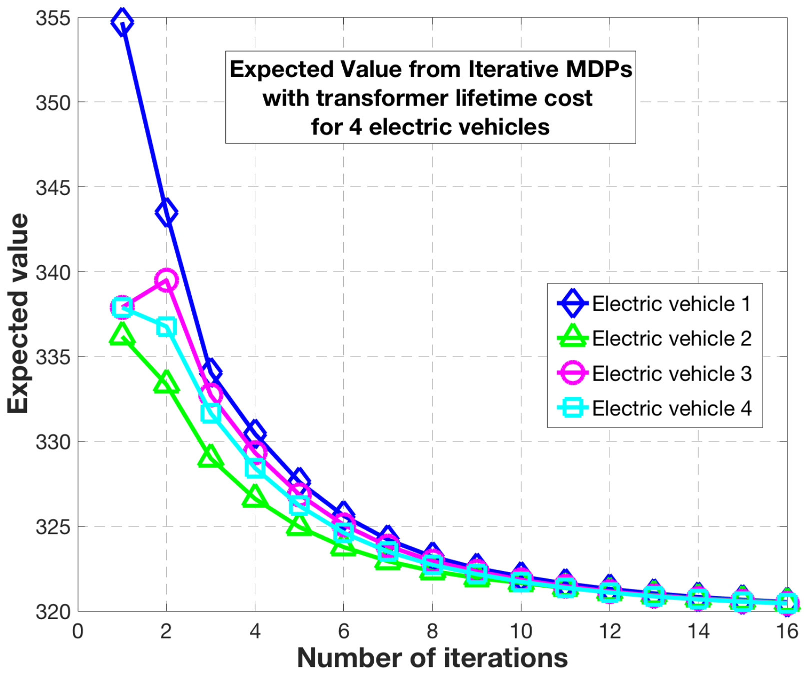


Figure 6.9: Expectation values from the iterative MDPs-based approach for $I = 4$ EVs, based on a forecast of the noncontrollable consumption with a noise based on $\text{SNR} = 4\,\text{dB}$.

6.5.2 *ELECTRICAL CONSUMPTION PAYMENT*

Here, we want to quantify the performance of different scheduling strategies on the Electricity Consumption Payment (ECP). The dynamic of the system state (6.40) represents the EVs total electricity bill corresponding to the total load consumption, whose parameters are [8]: $p = q = 1$, and the other values were estimated by solving a data-fitting problem in least-squares sense from the national electricity market of Australia for New South Wales territory [6]: $\alpha = 1$, $\beta = 11.91$ ¢.kWh^{-2}, $\gamma = -\beta$ and $z_t = 0$. A convenient way of measuring the ECP is formulated by the cost function as in [45, 105], which is used by each decision-maker i in the scheduling problem (deterministic or stochastic):

$$C_{x_1}^{\pi_i}(x_t, \ell_{i,t}, \ell_{-i,t}) \;\; := \;\; (\,\ell_{-i,t} \; + \; \ell_{i,t}\,)^{\eta}\, x_t\,,$$

where $\eta = 3$ is a coefficient indicating the impact of nonlinearity of the total load consumption. See [26] for details, and the initial electrical price is assumed to be $x_1 = 39.65$ ¢ taken from the data.

PERFORMANCE OF THE SCHEDULING STRATEGIES

In this section, we focus in the PoD defined in eq. (6.42) to measure the performance of the scheduling strategies. First, Figure 6.10 shows the PoD against the number of EVs, for a forecast noise based on SNR = 5 dB, which is an acceptable assumption of noise. The strategy built by the MDP methodology has the best efficiency with respect to the optimal case (compared with the centralized dynamical consumption strategy under a perfect forecast of the noncontrollable load consumption). Plug-and-Charge and the strategy of *Gan et al.* are not acceptable, they have the worst performance in terms of price of decentralization. Figure 6.11 shows the same results as those of Figure 6.10 but without these two strategies. The strategies of the iterative MDP, rectangular profiles and dynamic charging; seem to have a uniform behavior in term of price of decentralization as the number of EVs increases (they are stable) as it is shown in Figure 6.11. The method of valley-filling loses some performance up to $I = 10$ EVs and gains some efficiency after $I = 12$ EVs. Recall that this method tends to be a uniform consummation when it has already "filled the valleys" of the noncontrollable consumption (forecast). This is mainly confirmed by the Figure 6.12, which shows the PoD for $I = 15$ EVs. In such a case, the strategy by valley-filling becomes "independent of the noise", consuming uniformly over time for this number of EVs. In Figure 6.11, we observe that the PoD of the strategy from *Shinwari et al.*, is better as the number of EVs increases, but it is not better that the strategies of dynamic charging, rectangular profiles and MDPs-based.

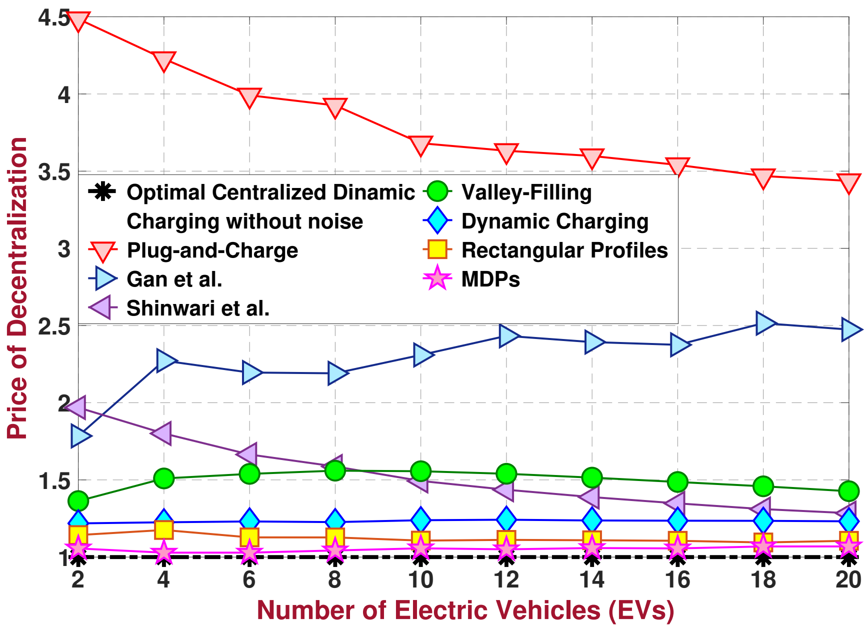


Figure 6.10: Price of decentralization (mean over scenarios) against the number of EVs (I), based on a forecast of the noncontrollable consumption with a noise based on SNR = 5 dB (decentralized case).

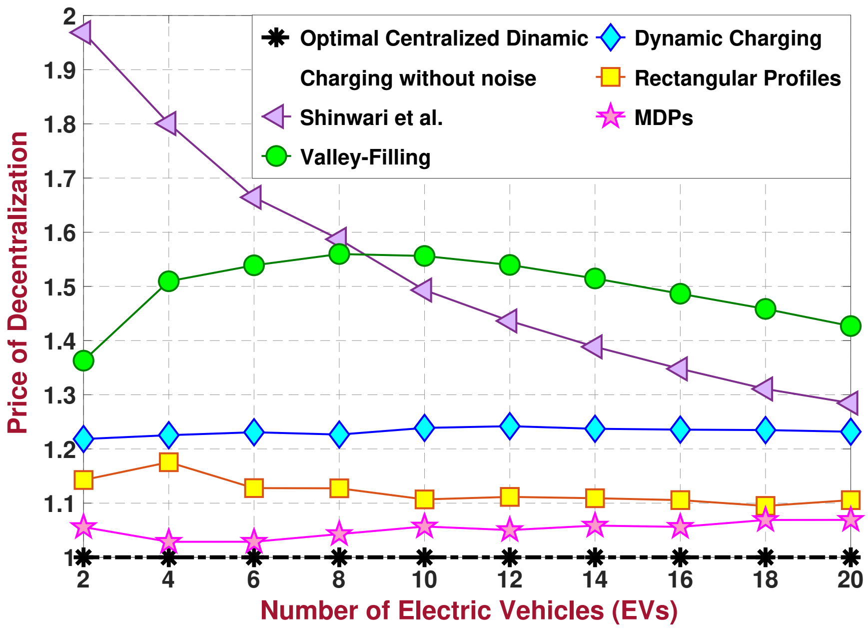


Figure 6.11: Price of decentralization (mean over scenarios) against the number of EVs (I), based on a forecast of the noncontrollable consumption with a noise based on SNR $= 5\,\text{dB}$ (decentralized case).

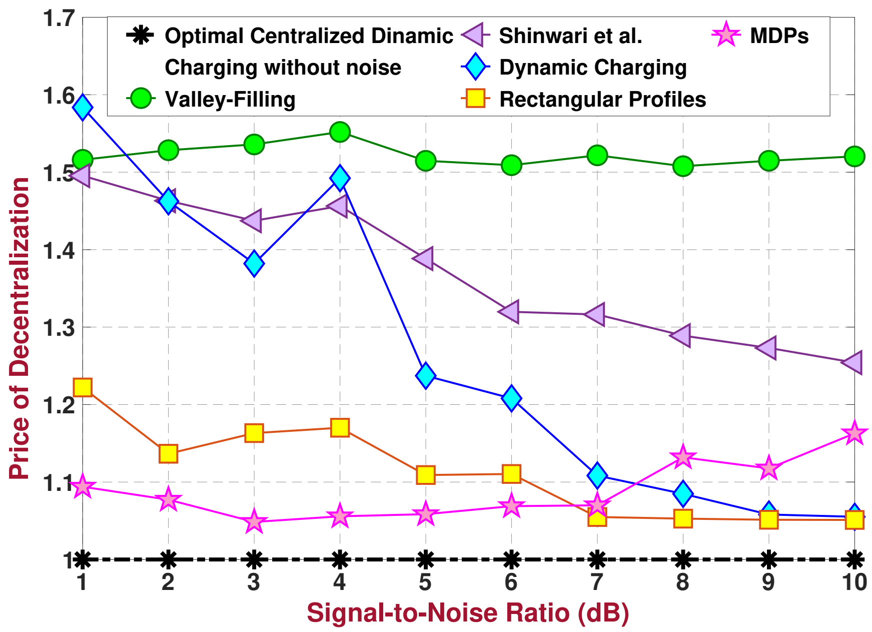


Figure 6.12: Price of decentralization (mean over scenarios) against forecast noises of the noncontrollable consumption, for $I = 15$ EVs (decentralized case).

Figure 6.12 shows the PoD against different forecast noises for $I = 15$ EVs. Note that when the forecast noise is small (for high values of SNR), the (decentralized) dynamic charging tens to be its centralized counterpart. The strategy of rectangular profiles has a better efficiency that dynamic charging and it is more robust to the noise. However, this not better than the strategy from the MDP-approach for high noise values. The approach of the iterative MDPs is more adaptable to the noise. Conversely, when the noise is small, the strategy MDP-based loss efficiency. This can be explained by the discretization made in all the variables.

6.6 DISCUSSION

In this Chapter 6, we have described the decentralized approach of the power consumption scheduling problem of Chapter 3. Here, each consumer is seen as a decision-maker and some information has to be coordinated among consumers to build appropriate power consumption strategies. The main reason for decentralization is to reduce the complexity of the problem in terms of information. In the centralized case, there is a single decision-maker that controls and builds all strategies of consumers. In this chapter, we have proposed decentralized strategies, making the procedure to be only relied on local information. The main idea was to use a technique so-called sequential best response dynamic among consumers to find the “best responses” in terms of power consumption strategies faces to the information of the total load consumption at each iteration. When the forecast of the noncontrollable load is stochastic, it is not yet clear how to make a proper algorithm. Here, we have proposed a promising technique based on iterative MDPs, having good performances in the numerical applications studied in this chapter. The formalization of the stochastic method through formal methods and the study of the respective convergence, remain open so far.

Part IV

CONCLUSION

This part presents the conclusions we have drawn from our research in this work and the suggestions of several ideas, computational challenges, approaches, and so on; for related future works.

7

DISCUSSION AND PERSPECTIVES

Abstract:

We first give a brief summary of our research in this chapter. We discuss the main models developed and results obtained in this work and how these fit in formal methods. In particular, from control designs and applications (as those based on smart grids), toward multi-constrained quantitative models (as those based on Markov decision processes). We examine some limitations of our work and how they could be addressed. We also review questions left open in this document. We close the thesis with a general vision of promising research directions linked to this work for short and long term prospects.

7.1 SUMMARY AND CONTRIBUTIONS

This work is well-justified by the need to construct system controllers (strategies) that are not only functionally correct but also efficient. Synthesizing appropriate strategies is not easy and classical development techniques based on testing (or trial-and-error) are largely inadequate. Through the use of formal methods, we have essentially provided mathematical formal models, algorithmic solutions and appropriate strategies to guarantee correctness in control designs and applications, mainly from the domain of smart grids. The power consumption scheduling problem studied in this work is an example of such an application area, and its design required a formal mathematical approach to ensure that it behaves correctly and in an optimal way (or at least with good performances). In simple terms, this is a scenario in which the consumers have a certain energy demand and want to have this demand to be fulfilled before a set deadline. Therefore, each consumer has to choose at each time the consumption power so that the accumulated energy reaches a desired level, adapting their decision-making to the constraints of the system, e.g, to those of the electricity Distribution Network (DN). For the purposes of numerical analysis, this scenario was considered a pool of Electric Vehicles (EVs), where we have examined technical and economic interactions between the EVs and the DN.

The paradigm of multi-constrained objectives in stochastic systems in general, and in Markov Decision Processes (MDPs) in particular, has recently arisen. It allows us to express various (quantitative or qualitative) objectives over a model and to synthesize strategies accordingly. In practice, the performance of reactive systems is impacted by interplays and trade-offs between several criteria, see for instance [15, 101] for some overviews. This new field of research is very rich and ambitious, with various types of objective combinations. For recent developments, one can cite for example: probability of conjunctions of objectives [62], trade-offs between expected value and variance [31, 83], percentile-query problem for various quantitative payoff functions [102], mean-payoff and energy games [36], conditional values-at-risk [76], and so on. The contributions of this thesis also participate in this new field of multi-criteria objective models, and it is shown through this work that the use of formal methods allows efficiency and robustness accuracy in the synthesis of strategies, in particular for the area of smart grids. This is mainly confirmed in the numerical applications made in this work.

7.1.1 *PART I: BACKGROUND*

A brief introduction of the core concepts of the research field and general context of this work was presented in the Part I. This part contains the following two chapters.

We have started the first part with the Chapter 1, which explains the three connected fields that were mainly covered in this thesis: reactive systems, formal methods and game theory. Formal methods were of particular interest in this work, since these are essential to assert the correctness and efficiency of strategies or controllers of systems that can be founded from a control, game or reactive modeling. MDPs are standard models in such fields and these were used mainly in this thesis for the modeling of a problem coming from smart grids, namely: the **power consumption scheduling problem**. Such a problem is a generalization of the EVs charging problem that was inspired first, from the modeling made in a deterministic setting reported in [19], and second, from the practical results of our work reported in [M1] in the stochastic setting.

The basic representation of the problem mentioned above, for a scenario in which several consumers have an energy requirement and the system state evolves with a dynamic law, was introduced in the Chapter 2. Precisely, we have first placed the context of the work of this thesis in the dynamic of research of smart grids, and second, introduced both important and relevant mathematical tools used mainly through this work. In particular, the background of stochastic methods represented on directed graphs and also, the type of strategies employed through this thesis, namely: pure, mixed and randomized strategies. In this work, we have settled two forms of decision-making through strategies for consumers:

(i) **Centralized**: in which there is a single decision-maker for all consumers, or

(ii) **Decentralized**: in which there are several decision-makers, each of them representing a consumer.

This gave rise to the next two parts of this document, which resp. cover the centralized and decentralized setting for the problem of power consumption scheduling.

7.1.2 *PART II: CENTRALIZED MODELING*

The Part II of this work has explored the centralized approach of the scheduling problem, in the sense that the strategies were assumed to be controlled and built by a single entity, which was responsible for satisfying individual requirements. This part begins with the Chapter 3 that generalizes our interesting practical results reported in [M1].

In the Chapter 3, we have settled the (centralized) modeling for the power consumption scheduling problem in a general mathematical form, to reduce the impact of consumption operations on the DN. From an application view, the centralization was understood as a centralized system operator of the DN, which controls and builds the strategies of consumers satisfying the requirements of the individual energy demands. This single decision-maker was attributed certain knowledge about the noncontrollable part of the total consumption to schedule the strategies. The scenario considered in this chapter was that a dayhead decision has to be made and a deterministic or stochastic forecast on the noncontrollable consumption was available.

First, when the forecast was deterministic, the strategies are reduced to be parameter vectors in which each component is a consumption power. Three sorts of strategies were provided in this case:

(i) **Rectangular profiles**: for the first approach, the strategies were imposed to be rectangular profiles, that is to say that we not only assume that each controllable consumption can only take two values (namely, the minimum and maximum power at which a consumer can be consuming), but also that when it is consuming, the consumption has to be uninterrupted. This approach was inspired from [20], but the centralized case was not analyzed there[1]. Here, we have modeled the centralized problem with such consumption methodology and used it for the numerical applications. Since the strategies were rectangular profiles here, each one boils down to a simple decision (namely, the time at which the consumption effectively starts), reducing the amount of variables in the making-decision, since the centralized controller chooses here a single parameter for each consumer (the consumption start time) and not the complete vectors of power consumption.

(ii) **Dynamical consumption**: in contrast with the previous method, the consumption profile was no longer rectangular and was arbitrary. The motivation for this was to have a better performance on the impact of consumption operations, but also to be able to control the system dynamic state. With rectangular profiles, the system was controlled in a one-shot manner computing the system values at each instant based on the current information. Here, the state dynamic was taken into account explicitly and the state was controlled. For instance, it was possible to guarantee the upper bound on the system state considered in the problem, which is not well suited with rectangular profiles (since it is less flexible). This approach was motivated from [19], but here we have provided a generalized method on the dynamic law of the system, and then to the one of [19] was a particular case. Under some conditions of such a dynamic (more precisely, convexity conditions), the problem boils down to a simple (convex) optimization problem.

(iii) **Valley-Filling**: The last approach was presented only for purposes of numerical comparison. This method replaces the power consumption problem by a valley-filling algorithm. This algorithm is a quite well-known technique (see e.g., [106]) to allocate a given additional energy demand over time (the one induced by the consumers here), given a primary demand (the noncontrollable consumption expressed here by the deterministic forecast). The idea is to consume when the noncontrollable demand is sufficiently low, and therefore it relies only on the minimization of the total

1 Among the possible reasons, may be the relative complexity of solving centralized problems. Since all variables and information are controlled by a single decision-maker. We have also observed it in the numerical applications at the end of this Chapter 3.

consumption and not on any other measure of impact over the DN.

Contrary to the previous methods, in which the effects of forecasting noises on the noncontrollable consumption has not been taken into account (since the forecast was a simple noisy vector of parameters), a stochastic forecast was then assumed. This provides a way to model the noncontrollable consumption by generating several scenarios, for which large databases exist (and precise statistics can be extracted). The resulting strategies can therefore adapt to different noncontrollable consumption events. After the general mathematical form of the (centralized) problem in a stochastic setting was presented, a discretization over has been considered leading to use of an MDP model (namely, a **multi-weighted MDP**). More precisely, an MDP with several requirements (as the one presented in our work reported in [M3]). After, inspired from [34], we have developed a method (namely, the unfolding of the MDP) to reduce the number of requirements in the model. Since it is very unrealistic and considerably harder to cover a priori all constraints of the model, the two constraints in probability were considered to be verified a posteriori by model checking method. Taking into account a priori the other constraints, the multi-constrained scheduling problem was then reduced to the well-known stochastic shortest path problem (see e.g., [14, 25]).

At the end of this chapter, a numerical analysis came into play to measure the performance of the provided strategies (namely, the impact of the consumption operations), where the scenario of consumers was considered a pool of EVs. We have examined here the technical and economic interactions between EVs and the DN. In particular, first to maximize the DN-transformer lifetime and second, to reduce the total electrical consumption payment. Each application was made independently and no trade-off was analyzed. In outline, the numerical applications showed that the strategy built by the MDP modeling was almost insensitive to forecasting noise. This is in line with our practical results of [M1]. Particularly, when the forecasting noise was high, the performance of the other methods decreased considerably and the MDP method is globally much more robust. When the forecasting noise is close to zero, i.e., under a "perfect forecast" (which is a very ambitious hypothesis), the dynamic charging method became optimal.

Note that all these resulting schemes developed in the deterministic and stochastic setting, have a high computational complexity due to the degree of information used to schedule (since, it was in a centralized way), in particular when the time-horizon and the number of consumers are large. This is the main purpose of Chapter 6 to overcome this problem "without losing performances".

While correctness and optimality are in the core of formal methods, applications domains like in smart grids (in particular, the power consumption scheduling problem of this work) have only seldomly been attacked through formal models. Formal methods come into play mainly in this Chapter 4 covering the research related to the existence of strategies for the general multi-constrained problem defined in this thesis.

This problem was inspired from the previous Chapter 3 and integrates both the "beyond worst-case" paradigm of [34], and the mix of probabilities and expectations as in [37]. This chapter has explained our theoretical results reported in [M3].

We have focused here on **multi-weighted MDPs**, particularly in those with a probability constraint over a weighted reachability condition, and a quantitative constraint on the expected value of a random variable defined using a weighted reachability condition. We have investigated the "cartography" of the problem when one parameter varies (a threshold), namely: the set of values of such a parameter for which the problem has or has not a solution. A partial cartography was obtained via an algorithmic approach based on two sequences of optimization problems, in which their solutions were assumed under **randomized strategies** (a method to obtain the synthesis was developed in [M3] and in the next Chapter 5). We have also discussed feasibility of the resulting approach to guarantee existence of appropriate strategies, in particular for the decision-making of a central controller of a system against the (stochastic) behavior of its environment (e.g., systems coming from control designs and applications in smart grids).

While in the previous Chapter 4, we have focused on the existence of a randomized strategy solution for the general multi-constrained problem, in this Chapter 5 we have investigated the synthesis of a solution for such a problem under **mixed strategies**. Particularly, in the existence and construction of a solution for each element in the two sequences of optimization problems (each element is an optimization problem), in which its objective function was a probability and the constraint was an expectation upper bounded by a threshold. Obtaining a mixed strategy here, it also allows us to define a randomized strategy solution for the general multi-constrained problem. This chapter explains our research reported in [M2] and in a forthcoming paper [M5].

The idea was first to fix an index in the two sequences (the same index for both), and define a general problem representing together the two problems taken from each sequence. In the previous chapter, such problems were defined for randomized strategies and the existence of a solution was analyzed. Here, we have defined a joint problem and we have analyzed its counterpart for mixed strategies. The idea was to find an optimal mixed strategy and then, to define an optimal randomized strategy for the initial problem. An optimal mixed strategy was obtained by solving a **dual optimization** problem.

Since mixed strategies can be understood as a convexification of pure strategies, we have also defined the pure strategy problem (that is the same, but the only difference is the sort of strategy to be used). We have identified then two primal problems: one for mixed strategies and another for pure strategies. In addition, we have defined their problem counterpart from duality optimization, that is, resp. the mixed and pure dual problems. Having these four problems in mind (pure/mixed primal problems and pure/mixed dual problems), we have shown that to solve the primal mixed problem, it is equivalent to solve the dual pure problem. An optimal mixed solution (for the primal mixed problem) was built by combining two pure strategies: one satisfying the constraint in the expectation and another one that does not. This mixed strategy gave the equality in the constraint in expectation (with the upper bound parameter). Since we focus in a strategy with the strict inequality in the constraint, we perturb such a mixed strategy to have another one that effectively satisfies the strict inequality and to be near to the optimum. With this strategy in mind, we use a theorem from game theory (namely, the Kuhn's theorem that can be applied in this context), to define a randomized strategy solution of the initial problem.

7.1.3 *PART III: DECENTRALIZED MODELING*

The Part III of this work has discussed the decentralized approach of the main problem of interest: the problem of power consumption scheduling. Contrary to the Part II, the strategies for consumers are controlled and built here by several decision-makers, each of them representing a consumer. This part has explained our work reported in [M4], the one of a forthcoming paper [M5], and also it is in line with our practical results of [M1]. This part contains the following chapter.

The Chapter 6 has described the general mathematical modeling of the power consumption scheduling problem, assuming that the consumption operations had been decentralized, in the sense that:

(i) each consumer was seen as a decision-maker, so that each of them was free to make its own decision in terms of choosing its consumption power (i.e., controlling and building its own strategy), and

(ii) it had to be decentralized information-wise, i.e., the scheduling algorithm or procedure (implemented by a machine, which is the most common scenario), only relied on local information.

The main idea was then to use a technique so-called sequential Best Response Dynamic (BRD) among consumers. BRD acted as a coordination mechanism through decision-makers to ensure that the scheduled strategies were consistent with the problem. To handle the decentralized problem and to make an effective calculation of the strategies, the scenario considered in this chapter was that a day-head decision has to be made and a deterministic or stochastic forecast on the noncontrollable consumption was available as in the Chapter 3. To schedule a strategy for a consumer, it was attributed certain knowledge about the total consumption, so that a consumer updated its strategy in a round-robin manner by solving the problem from "its point of view". Precisely, this allowed to built a way to coordinate information among the decision-makers to determine their "best response behaviours", by considering their "opponents" are part of the noncontrollable consumption. This was the sequential BRD used in this work.

When the forecast was deterministic, the strategies were reduced to be parameter vectors in which each component is a consumption power and the iterative information among consumers was a simple vector known. In such a case, the three sorts of strategies provided in the Chapter 3 are used to solve the problem. However, these were built here by sequential BRD. On the other hand, when the forecast was stochastic, an iterative MDP method was developed, based on the MDP technique of the Chapter 3. The iterative information among consumers was a known vector of random variables. Note that solving problems in which several decision-makers must operate collaboratively in sequential (stochastic) environments is an important challenge. Here we have developed an iterative technique based on MDPs and BRD. However, there is no formal proof of optimality, contrary to what we usually want in formal methods. This remains an open problem.

At the end of this chapter, a numerical analysis came into play to measure the performance of the provided decentralized strategies based on BRD. The scenario of consumers was considered a pool of EVs. We have examined here the technical and economic interactions between EVs and the DN. In particular, first to maximize the DN-transformer lifetime and second, to reduce the total electrical consumption payment. Each application was made independently and no trade-off was analyzed. In line with the Chapter 3, numerical applications showed that the MDP method is almost insensitive to noise and this is globally much more robust than the other methods. Other decentralized techniques were also used to compare the performances of our work. These are: *Gan et al.* [56] and *Shinwari et al.* [106]. However, our MDP methodology remained more robust.

7.2 FUTURE RESEARCH DIRECTIONS

First, in the numerical applications in the Chapter 3 and Chapter 6, we have assumed to have an available forecast on the noncontrollable part of the total load consumption to schedule strategies. A potentially interesting extension could be the study of suitable methods to perform a "correct" forecast of such an environment. This can be based, e.g., on reinforcement learning (see e.g, [108]). For instance, in [91] the probability transitions of the Markov model is not assumed known a priori and the centralized controller learns continuously to adapt the individual consumer preferences and pricing variations over time. Concerning several decision-makers, some overviews of "multi-agent" reinforcement learning algorithms (fully cooperative, fully competitive, etc.) have been reviewed in [35]. However, (as in smart grid) formal methods have only seldomly been attacked in such a field. Some approximations have been studied for example in [55], but only safety guarantees from formal verification have been given.

Second, the main thing to discuss from the decentralized application of the Chapter 6, is the guarantee of the convergence of the method based on iterative MDPs. Even if we have observed such behavior numerically, an interesting extension should be to study a formalism of such a proposed method. This method can be seen as if each decision-maker has its own learning process. Then, they determine their best-response behaviors considering their opponents are part of the environment. From a game theory view, this often results in local optima [33, 66, 81], but (to our knowledge) no formal methods have been developed.

Third, in Chapter 4 we have identified realistic conditions on the structure of the multi-weighted MDP under which our algorithmic technique was almost-complete. The case of MDPs not satisfying these conditions remain open, but we believe that our approximation techniques gives interesting information which suffice for practical applications such as in smart grids. In addition, under one of these conditions mentioned there, we can immediately extend to multi-weights with worst-case constraints. However, in a more general setting where the costs can take any value, this could not be solved, which has to be put in parallel with the undecidability result of [102]. A nice continuation of our work would consist in computing (approximations of) Pareto-optimal solutions in the multi-weighted MDP setting. Improving the complexity and practicality of our approach is also on our agenda for future work.

Fourth, some studies/applications in the area of smart grid (in particular in the EVs-charging problem) that we can focus on the use formal methods in future works, could be included, e.g., to consider that the departure time of EVs is stochastic [87], to assume that the stochastic environment is the wind power generation [67, 80], to define a stochastic game taking into account the dynamics of electrical systems (heating, air conditioning, etc.) and to analyze optimal equilibria [54], to design a model with uncertain energy demands [104], to consider a large population of EVs in a energy game [82], and so on. In addition, it is also possible to focus on learning methods over stochastic environments, as in [39] for the EVs-charging problem with forecasted prices, or as in [13] in which a learning algorithm about the real-time energy price was proposed, etc. Where again, formal methods need to be applied to guarantee correctness.

Part V

APPENDIX

This Part includes the synthesis in French of this work and the mathematical proofs of our research.

A

SYNTHÈSE EN FRANÇAIS

1.1 INTRODUCTION

Au sein de la communauté scientifique, l'étude des réseaux d'énergie ou « smart grids » suscite un vif intérêt puisque ces infrastructures deviennent de plus en plus importantes dans notre monde moderne [88]. Des outils mathématiques avancés et complexes sont nécessaires afin de bien concevoir et mettre en œuvre ces réseaux, dont la précision et l'optimalité sont deux caractéristiques essentielles pour leur conception [65]. Cela motive fortement le travail développé dans cette thèse.

A.1.1 *INSPIRATION ET CONTEXTE*

Dans cette thèse, nous étudions le problème général de planification de consommation d'énergie. En termes simples, il s'agit d'un scénario dans lequel les consommateurs ont besoin d'une certaine quantité d'énergie et souhaitent que cette demande soit satisfaite dans une période spécifique (e.g., un Véhicule Électrique (VE) doit être rechargé dans une fenêtre de temps définie par son propriétaire). Par conséquent, chaque consommateur doit choisir (par un système informatisé) une puissance de consommation à chaque instant afin que l'énergie finale accumulée atteigne un niveau souhaité. La manière dont les puissances de consommation sont choisies est obtenue par l'application d'une « stratégie » qui prend en compte à chaque instant les informations pertinentes d'un consommateur (e.g., l'énergie accumulée pour recharge le VE) afin de choisir un niveau de consommation approprié.

Nous nous concentrons principalement sur la synthèses des stratégies appropriées et optimales pour réduire l'impact des consommateurs sur le réseau d'énergie et qui sont confrontés aux incertitudes de l'environnement du système électrique, à savoir: la part non-contrôlable de la consommation électrique totale (e.g., la consommation hors VEs dans le quartier résidentiel). Nous étudions les cas où cette consommation non-contrôlable est déterministe ou stochastique. En synthétisant des stratégies fonctionnellement correctes, nous pouvons donner de garanties de précision du système. En particulier, la consommation contrôlable peut s'adapter aux contraintes du réseau d'énergie (e.g., pour ne pas dépasser la température maximale d'arrêt du transformateur électrique) et aux objectifs des consommateurs (e.g., tous les VEs soient rechargés en minimisant le coût total de la consommation). Dans ce chapitre, nous analysons une procédure informatique pour concevoir des stratégies selon une approche centralisée (dans laquelle il n'y a qu'un seul décideur qui contrôle toutes les stratégies des consommateurs)[1] en utilisant des méthodes formelles [48], la théorie des jeux [84] et l'optimisation [30].

1 Pour le cas décentralisé (dans laquelle il y a plusieurs contrôleurs, chacun représentant un consommateur) voir le Chapitre 6.

Nous étudions aussi la généralisation théorique de ce problème d'énergie, qui peut être vu comme un système stochastique avec des multi-objectifs sous contraintes. En particulier, nous pouvons exprimer une limite supérieure de la consommation totale sur le réseau (codée comme une condition d'accessibilité dans notre modèle), une contrainte imposée à la demande d'énergie des consommateurs en tant qu'objectif quantitatif d'accessibilité, et divers critères d'optimisation des coûts (e.g., minimiser le vieillissement du transformateur électrique). Nous caractérisons l'existence d'une stratégie à travers des problèmes d'optimisation sous-jacents qui sont résolus à l'aide des problèmes duales associés, et montrons comment construire une telle stratégie. Comme les valeurs optimales des problèmes initial et dual sont différentes pour des stratégies pures (écart de dualité), nous étendons ces problèmes aux stratégies mixtes et prouvons qu'il n'y a pas d'écart de dualité dans un tel contexte. Nous montrons enfin qu'on peut construire une solution (stratégie) mixte à travers d'au plus deux stratégies pures et développons an algorithme pour telle construction.

1.2 PROBLÈME D'ÉNERGIE

Nous considérons un Réseau de Distribution (RD) comprenant un transformateur auquel deux groupes sont connectés : un ensemble d'appareils électriques contrôlables ou de consommateurs (e.g., des Véhicules Électriques (VEs)), dénoté par $\mathcal{I} := \{1, \ldots, I\}$, $I \in \mathbb{N}$, et un ensemble d'autres appareils électriques. Ce dernier groupe est supposé induire une consommation indépendante de $\mathcal{I}$ et est donc appelée consommation non-contrôlable. En supposant que le temps est discrétisé avec un pas de temps Δ_t et indexé par $t \in \mathcal{T} := \{0, 1, \ldots, T\}$, $T \in \mathbb{N}$, la **consommation totale** à t peut s'écrire de la manière suivante :

E.g., si chaque t représente un créneau de $30\,\mathrm{min}$, alors $\Delta_t = 0.5\,\mathrm{h}$

$$\ell_t \;:=\; \ell_{0,t} + \sum_{i \in \mathcal{I}} \ell_{i,t} \;\leq\; \ell^{\max} , \tag{A.1}$$

où $\ell^{\max} \in \mathbb{R}_0^+$ est la puissance maximale admissible du transformateur du RD, $\ell_{0,t} \in \mathbb{R}_0^+$ est la **consommation non-contrôlable réelle** (inconnue à priori) et $\ell_{i,t} \in \mathbb{R}_0^+$ est la **consommation contrôlable** du consommateur $i \in \mathcal{I}$, telle que :

La puissance minimale fait référence, e.g., lorsque i consomme en mode veille

$$\ell_i^{\min} \;\leq\; \ell_{i,t} \;\leq\; \ell_i^{\max} , \tag{A.2}$$

où $\ell_i^{\min}, \ell_i^{\max} \in \mathbb{R}_0^+$ sont resp. la puissance de **consommation minimale et maximale**. Une contrainte de **besoin d'énergie** $e_i \in \mathbb{R}_0^+$ est considérée pour chaque $i \in \mathcal{I}$:

E.g., $e_i = 24\,\mathrm{kWh}$ correspond à une charge complète d'un VE i

$$\Delta_t \sum_{t \in \mathcal{T}} \ell_{i,t} \;\geq\; e_i . \tag{A.3}$$

Pour le système d'intérêt (le RD), l'état est dénoté à $t \in \mathcal{T}$ par $x_t \in \mathbb{R}$, et il est requis de maintenir x_t borné supérieurement par un seuil donné $x^{\max} \in \mathbb{R}_0^+$, c.-à-d.,

E.g., si x_t représente la température du transformateur du RD, x_t doit rester au plus égale à la température maximale $x^{\max}$

$$x_t \;\leq\; x^{\max} . \tag{A.4}$$

Un modèle général (non linéaire) est supposé pour la **dynamique de l'état du système**. Il est exprimé par :

$$x_{t+1} \;=\; f(x_t, \ell_t, \ell_{0,t}) \tag{A.5}$$

pour chaque $t \in \mathcal{T}$, où $x_1 \in \mathbb{R}_0^+$ est une condition initiale donnée de l'état du système et f est une fonction connue. Une hypothèse pratique pour les algorithmes étudiés dans cette thèse, est la suivante.

Hypothèse A.2.1

La fonction f de la dynamique de l'état du système (A.5), est une fonction qui dépend de la consommation totale (A.1).

STRATÉGIES

La manière dont la consommation contrôlable d'un consommateur est choisie, est obtenue par l'application d'une stratégie à chaque instant t. Les stratégies sont conçues dans cet Appendix A selon une approche centralisée, c.-à-d., qu'il n'y a qu'un seul décideur (système informatisé centralisé) qui contrôle toutes les stratégies des consommateurs. L'information qu'il prend en compte pour définir les stratégies est l'historique (également appelé chemin) du système. Un **chemin du système** est représenté à t par :

$$\omega_t \quad := \quad (x_1, \boldsymbol{\ell}_1, x_2, \boldsymbol{\ell}_2, \ldots, x_{t-1}, \boldsymbol{\ell}_{t-1}, x_t) \ , \tag{A.6}$$

où $\boldsymbol{\ell}_t := (\ell_{i,t})_{i \in \mathcal{I}}$ est le **vecteur de consommation contrôlable** des consommateurs à t. Une **stratégie** (centralisée) pour les consommateurs est définie à l'instant t par :

$$\boldsymbol{\pi}_t(\omega_t) \quad = \quad \boldsymbol{\ell}_t \ ,$$

et nous dénotons un profil des stratégies par $\boldsymbol{\pi} := (\boldsymbol{\pi}_t)_{t \in \mathcal{T}}$. Notez que cette définition de stratégie est conforme à celle d'une stratégie pure vue dans la Section 2.6.3. Rappelons que, dans ce contexte, $\boldsymbol{\pi}$ est **sans mémoire** ou de **Markov** si pour chaque $t \in \mathcal{T}$,

$$\boldsymbol{\pi}_t(\omega_t) \quad = \quad \boldsymbol{\pi}_t(\omega'_{t'}) \ , \tag{A.7}$$

où $\omega_t = (x_1, \boldsymbol{\ell}_1, x_2, ..., \boldsymbol{\ell}_{t-1}, x_t)$ et $\omega'_{t'} = (x_1, \boldsymbol{\ell}'_1, x'_2, ..., \boldsymbol{\ell}'_{t'-1}, x'_{t'})$ sont des chemins finis du système, tels que $x_t = x'_{t'}$. Autrement dit, la seule information pertinente qu'une stratégie de Markov a besoin est contenu dans l'état actuel du système.

PROBLÈME D'ÉNERGIE

Le problème d'intérêt de planification de la consommation d'énergie pour les consommateurs est formulé comme suit :

Problème de Planification de Consommation d'Énergie

$$\min_{\boldsymbol{\pi}} \quad \sum_{t \in \mathcal{T}} C^{\boldsymbol{\pi}}_{x_1}(x_t, \boldsymbol{\ell}_t, \ell_{0,t}) \tag{A.8}$$
$$\text{s.t.} \quad (\text{A.1}),\ (\text{A.2}),\ (\text{A.3}),\ (\text{A.4}),\ (\text{A.5}) \ .$$

Pour résoudre ce problème, nous supposons qu'une prévision de la consommation non-contrôlable réelle $(\ell_{0,t})_{t \in \mathcal{T}}$ est disponible, qui peut être déterministe ou stochastique (voir Section 2.4 pour plus de détails). De cette manière, un profil des stratégies optimales $\boldsymbol{\pi}^*$ est construit hors-ligne pour définir la consommation contrôlable optimale qui sera exécutée. Une fois que $\boldsymbol{\pi}^*$ est définie, il peut être mise en ligne.

A.2.1 *MÉTHODOLOGIE POUR LE CAS DÉTERMINISTE*

Dans cette section, une **prévision déterministe** de la consommation non-contrôlable réelle sera disponible, laquelle intervient en tant qu'un profil de paramètres dans le problème (A.8). Cette prévision sera dénotée par $(\widetilde{\ell}_{0,t})_{t\in\mathcal{T}}$. Tout le problème est donc de choisir la suite des niveaux de puissances de consommation $(\ell_{i,t})_{t\in\mathcal{T}}$ pour chaque consommateur $i \in \mathcal{I}$ en résolvant (A.8). Nous proposons trois types de stratégies de consommation dans ce qui suit.

PROFILS DE CONSOMMATION RECTANGULAIRES

Pour chaque consommateur $i \in \mathcal{I}$, nous supposons, premièrement, que la consommation $\ell_{i,t}$ à $t \in \mathcal{T}$ ne peut que prendre les valeurs $\ell_i^{\min}$ ou $\ell_i^{\max}$, et, deuxièmement, que lorsque $\ell_{i,t} = \ell_i^{\max}$, la consommation doit être ininterrompue jusqu'à ce que la contrainte de besoin d'énergie (A.3) soit satisfaite. Cela conduit à des profils de consommation rectangulaires que nous dénotons par :

E.g., si pour i, $e_i = 6\,\mathrm{kWh}$, $\ell_i^{\min} = 0\,\mathrm{kW}$, $\ell_i^{\max} = 3\,\mathrm{kW}$, $T = 5$ et $\Delta_t = 1$, alors $\boldsymbol{\ell}_i = (0,0,3,3,0)$ est un profil rectangulaire pour i

$$\boldsymbol{\ell}_i \in \left\{ (\ell_{i,t})_{t\in\mathcal{T}} \in \{\ell_i^{\min},\, \ell_i^{\max}\}^T \;\middle|\; \begin{array}{l} \forall t \in \{t_i^{\text{start}}, \dots, t_i^{\text{stop}}\} \subseteq \mathcal{T},\ \ell_{i,t} = \ell_i^{\max} \\ \forall t \in \mathcal{T} \setminus \{t_i^{\text{start}}, \dots, t_i^{\text{stop}}\}\ ,\ \ell_{i,t} = \ell_i^{\min} \end{array} \right\} \tag{A.9}$$

où $t_i^{\text{start}}, t_i^{\text{stop}} \in \mathcal{T}$ représentent resp. le temps de début et d'arrêt de la consommation, tels que $t_i^{\text{start}} \le t_i^{\text{stop}}$. Par conséquent, une stratégie de planification de la consommation d'énergie est réduite à choisir t_i^{start}. En pratique, à partir de la contrainte de besoin d'énergie (A.3), chaque $t_i^{\text{start}} \in \mathcal{T}$ est limité à être :

Ici, nous ne prendrons pas en compte la consommation due à la puissance minimale $\ell_i^{\min}$

$$t_i^{\text{start}} \quad \le \quad T \; - \; \frac{e_i}{\ell_i^{\max}} \ , \tag{A.10}$$

où t_i^{stop} peut être choisi comme le temps minimum tel que :

$$(t_i^{\text{stop}} \; - \; t_i^{\text{start}})\, \ell_i^{\max} \quad \ge \quad e_i \ . \tag{A.11}$$

Avec un petit abus de notations, nous écrivons une stratégie de profils rectangulaires par $\boldsymbol{\pi}^{\text{RP}} = (\, t_1^{\text{start}},\ \dots,\ t_I^{\text{start}}\,)$. Ainsi, le problème (A.8) peut s'écrire dans ce contexte comme :

Problème d'Énergie - Profils Rectangulaires

$$\min_{\boldsymbol{\pi}^{\text{RP}}} \quad \sum_{t\in\mathcal{T}} C_{x_1}^{\boldsymbol{\pi}^{\text{RP}}}(\widetilde{x}_t,\, \boldsymbol{\ell}_t\,;\, \widetilde{\ell}_{0,t}) \tag{A.12}$$
$$\text{s.t.} \quad (\text{A.1}),\ (\text{A.4}),\ (\text{A.5}),\ (\text{A.9}),\ (\text{A.10}),\ (\text{A.11})\ .$$

Puisque l'état du système est complètement déterminé sous l'hypothèse d'une prévision déterministe, le problème (A.12) est réduit à un problème d'optimisation.

CONSOMMATION DYNAMIQUE

Cette méthode prend explicitement l'évolution de l'état du système (A.5), laquelle peut être exprimée comme une seule fonction dépendant de la condition initiale x_1, la prévision déterministe $(\widetilde{\ell}_{0,t})_{t\in\mathcal{T}}$ et la suite $(\boldsymbol{\ell}_1, \ldots, \boldsymbol{\ell}_t)$ pour chaque $t \in \mathcal{T}$. Cette observation nous permet de convertir le problème (A.8) en un problème d'optimisation standard [30] en définissant des fonctions $(g_t)_{t\in\mathcal{T}}$, telles que $g_1(x_1) := x_1$ et

$$\begin{aligned} & g_{t+1}(x_1, \boldsymbol{\ell}_1, \ldots, \boldsymbol{\ell}_t\,;\, \widetilde{\ell}_{0,1}, \ldots, \widetilde{\ell}_{0,t}) \\ & \qquad := \; f\big(g_t(x_1, \boldsymbol{\ell}_1, \ldots, \boldsymbol{\ell}_{t-1}\,;\, \widetilde{\ell}_{0,1}, \ldots, \widetilde{\ell}_{0,t-1}),\, \boldsymbol{\ell}_t;\, \widetilde{\ell}_{0,t}\big)\;. \end{aligned} \tag{A.13}$$

De cette façon, la contrainte (A.4) peut s'écrire dans ce contexte comme :

$$g_t(x_1, \boldsymbol{\ell}_1, \ldots, \boldsymbol{\ell}_{t-1}\,;\, \widetilde{\ell}_{0,1}, \ldots, \widetilde{\ell}_{0,t-1}) \;\leq\; x^{\max}\;. \tag{A.14}$$

Tout le problème est donc de trouver une stratégie de consommation dynamique (dénotée ici par $\boldsymbol{\pi}^{\mathrm{DC}}$) en résolvant le problème suivant d'optimisation standard :

Problème d'Énergie - Consommation Dynamique

$$\begin{aligned} \min_{\boldsymbol{\pi}^{\mathrm{DC}}} \; & \sum_{t\in\mathcal{T}} C_{x_1}^{\boldsymbol{\pi}^{\mathrm{DC}}}\Big(g_t(x_1, \boldsymbol{\ell}_1, \ldots, \boldsymbol{\ell}_{t-1}\,;\, \widetilde{\ell}_{0,1}, \ldots, \widetilde{\ell}_{0,t-1}),\, \boldsymbol{\ell}_t\,;\, \widetilde{\ell}_{0,t}\Big) \\ & \text{s.t.} \;\; (A.1),\; (A.2),\; (A.3),\; (A.13),\; (A.14)\;. \end{aligned} \tag{A.15}$$

Ce problème est un problème d'optimisation convexe si $C_{x_1}, g_1, \ldots, g_t$ et f sont convexes

MÉTHODE DE VALLEY-FILLING

L'idée de l'algorithme de valley-filling est de positionner le vecteur de la consommation contrôlable agrégée, dénoté par :

$$\boldsymbol{\pi}^{\mathrm{VF}} \;=\; \left(\sum_{i\in\mathcal{I}} \ell_{i,1},\; \ldots,\; \sum_{i\in\mathcal{I}} \ell_{i,T}\right)\;,$$

afin de remplir les vallées de la courbe dessinée par la consommation non-contrôlable $(\widetilde{\ell}_{0,t})_{t\in\mathcal{T}}$. Puisque cette méthode contrôle à chaque instant t la somme des consommations contrôlables, nous pouvons réécrire la contrainte de la demande d'énergie (A.3) par :

$$\Delta_t \sum_{t\in\mathcal{T}}\left(\sum_{i\in\mathcal{I}} \ell_{i,t}\right) \;\geq\; \sum_{i\in\mathcal{I}} e_i\;. \tag{A.16}$$

En suivant la définition classique de cet algorithme (voir par exemple [94, 106]), nous écrivons le problème (A.8) dans ce contexte par :

Problème d'Énergie - Valley-Filling

$$\begin{aligned} \max_{\boldsymbol{\pi}^{\mathrm{VF}}} \; & \sum_{t\in\mathcal{T}} \Phi^{\boldsymbol{\pi}^{\mathrm{VF}}}\left(\widetilde{\ell}_{0,t} + \sum_{i\in\mathcal{I}} \ell_{i,t}\right) \\ & \text{s.t.} \;\; (A.2),\; (A.16)\;. \end{aligned} \tag{A.17}$$

Ici, Φ est une fonction strictement convexe

La solution de ce problème est donnée explicitement en utilisant la méthode des multiplicateurs de Lagrange [24, 30] :

$$\sum_{i\in\mathcal{I}} \ell_{i,t} \;=\; \min\left\{\sum_{i\in\mathcal{I}} \ell_i^{\max},\ \max\left\{\sum_{i\in\mathcal{I}} \ell_i^{\min},\ \mu - \widetilde{\ell}_{0,t}\right\}\right\} ,$$

où $\mu \in \mathbb{R}_0^+$ est un seuil (le niveau de remplissage par « l'eau ») à définir pour que la contrainte de besoin d'énergie (A.16) soit vérifiée.

A.2.2 *MÉTHODOLOGIE POUR LE CAS STOCHASTIQUE*

Les méthodes précédentes ont été conçues en supposant une prévision déterministe de la consommation non-contrôlable, (laquelle représente un simple scénario donné par un vecteur de paramètres). Dans cette section, une **prévision stochastique** est disponible pour résoudre le problème (A.8). Cette prévision sera dénotée par $(\widetilde{L}_{0,t})_{t\in\mathcal{T}}$, qui est une collection finie de variables aléatoires i.i.d. définies à travers d'une distribution de probabilité P_0.

Supposons que $\boldsymbol{L}_t = (L_{i,t})_{i\in\mathcal{I}}$ est la sequence qui prend le vecteur de consommation contrôlable $\boldsymbol{\ell}_t = (\ell_{i,t})_{i\in\mathcal{I}}$ à $t \in \mathcal{T}$. Nous écrivons (de manière stochastique) la consommation totale (A.1) comme suit :

$$\widetilde{L}_t \;=\; \widetilde{L}_{0,t} + \sum_{i\in\mathcal{I}} L_{i,t} \, . \tag{A.18}$$

Ainsi, le comportement de l'état du système est également stochastique et nous le dénotons par $\widetilde{X}_t$. Supposons maintenant que l'état du système à t est x_t et qu'un vecteur de consommation contrôlable $\boldsymbol{\ell}_t$ est sélectionné. La probabilité que l'état du système soit x_{t+1} à l'instant $t+1$ (suivant l'évolution f dans (A.5)) est calculée comme :

$$\begin{aligned} P\Big[\widetilde{X}_{t+1} = \widetilde{x}_{t+1} \;\Big|\; \widetilde{X}_t = \widetilde{x}_t,\, \boldsymbol{L}_t = \boldsymbol{\ell}_t\Big] & \\ = P_0\Big[\,\widetilde{L}_{0,t} \in \{\,\widetilde{\ell}_{0,t} \,|\, \widetilde{x}_{t+1} = f(\widetilde{x}_t, \boldsymbol{\ell}_t; \widetilde{\ell}_{0,t})\} \;\Big|\; \widetilde{X}_t = \widetilde{x}_t,\ \boldsymbol{L}_t = \boldsymbol{\ell}_t\Big] \, . \end{aligned} \tag{A.19}$$

La contrainte de la puissance maximale $\ell^{\max}$ admissible du transformateur du RD et de la borne supérieur $x^{\max}$ de la valeur de l'état du système, peuvent s'écrire maintenant sur $\mathcal{T}$ comme suit :

$$\prod_{t\in\mathcal{T}} P_0\Big[\,\widetilde{L}_t \;\le\; \ell^{\max} \;\Big|\; \boldsymbol{L}_t = \boldsymbol{\ell}_t\Big] \;\ge\; 1 - \varepsilon_\ell \tag{A.20}$$

$$\prod_{t\in\mathcal{T}} P\Big[\widetilde{X}_{t+1} \;\le\; x^{\max} \;\Big|\; \widetilde{X}_t = \widetilde{x}_t,\, \boldsymbol{L}_t = \boldsymbol{\ell}_t\Big] \;\ge\; 1 - \varepsilon_x \tag{A.21}$$

où $(\varepsilon_\ell, \varepsilon_x) \in [0,1]^2$ représente le risque en probabilité de dépasser resp. les valeurs $\ell^{\max}$ et $x^{\max}$. En réécrivant la contrainte de la demande d'énergie (A.3) sous la forme :

$$\Delta_t \sum_{t\in\mathcal{T}} L_{i,t} \;\ge\; e_i \, , \tag{A.22}$$

et la contrainte sur la consommation contrôlable (A.2) par :

$$\ell_i^{\min} \;\le\; L_{i,t} \;\le\; \ell_i^{\max} \, , \tag{A.23}$$

le problème (A.8) peut être réécrit dans ce contexte comme suit :

Problème d'Énergie - Cas Stochastique

$$\min_{\pi} \quad \sum_{t \in \mathcal{T}} \mathbb{E}^{\pi}_{x_1} \left[C^{\pi}(\widetilde{X}_t,\, \boldsymbol{L}_t\,;\, \widetilde{L}_{0,t}) \right] \tag{A.24}$$

$$\text{s.t.} \quad (\text{A.18}),\ (\text{A.19}),\ (\text{A.20}),\ (\text{A.21}),\ (\text{A.22}),\ (\text{A.23})\ .$$

Pour résoudre ce problème et faire un calcul effectif, une discrétisation sur la consommation sera considérée. Ce problème peut donc être modélisé à l'aide d'un processus de décision de Markov [96] lorsque des statistiques (appropriées) sont disponibles.

APPROCHE PAR UN PROCESSUS DE DÉCISION DE MARKOV

D'après la Définition 2.6.9, nous utilisons un Processus de Décision de Markov (MDP) double-pondéré dans ce contexte. Il est de la forme :

$$\mathcal{M} \quad = \quad \Big(\mathcal{X} \times \mathcal{L},\, (x_1, 0),\, \mathcal{A},\, P,\, C,\, C_{\mathcal{I}} \Big)\ .$$

où $\mathcal{X} \times \mathcal{L}$ définit l'espace des états, $\mathcal{X}$ est l'ensemble fini d'états du système du RD, $\mathcal{L}$ est l'ensemble (fini) représentant les valeurs de la consommation totale (A.18), $(x_1, 0)$ est l'état initial, $\mathcal{A}$ est l'espace fini d'actions (c.-à-d., de la consommation contrôlable des consommateurs), P est la probabilité de transition entre les états, C es une fonction de coût entre les transitions des états (celle du problème (A.24)), et $C_{\mathcal{I}}$ est une autre fonction de coût définie comme l'énergie agrégée :

$$C_{\mathcal{I}}(\,\boldsymbol{L}_t\,) \quad = \quad \Delta_t \sum_{i \in \mathcal{I}} L_{i,t}\ .$$

Dans ce qui suit, nous déplions $\mathcal{M}$ afin de simplifier la contrainte de la demande d'énergie des consommateurs dans un objectif quantitatif d'accessibilité. D'après la Définition 2.6.10, le dépliage de $\mathcal{M}$ est la structure suivante :

$$\mathcal{M}_T \quad = \quad \Big(\mathcal{S}_T,\, s_1,\, \mathcal{A},\, P_T,\, \boldsymbol{C} \Big)\ , \tag{A.25}$$

où $\mathcal{S}_T = \mathcal{X} \times \mathcal{L} \times \mathcal{E} \times \mathcal{T}$ est l'espace des états, où $\mathcal{E}$ représente l'espace de l'énergie accumulée des consommateurs :

$$\mathcal{E} \quad = \quad \left[0,\, T \sum_{i \in \mathcal{I}} \ell_i^{\max} \right]\ ,$$

$s_1 = (x_1, 0, 0, 1)$ est l'état initial, $\mathcal{A}$ est l'espace des actions défini par :

$$\mathcal{A} \quad = \quad \bigcup_{t \in \mathcal{T}} \left\{ \ell_{\mathcal{I},t} \;=\; \sum_{i \in \mathcal{I}} \ell_{i,t} \;\Big|\; \sum_{i \in \mathcal{I}} \ell_i^{\min} \;\leq\; \ell_{\mathcal{I},t} \;\leq\; \sum_{i \in \mathcal{I}} \ell_i^{\max} \right\}\ ,$$

$P_T : \mathcal{S}_T \times \mathcal{A} \to \mathcal{D}(\mathcal{S}_T)$ est la probabilité de transition définie par :

$$P_T(s_t, \ell_{\mathcal{I},t})(s_{t+1}) \quad = \quad P(\mathrm{proj}_1(s_t), \ell_{\mathcal{I},t})(\mathrm{proj}_1(s_{t+1}))$$

Ici, proj_j *est la* $j^{\text{ème}}$*-projection sur les états,* $j = 1, 2, 3, 4$

$$\begin{aligned}
\text{si}\ (i) \quad & \mathrm{proj}_1(s_{t+1}) \;=\; f(\mathrm{proj}_1(s_t),\, \ell_{\mathcal{I},t}\,;\, \widetilde{\ell}_{0,t})\,,\\
(ii) \quad & \mathrm{proj}_2(s_{t+1}) \;=\; \widetilde{\ell}_{0,t} + \ell_{\mathcal{I},t}\,,\\
(iii) \quad & \mathrm{proj}_3(s_{t+1}) \;=\; \min\{\, e_{\mathcal{I}},\, \mathrm{proj}_3(s_t) + \Delta_t\, \ell_{\mathcal{I},t}\,\}\,, \text{ où}\\
& e_{\mathcal{I}} \;=\; \sum_{i \in \mathcal{I}} e_i\,, \qquad\qquad (A.26)\\
(iv) \quad & \mathrm{proj}_4(s_{t+1}) \;\le\; T,
\end{aligned}$$

et $P_T(s_t, \ell_{\mathcal{I},t})(s_{t+1}) = 0$ dans le cas contraire. Notez que (i) est implicitement inclus dans (A.19). Enfin, la fonction de coût $\boldsymbol{C}$ dans $\mathcal{M}_T$ est la même que C de $\mathcal{M}$.

Puisque nous sommes intéressés à atteindre la énergie accumulée (A.26) pour satisfaire la demande des consommateurs, nous définissons un ensemble objectif $\mathcal{G} \subset \mathcal{S}_T$ comme suit :

$$\mathcal{G} \;:=\; \Big\{ s_t \in \mathcal{S}_T \;\mid\; \mathrm{proj}_3(s_t) = e_{\mathcal{I}} \;\text{ and }\; \mathrm{proj}_4(s_t) = T \Big\}\,.$$

Nous voudrons trouver une stratégie de sorte que $\mathcal{G}$ est atteint. Cela est réduit au problème du plus court chemin, pour lequel l'existence et la construction d'une stratégie optimale peut être décidée en temps POLYNOMIAL, voir par exemple [14, 25].

A.2.3 APPLICATION NUMÉRIQUE

Nous nous plaçons dans un RD alimenté par un transformateur moyenne-tension/basse-tension 20 kV / 400 V de puissance maximale $\ell^{\max} = 90\,\mathrm{kW}$ dans un quartier résidentielle qui contient une trentaine de foyers. Nous assumons que l'état du RD (A.5) évolue avec la dynamique suivante :

$$x_{t+1} \;=\; \alpha x_t + \beta \Big(\ell_{0,t} + \sum_{i \in \mathcal{I}} \ell_{i,t} \Big)^2 + \gamma \Big(\ell_{0,t-1} + \sum_{i \in \mathcal{I}} \ell_{i,t-1} \Big)^2 + z_t\,.$$

L'état du système représente la température du point chaud du transformateur du RD, dont la température nominale est $x_1 = 98\,°\mathrm{C}$ (température de référence) et la température maximale d'arrêt (shut-down) est $x^{\max} = 150\,°\mathrm{C}$. Les paramètres du modèle précédent sont [70] : $\alpha = 0.83$, $\beta = 30.91\,°\mathrm{C.kW}^{-2}$, $\gamma = -19.09\,°\mathrm{C.kW}^{-2}$ et z_t est définie par $z_t = 0.17\,(8.47 + x_t^{\mathrm{amb}})$, où x_t^{amb} est la température ambiante à t.

Des données correspondant à $(x_t^{\mathrm{amb}})_{t \in \mathcal{T}}$ *sont obtenues de [5]*

DURÉE DE VIE DU TRANSFORMATEUR

Le transformateur a été dimensionné de telle sorte qu'il vive 40 ans sans la consommation contrôlable. Le temps de vie du transformateur est défini inversement proportionnel au vieillissement total :

$$\text{Lifetime} \;:=\; T\, \frac{40}{\sum_{t \in \mathcal{T}} C_{x_1}^{\pi}(x_t, \boldsymbol{\ell}_t, \ell_{0,t})}\,,$$

où la fonction de coût C est considérée comme le modèle du vieillissement instantané. Un modèle bien admis qui mesure la vitesse de dégradation par rapport à la température nominale du transformateur est le suivant [68] : $C_{x_1}^{\pi}(x_t) = \exp(a\, x_t + b)$, où $a = 0.12\,°\mathrm{C}^{-1}$ et $b = -11.32$.

CONSOMMATION CONTRÔLABLE

Les simulations sont effectuées sur un temps discrétisé, où chaque t correspond un créneau de $30\,\text{min}$, de sorte que $\Delta_t = 0.5\,\text{h}$. Les opérations de la consommation contrôlable ont lieu dans une fenêtre de temps de $17\,\text{h}$ à $7\,\text{h}$ du lendemain matin. Le reste de la journée (de $7\,\text{h}$ à $17\,\text{h}$ du lendemain matin), la consommation totale sur le RD est constitué uniquement de la consommation non-contrôlable. Ainsi, $T = 48$. Les consommateurs sont considérés ici comme un ensemble de Véhicules Électriques (VEs), de sorte que chaque $i \in \mathcal{I}$ représente un VE. Les puissances de charge minimale et maximale induites par un VE i sont resp. $\ell_i^{\min} = 0\,\text{kW}$ et $\ell_i^{\max} = 3\,\text{kW}$. Les temps d'arrivée et de départ des VEs sont choisis de manière aléatoire selon une distribution normal $\mathcal{N}(4, 1.5)$ et $\mathcal{N}(28, 0.75)$ respectivement. La demande d'énergie de chaque VE est obtenue à partir de la distance à parcourir pour chacun d'eux lors du prochain voyage :

Ceci correspond à une charge résidentielle typique

$$e_i = \lambda_i \frac{24\,\text{kWh}}{150\,\text{km}} (29.4 + 8)\,\text{km} ,$$

De cette manière, e_i représente entre un $40\,\%$ à $80\,\%$ de la capacité d'un Renault Zoe, comme dans [82]

où $24\,\text{kWh}$ correspond à une charge complète de la batterie d'un VE (e.g., d'un Renault Zoe), $150\,\text{km}$ est la distance parcourue (en moyenne), $(29.4 + 8)\,\text{km}$ est la distance journalière moyenne parcourue ($29.4\,\text{km}$ pour faire un aller-retour au travail par exemple, et $8\,\text{km}$ pour un autre voyage), et λ_i est une marge prise par les consommateurs pour être sûr de ne pas manquer d'énergie lors de la conduite. Ce dernier est généré aléatoirement entre $\{1.5,\ 2,\ 2.5,\ 3\}$.

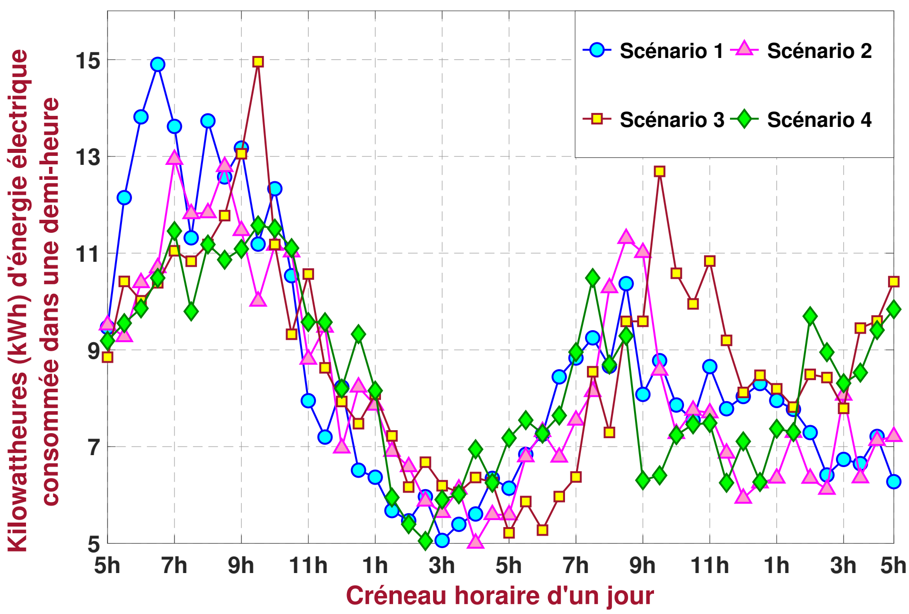


Figure A.1: Quatre scénarios de la consommation non-contrôlable sur un jour.

CONSOMMATION NON-CONTRÔLABLE

Pour représenter le profil de la consommation non-contrôlable réelle $(\ell_{0,t})_{t\in\mathcal{T}}$, quatre scénarios sont générés. Premièrement, nous utilisons une base de données de consommation historique extraite de l'opérateur du réseau d'électricité australien « Ausgrid » de la ville de Sydney [3], dans laquelle nous avons choisi au hasard 30 ménages pour représenter un district. Deuxièmement, nous utilisons une méthode de classification soustractive [40] avec une plage d'influence de 0.9 pour générer quatre classes représentatives, où chacune représentant un scénario de $(\ell_{0,t})_{t\in\mathcal{T}}$, e.g., pour une saison de l'année. La Figure A.1 montre une représentation graphique des quatre scénarios générés. Un élément important est de savoir comment évaluer l'impact de prévision de la consommation non-contrôlable réelle, ici sur chaque scénario $(\ell^k_{0,t})_{t\in\mathcal{T}}$, $k = 1, 2, 3, 4$. Dans cette section, nous considérons qu'une prévision sur un scénario peut être soit déterministe soit stochastique. Lorsqu'elle est déterministe (resp. stochastique), elle est un vecteur avec un bruit (resp. une variable aléatoire). Dans les deux cas, le bruit de la prévision est supposé être un bruit additif $\mathcal{N}(0, \sigma_k^2)$, où la variance σ_k^2 est connue. Ainsi, les modèles (2.4) et (2.6) seront utilisés. Considérant qu'un scénario k est fixé, la variance du bruit est obtenue via le « forecasting signal-to-noise ratio » (SNR) exprimé en décibel (dB), ce qui permet de mesurer dans quelle mesure la consommation non-contrôlable réelle du scénario k peut être prédite. Un SNR est défini par :

$$\mathrm{SNR}_k \quad := \quad 10 \log_{10} \left(\frac{1}{T\sigma_k^2} \sum_{t\in\mathcal{T}} (\ell^k_{0,t})^2 \right).$$

Lorsque l'approche stochastique est adoptée, une discrétisation est considérée pour modéliser le problème dans un MDP $\mathcal{M}_T$ (A.25), où un histogramme normalisé est utilisé avec un paramètre $b = 0.05$ pour ajuster la largeur des classes générés. L'espace des actions $\mathcal{A}$ de $\mathcal{M}_T$ est discrétisé selon les puissances de charge minimale et maximale, de sorte que $\mathcal{A} = \{0, \Delta_\mathcal{A}, ..., I\Delta_\mathcal{A}\}$, où $\Delta_\mathcal{A} = 3$.

PERFORMANCE DES STRATÉGIES DE CONSOMMATION

Nous présentons dans la Figure A.2 et la Figure A.3 les performances des différents stratégies de consommation d'énergie pour la recharge des VEs, obtenus sur du vieillissement du transformateur (équipement coûteux du RD). Premièrement, la Figure A.2 montre que la méthode classique de Brancher-et-Charger a un impact très significatif sur le transformateur. Cette stratégie n'est donc pas acceptable. Dans cette Figure A.2, la prévision sur la consommation hors VEs est supposée avoir un bruit basé sur un $\mathrm{SNR} = 15\,\mathrm{dB}$ (hypothèse ambitieuse de prévision presque-parfaite). La stratégie construite à l'aide d'un MDP a une performance similaire à celle de consommation dynamique, cette dernière est optimale lorsque le bruit de prévision est proche de zero (SNR grand). Les autres deux stratégies ont de performances proches à celle du MDP, cependant puisque la stratégie des profils rectangulaires est moins flexible (car la consommation est ininterrompue), elle perd un peu en performance. Deuxièmement, pour des cas plus réalistes de prévision, nous observons dans la Figure A.3 que la stratégie construite par la méthode MDP est beaucoup plus robuste que les autres méthodes considérées et presque insensible aux bruits de prévision. Bien que les autres stratégies s'améliorent lorsque la prévisions est de meilleure qualité (SNR augmente), leurs performances diminuent et sont considérablement « chaotiques» pour des valeurs inférieures à SNR $= 6$ dB.

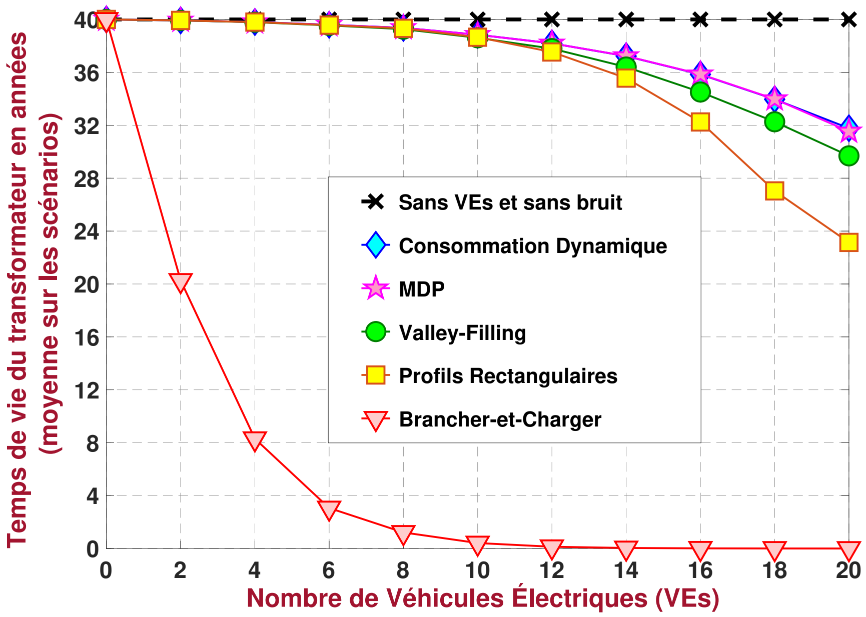


Figure A.2: Temps de vie du transformateur du RD en fonction du nombre de Véhicules Électriques (I), sous l'hypothèse d'une prévision imparfaite sur la consommation non-contrôlable réelle (dans chaque scénario) basée sur un SNR $= 15$ dB. (Cas centralisé).

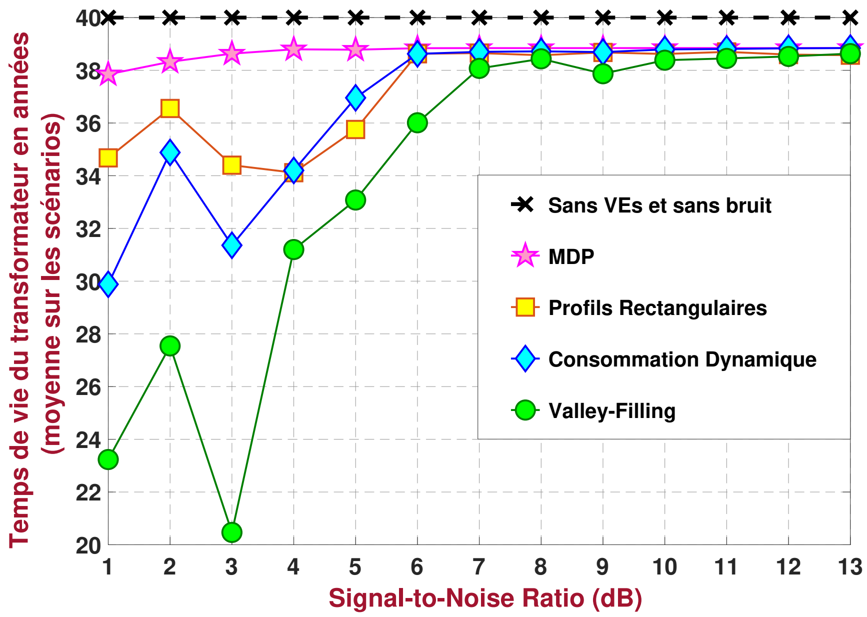


Figure A.3: Temps de vie du transformateur du RD en fonction des bruits de prévision sur le la consommation non-contrôlable réelle (dans chaque scénario), pour $I = 10$ VEs. (Cas centralisé).

1.3 PROBLÈME GÉNÉRAL SOUS-JACENT

Dans cette section, nous nous intéressons à l'existence et la synthèse des stratégies dans les MDPs doublement pondérés, qui satisfont à la fois une contrainte en probabilité sur une condition d'accessibilité pondérée, et une contrainte quantitative sur la valeur espérée d'une variable aléatoire définie à l'aide d'une autre condition d'accessibilité. Cette étude généralise le problème d'énergie vu dans la Section A.2 (tel que la recharge de VEs). Nous étudions ici l'ensemble des valeurs d'un paramètre (un seuil) pour lequel le problème général a une solution, et montrons comment une caractérisation partielle de cet ensemble peut être obtenue via deux séquences de problèmes d'optimisation.

POSITION DU PROBLÈME

Le problème (centralisé) de la planification de consommation d'énergie (e.g., la recharge des VEs) consiste en l'ordonnancement de la consommation contrôlable dans un intervalle de temps en présence d'une consommation non-contrôlable incertaine de manière à minimiser l'impact de la consommation totale sur le RD, voir (A.8). Une première contrainte sur le transformateur du RD est donnée par sa puissance maximale admissible à chaque instant (A.1), et une deuxième peut représenter le besoin d'énergie (A.3). La consommation contrôlable peut donc être aussi considérée comme une première fonction de coût. En utilisant un modèle standard pour le vieillissement du transformateur (e.g., voir [M1]), nous pouvons exprimer ce dernier sous la forme d'une deuxième fonction de coût. Ainsi, un MDP doublement pondéré peut être construit, voir Définition 2.6.9. Dans ce contexte, nous combinons (voir Définition A.3.1) une condition d'accessibilité sûre vers un ensemble cible (en représentant, e.g., la recharge des VEs), une contrainte en probabilité sur la proportion de chemins ayant une valeur supérieure à certain seuil avec la première fonction de coût (e.g., contraint sur la consommation totale sur le RD), et une contrainte sur la valeur espérée d'une fonction du deuxième coût (e.g., le vieillissement du transformateur).

Ici, $\Delta[\Pi]$ est l'ensemble des stratégies mixtes, voir Section 2.6.3, $\Omega_{x_0}^{\sigma_\varepsilon}$ est l'ensemble des chemins à partir de x_0 selon σ_ε fixée, voir Section 2.6.4, et TS est la fonction de la somme tronquée, voir Définition 2.6.7

Définition A.3.1 : Problème Général

Étant donné un MDP double pondéré $\mathcal{M} = (\mathcal{X}, \mathcal{A}, P, (C_i)_{i=1}^2)$, un état initial $x_0 \in \mathcal{X}$, un ensemble objectif[a] $\mathcal{G} \subset \mathcal{X}$, et deux seuils $\nu_1, \nu_2 \in \mathbb{Q}$. Pour chaque $\varepsilon \in [0,1] \cap \mathbb{Q}$, le problème général $\mathbf{Pb}(\varepsilon)$ est défini par[b] : il existe $\sigma_\varepsilon \in \Delta[\Pi]$, telle que chaque $\omega \in \Omega_{x_0}^{\sigma_\varepsilon}$ atteint $\mathcal{G}$,

$$\mathbb{P}_{x_0}^{\sigma_\varepsilon}[\,\mathrm{TS}_1^{\mathcal{G}} \geq \nu_1\,] \geq 1-\varepsilon \qquad \text{et} \qquad \mathbb{E}_{x_0}^{\sigma_\varepsilon}[\,\mathrm{TS}_2^{\mathcal{G}}\,] \leq \nu_2$$

a L'ensemble goal $\mathcal{G}$ est absorbant par Hypothèse 2.6.8.

b Pour être plus précis, le problème devrait être noté comme $\mathbf{Pb}_{\mathcal{M},\nu_1,\nu_2}(\varepsilon)$, car il dépend sur $\mathcal{M}$, ν_1 and ν_2 fixés. Pour relâcher la notation, nous écrivons simplement $\mathbf{Pb}(\varepsilon)$.

Notre objectif est de calculer les valeurs de ε pour lequel $\mathbf{Pb}(\varepsilon)$ a une solution. Nous supposons qu'il existe une stratégie satisfaisant la contrainte de l'espérance. Sinon, $\mathbf{Pb}(\varepsilon)$ n'a aucune solution quel que soit ε. Ce dernier problème est réduit au problème du « plus court chemin », où l'existence d'une stratégie γ-optimale, pure et sans mémoire (c.-à-d., de Markov (A.7)) π_γ, telle que $\mathbb{E}_{x_0}^{\pi_\gamma}[\mathrm{TS}_2^{\mathcal{G}}] \leq \nu_2 + \gamma$, où $\gamma > 0$, peut être décidée en temps POLYNOMIAL, voir par exemple [14, 25]. Dans ce qui suit, nous définissons la fonction suivante pour caractériser le problème $\mathbf{Pb}(\varepsilon)$.

Définition A.3.2 : Cartography

Nous appellerons **cartographie** de $\mathbf{Pb}(\varepsilon)$ à la fonction qui associe à chaque $\varepsilon \in [0,1]$, soit VRAI si $\mathbf{Pb}(\varepsilon)$ a une solution, soit FAUX sinon.

Nous décrivons maintenant un algorithme pour cartographier approximativement $\mathbf{Pb}(\varepsilon)$ sur l'intervalle $[0,1]$.

A.3.1 *CARTOGRAPHIE APPROXIMÉE*

Nous introduisons deux problèmes d'optimisation liés à $\mathbf{Pb}(\varepsilon)$, à partir desquels nous obtenons des informations sur les valeurs de ε. Notre approche caractérise partiellement l'intervalle $[0,1]$, cependant sous certaines conditions dans la structure de $\mathcal{M}$, la cartographie est presque complète [M3]. Nous dénotons l'espace de probabilité induit à partir de $\mathcal{M}$ par $(\Omega_{x_0}, \mathcal{B}(\Omega_{x_0}), \mathbb{P}_{x_0})$, voir Section 2.6.2 pour plus de détails. Pour $T \in \mathbb{N}_0$, nous définissons deux événements sur les chemins :

$\mathrm{proj}_t^{\mathcal{X}}$ est la $t^{\text{ème}}$-projection d'un chemin sur les états dans $\mathcal{X}$

$$\mathcal{E}_T \;:=\; \left\{ \omega \in \Omega_{x_0} \;\mid\; \exists\, t \leq T \;:\; \mathrm{proj}_t^{\mathcal{X}}(\omega) \in \mathcal{G} \right\},$$

c.-à-d., $\mathcal{E}_T$ représente l'ensemble mesurable de chemins qui atteignent $\mathcal{G}$ dans au plus T-pas. Nous dénotons aussi le complement de $\mathcal{E}_T$ par $\overline{\mathcal{E}}_T$. Ce dernier est l'ensemble mesurable de chemins qui n'atteignent pas $\mathcal{G}$ dans les premiers T-pas, c.-à-d.,

$$\overline{\mathcal{E}}_T \;=\; \left\{ \omega \in \Omega_{x_0} \;\mid\; \mathrm{proj}_t^{\mathcal{X}}(\omega) \notin \mathcal{G} \;,\; \forall\, t \leq T \right\}.$$

Nous écrivons aussi l'ensemble suivant :

$$\overline{\mathcal{E}}_{\alpha,T} \;:=\; \begin{cases} \overline{\mathcal{E}}_T & \text{if } \alpha = 1 \\ \varnothing & \text{if } \alpha = 0 \end{cases}$$

PROBLÈMES D'OPTIMISATION

Nous considérons une **fonction objective** pour $T \in \mathbb{N}$ fixé. Pour une stratégie pure $\pi \in \Pi$, elle est définie par :

Pour des stratégies mixtes $\sigma \in \Delta[\Pi]$, la définition est équivalente pour $\mathbb{J}_{\alpha,T}(\sigma)$.

$$\mathbb{J}_{\alpha,T}(\pi) \;:=\; \mathbb{P}_{x_0}^{\pi}[\,(\mathcal{E}_T \cap \mathrm{TS}_1^{\mathcal{G}} < \nu_1) \cup \overline{\mathcal{E}}_{\alpha,T}\,], \tag{A.27}$$

Ainsi, les deux problèmes d'optimisation peuvent être regroupés sous le paramètre $\alpha \in \{0,1\}$ comme suit :

Notez que par l'hypothèse de l'existence de π_γ, ce problème est faisable

Problème Primal - Stratégies Pures

$$[\textbf{P-PS}] \qquad \min_{\pi} \ \mathbb{J}_{\alpha,T}(\pi) \qquad \text{s.t.} \ \mathbb{E}^{\pi}_{x_0}[\,\mathrm{TS}^{\mathcal{G}}_2\,] \ \leq \ \nu_2 \tag{A.28}$$

Puisque les stratégies mixtes sont vues comme des combinaisons convexes des stratégies pures, il est facile de voir que le problème (A.28) peut être écrit pour des stratégies mixtes comme suit :

Problème Primal - Stratégies Mixtes

$$[\textbf{P-MS}] \qquad \min_{\sigma} \ \sum_{\pi} \sigma(\pi)\, \mathbb{J}_{\alpha,T}(\pi) \qquad \text{s.t.} \ \sum_{\pi} \sigma(\pi)\, \mathbb{E}^{\pi}_{x_0}[\,\mathrm{TS}^{\mathcal{G}}_2\,] \ \leq \ \nu_2 \tag{A.29}$$

Nous écrivons la **valeur optimale** de (A.28) (resp. de (A.29)) comme $\mathbb{J}^{\mathrm{p}*}_{\alpha,T}$ (resp. $\mathbb{J}^{\mathrm{m}*}_{\alpha,T}$). Notez que $\mathbb{J}^{\mathrm{m}*}_{\alpha,T} \leq \mathbb{J}^{\mathrm{p}*}_{\alpha,T}$, car des stratégies mixtes nous permettent d'obtenir une valeur optimale qui pourrait être meilleure que sous des stratégies pures. Pour les deux contextes des stratégies, nous notons simplement $\mathbb{J}^{*}_{\alpha,T}$. Il n'est pas difficile de voir que pour chaque $T \in \mathbb{N}$, $\mathbb{J}^{*}_{0,T} \leq \mathbb{J}^{*}_{1,T}$; et que la séquence $(\mathbb{J}^{*}_{\alpha,T})_{T\in\mathbb{N}}$ est croissante pour $\alpha = 0$ et décroissante pour $\alpha = 1$. Puisque cette dernière est bornée, elle est donc convergente [M3]. Nous écrivons sa limite par $\mathbb{J}^{*}_{\alpha,\infty} := \lim_{T\to+\infty} \mathbb{J}^{*}_{\alpha,T}$.

APPROCHE PRESQUE COMPLÈTE

Les deux théorèmes suivants caractérisent la cartographie approximative du problème $\mathbf{Pb}(\varepsilon)$.

Théorème A.3.3

Si $\varepsilon \ < \ \mathbb{J}^{*}_{1,\infty} \Rightarrow$ le problème $\mathbf{Pb}(\varepsilon)$ n'a pas de solution.

Preuve : Voir la Preuve B.1.3 dans l'Appendice B.1. ∎

Théorème A.3.4

Pour chaque $T \in \mathbb{N}$, le problème $\mathbf{Pb}(\varepsilon)$ a une solution $\forall\, \varepsilon > \mathbb{J}^{*}_{1,T}$, et n'a pas de solution $\forall\, \varepsilon < \mathbb{J}^{*}_{0,T}$.

Preuve : Voir la Preuve B.1.4 dans l'Appendice B.1. ∎

La Figure A.4 résume l'analyse des lemmes précédents. Théoriquement (Théorème A.3.3), elle est presque-complète, puisque le seul statut de $\varepsilon = \mathbb{J}^{*}_{1,\infty}$ reste incertain. Cependant, il reste à discuter une chose pour faire un calcul effectif (Theorem A.3.4) : les limites $\mathbb{J}^*_{0,\infty}$ et $\mathbb{J}^*_{1,\infty}$ sont à priori inconnues par calcul, d'où la cartographie n'est pas efficace. L'idée est alors d'utiliser la séquence $(\mathbb{J}^*_{\alpha,T})_{T\in\mathbb{N}}$ pour chaque α, pour se rapprocher aux limites. Cependant, si les deux limites coïncident, nous aurions une approche presque complète et donc un algorithme couvrant presque l'intervalle $[0,1]$ avec soit la ligne rouge (où il n'y a pas de solution), soit la ligne verte (où il y a une solution). Il existe des situations (relatives aux cycles dans $\mathcal{M}$) dans lesquelles on peut montrer que $\mathbb{J}^*_{0,\infty} = \mathbb{J}^*_{1,\infty}$, ce qui permet de réduire la partie inconnue de la cartographie à un seul singleton, c.-à.-d., $\varepsilon = \mathbb{J}^*_{0,\infty} = \mathbb{J}^*_{1,\infty}$.

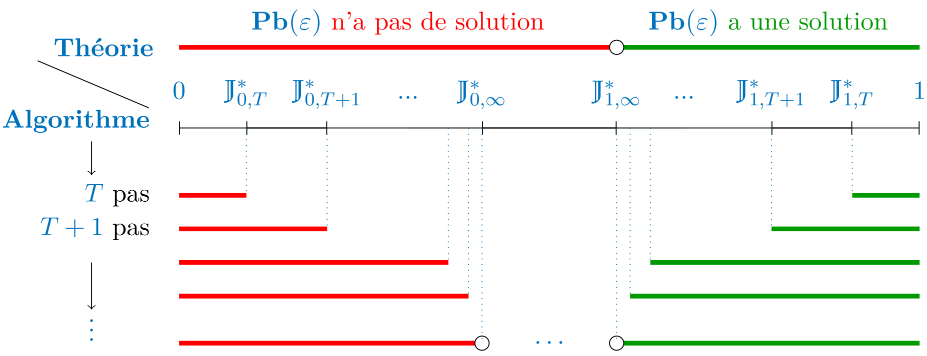


Figure A.4: Cartographie partielle du problème **Pb**(ε).

Théorème A.3.5

Si tous les cycles ont des coûts positifs sous C_i, pour au moins un $i \in \{1,2\}$; alors $\mathbb{J}^*_{0,\infty} = \mathbb{J}^*_{1,\infty}$.

Lorsque nous supposons que les cycles ont des coûts positifs, nous le pensons pour chaque cycle dans le MDP $\mathcal{M}$

Preuve : Voir la Preuve B.1.5 dans l'Appendice B.1. ∎

1.4 PROBLÈMES D'OPTIMISATION SOUS-JACENTS

Dans cette section, nous nous concentrons sur la solution aux problèmes d'optimisation de la Section A.3. Pour cela, nous fixons $T \in \mathbb{N}$ et déplions le MDP double pondéré $\mathcal{M}$ du problème **Pb**(ε) (voir Définition A.3.1) jusqu'à la profondeur T comme un arbre en gardant une copie de $\mathcal{M}$ sous chaque feuille. Cette nouvelle structure sera dénotée $\mathcal{M}_T$. À partir de chaque feuille, nous utiliseront toujours la stratégie du plus court chemin π_γ, de sorte que les problèmes d'optimisation sont réduits à trouver une solution $\forall t < T$. Nous notons $\overline{\pi} := \pi\pi_\gamma$ pour une stratégie qui joue depuis $t=0$ jusqu'à $t = T-1$ comme une stratégie pure $\pi \in$ sur $\mathcal{M}_T$, et comme π_γ pour chaque $t \geq T$. L'ensemble de ce type des stratégies $\overline{\pi}$ est dénoté par $\underline{\Pi}_\gamma \subset \Pi$.

Les solutions des problèmes d'optimisation sont obtenues en résolvant des problèmes duales sous-jacents comme illustré ci-dessous.

A.4.1 *APPROCHE LAGRANGIENNE*

Nous définissons la **fonction Lagrangienne** pour des stratégies pures $\overline{\pi} \in \underline{\Pi}_\gamma$ et pour un paramètre $\lambda \in \mathbb{R}_0^+$ (appelé **variable duale**), par :

$$\mathbb{L}_{\alpha,T}(\overline{\pi}, \lambda) \;:=\; \mathbb{J}_{\alpha,T}(\overline{\pi}) \;+\; \lambda\Big(\mathbb{E}^{\overline{\pi}}_{x_0}[\mathrm{TS}_2^{\mathcal{G}}] \;-\; \nu_2\Big)\,. \tag{A.30}$$

Comme dans (A.29), il n'est pas difficile de voir que la version de (A.30) pour des stratégies mixtes $\sigma \in \Delta[\underline{\Pi}_\gamma]$, peut être exprimée par :

$$\mathbb{L}_{\alpha,T}(\sigma, \lambda) \;=\; \sum_{\overline{\pi}} \sigma(\overline{\pi})\, \mathbb{L}_{\alpha,T}(\overline{\pi}, \lambda)\,. \tag{A.31}$$

En prenant le minimum sur des stratégies pures (resp. mixtes) dans (A.30) (resp. dans (A.31)), nous dénotons la **fonction Lagrangienne dual** pour des stratégies pures et mixtes, resp. comme :

$$\begin{aligned} \mathbb{L}^{\mathrm{p}}_{\alpha,T}(\lambda) \;&:=\; \min_{\overline{\pi}} \;\mathbb{L}_{\alpha,T}(\overline{\pi}, \lambda) \\ \mathbb{L}^{\mathrm{m}}_{\alpha,T}(\lambda) \;&:=\; \min_{\sigma} \;\mathbb{L}_{\alpha,T}(\sigma, \lambda) \end{aligned} \tag{A.32}$$

Il se trouve que les fonctions (A.30) et (A.31) sont toujours concaves en λ, et ainsi, les fonctions dans (A.32) aussi. Nous définissons le **problème dual** pour des stratégies pures et mixtes, resp. par :

Problème Dual - Stratégies Pures et Mixtes

$$\begin{aligned} &[\textbf{D-PS}] \qquad \sup_{\lambda\geq 0} \;\mathbb{L}^{\mathrm{p}}_{\alpha,T}(\lambda) \\ &[\textbf{D-MS}] \qquad \sup_{\lambda\geq 0} \;\mathbb{L}^{\mathrm{m}}_{\alpha,T}(\lambda) \end{aligned} \tag{A.33}$$

Dans ce qui suit, la **valeur dual optimale** pour les problèmes dans (A.33) pour des stratégies pures et mixtes sera notée resp. par $\mathbb{L}^{\mathrm{p}*}_{\alpha,T}$ et $\mathbb{L}^{\mathrm{m}*}_{\alpha,T}$. Dans ce contexte, le « théorème de dualité faible » peut s'appliquer [23], montrant que :

$$\mathbb{J}^{\mathrm{p}*}_{\alpha,T} \;-\; \mathbb{L}^{\mathrm{p}*}_{\alpha,T} \;\geq\; 0 \qquad \text{et} \qquad \mathbb{J}^{\mathrm{m}*}_{\alpha,T} \;-\; \mathbb{L}^{\mathrm{m}*}_{\alpha,T} \;\geq\; 0\,.$$

Quand l'inégalité est stricte nous disons qu'il n'existe qu'une **dualité faible**, et s'il y a l'égalité nous disons qu'il y a une **dualité forte**. On peut montrer qu'il y a une dualité faible sous des stratégies pures et une dualité forte sous des stratégies mixtes. Dans le dernier cas, le problème primal et dual sont équivalentes, dans le sens qu'une stratégie optimale peut être construite à l'aide d'une solution au problème dual. Il est également possible de montrer que les valeurs duales optimales des problèmes duales pour des stratégies pures et mixtes sont égales. Ainsi, nous pouvons résoudre (A.33) pour des stratégies pures (problème sans contrainte) et donc, la variable dual optimale λ^* sera aussi une solution au problème dual pour des stratégies mixtes.

Proposition A.4.1

Soit $\sigma_\alpha^* \in \Delta[\underline{\Pi}_\gamma]$, $\lambda^* \geq 0$ et $\nu_2 \in \mathbb{R}$ un seuil. Les affirmations suivantes sont équivalentes:

(*i*) σ_α^* est une (stratégie mixte) solution au problème P-MS et λ^* est une (variable dual) solution au problème D-MS.

(*ii*) $\sigma_\alpha^* \in \arg\min_\sigma \mathbb{L}_{\alpha,T}(\sigma,\lambda^*)$, $\mathbb{E}_{x0}^{\sigma_\alpha^*}[\,\mathrm{TS}_2^{\mathcal{G}}] \leq \nu_2$, et

$$\lambda^* \Big(\mathbb{E}_{x_0}^{\sigma_\alpha^*}[\,\mathrm{TS}_2^{\mathcal{G}}] - \nu_2 \Big) = 0 \tag{A.34}$$

Preuve : Voir la Preuve B.2.9 dans l'Appendice B.2, avec $\nu_2 = \nu'$. ■

En plus de la proposition précédente, nous pouvons montrer ce qui suit :

$$\begin{aligned} \mathbb{L}_{\alpha,T}^{\mathrm{m}*} = \mathbb{L}_{\alpha,T}^{\mathrm{m}}(\lambda^*) = \mathbb{L}_{\alpha,T}^{\mathrm{p}*} = \mathbb{L}_{\alpha,T}^{\mathrm{p}}(\lambda^*) = \mathbb{J}_{\alpha,T}^{\mathrm{m}*} &= \mathbb{J}_{\alpha,T}^{\mathrm{m}}(\sigma_\alpha^*) \\ &\leq \mathbb{J}_{\alpha,T}^{\mathrm{p}*} \end{aligned}$$

Nous nous concentrons sur les conditions suffisantes de la Proposition A.4.1 pour trouver une solution optimale au problème P-MS. Notez que si nous trouvons une stratégie $\sigma_\alpha^* \in \arg\min_\sigma \mathbb{L}_{\alpha,T}(\sigma,\lambda^*)$, telle que $\mathbb{E}_{x_0}^{\sigma_\alpha^*}[\mathrm{TS}_2^{\mathcal{G}}] \leq \nu_2$ avec $\lambda^* = 0$, alors σ_α^* est une solution au problème P-MS. Dans ce cas, nous aurons $\mathbb{J}_{\alpha,T}^{\mathrm{m}*} = \mathbb{L}_{\alpha,T}^{\mathrm{p}}(0)$ et donc, la stratégie σ_α^* choisira une stratégie pure avec une probabilité égale à un, c.-à.-d., une stratégie pure résoudre le problème P-PS et aussi P-MS, voir Figure A.5. Dans le cas contraire, quand $\lambda^* > 0$, nous montrons qu'une stratégie mixte solution au problème P-MS peut être construite à l'aide d'au plus deux stratégies pures.

Proposition A.4.2

Il existe deux stratégies pures $\overline{\pi}'_\alpha$, $\overline{\pi}''_\alpha \in \arg\min_{\overline{\pi}} \mathbb{L}_{\alpha,T}(\overline{\pi},\lambda^*)$ pour une variable duale $\lambda^* > 0$, définissant une stratégie mixte $\sigma_\alpha^* \in \Delta[\,\{\overline{\pi}'_\alpha, \overline{\pi}''_\alpha\}\,]$ solution au problème P-MS.

Preuve : Voir la Preuve B.2.11 dans l'Appendice B.2. ■

La Figure A.5 montre une représentation graphique de la Proposition A.4.2.

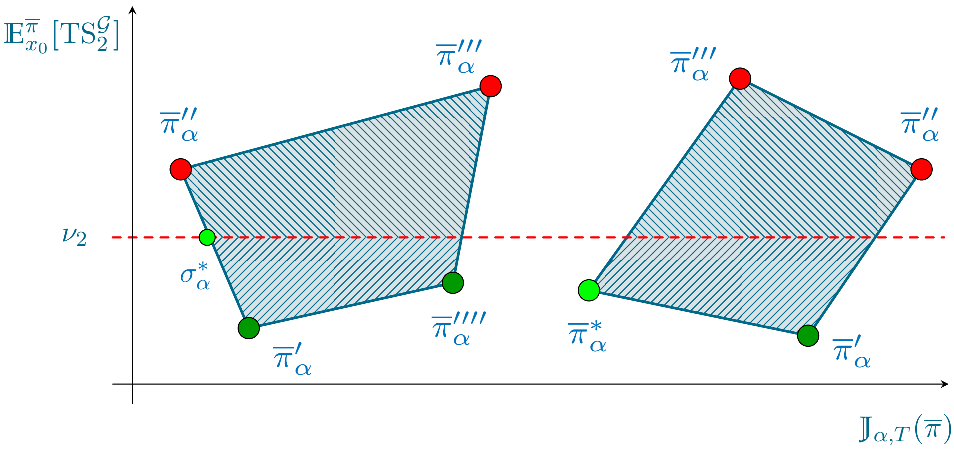


Figure A.5: Sur la gauche, σ_α^* est optimale, combinant les stratégies pures $\overline{\pi}_\alpha'$ et $\overline{\pi}_\alpha''$. Sur la droite, $\overline{\pi}_\alpha^*$ est optimale (solution du problème sans contrainte).

APPROCHE ALGORITHMIQUE

Nous examinons l'Algorithme 1 pour expliquer la Proposition A.4.2. Tout d'abord, nous résolvons le problème P-PS de (A.28) sans contrainte (cas $\lambda^* = 0$) et calculons la valeur de l'espérance dans la contrainte. Si elle est inférieure ou égale à ν_2, alors une stratégie pure est une solution (cas à droite de la Figure A.5). En supposant maintenant que cette stratégie, qui résolve le problème P-PS sans contrainte, ne satisfait pas la contrainte, on sera donc dans le cas $\lambda^* > 0$. Dans ce cas, on s'approche vers λ^* à travers la méthode de dichotomie, en générant deux suites $(\underline{\lambda}_n)_{n\in\mathbb{N}}$ et $(\overline{\lambda}_n)_{n\in\mathbb{N}}$, telles que $\underline{\lambda}_n \nearrow \lambda^*$ et $\overline{\lambda}_n \searrow \lambda^*$. Nous initialisons les suites comme $\underline{\lambda}_0 := 0$ et $\overline{\lambda}_0 := \lambda^+ > 0$. Si λ^+ n'est pas assez grande, nous cherchons une autre λ^+ jusqu'à avoir une stratégie pure qui satisfasse la contrainte. Ainsi, on aura $\overline{\pi}_{\underline{\lambda}_0}$ qui satisfait la contrainte et $\overline{\pi}_{\overline{\lambda}_0}$ qui ne la satisfait pas. Ensuite, on passe à la construction d'une stratégie mixte σ_α^*. Pour chaque $\underline{\lambda}_n$ et $\overline{\lambda}_n$ fixés, on calcule $\overline{\pi}_{\underline{\lambda}_n}$ et $\overline{\pi}_{\overline{\lambda}_n}$ à travers de la récursion classique de Bellman [21], en minimisant (A.30), qui est un problème (pour des stratégies pures) sans contrainte et avec un horizon fini. Cette approche générera deux séquences de valeurs espérées de la contrainte. Ces dernières sont des séquences monotones, comme on le montre ci-dessous.

Par hypothèse, il existe au moins une stratégie satisfaisant la contrainte

Lemma 1.4.3

Supposons que nous pouvons obtenir $\overline{\pi}_{\alpha,\lambda}^* \in \arg\min_{\overline{\pi}} \mathbb{L}_{\alpha,T}(\overline{\pi},\lambda)$ pour chaque $\lambda \geq 0$. Alors, $\lambda \mapsto \mathbb{E}_{x_0}^{\overline{\pi}_{\alpha,\lambda}^*}[\mathrm{TS}_2^{\mathcal{G}}]$ est non-croissant.

Preuve : Voir la Preuve B.2.3 dans l'Appendice B.2. ∎

Puisque par la méthode de dichotomie, $\lambda_n = (\overline{\lambda}_{n-1} + \underline{\lambda}_{n-1})/2$ est le point médian de l'intervalle $[\underline{\lambda}_{n-1}, \overline{\lambda}_{n-1}]$ à l'itération n, on peut montrer que $|\lambda_n - \lambda^*| \leq (\overline{\lambda}_0 - \underline{\lambda}_0)/2^{n-1}$, et alors que

$$\left| \mathbb{J}^{\mathrm{m}}_{\alpha,T}(\sigma^*_\alpha) \;-\; \mathbb{J}^{\mathrm{m}*}_{\alpha,T} \right| \;\leq\; E_n \,\frac{\overline{\lambda}_0 \,-\, \underline{\lambda}_0}{2^{n-1}} \;,$$

où $E_n \in \mathbb{Q}$ est en fonction des valeurs espérées avec $\pi_{\underline{\lambda}_n}$ et $\pi_{\overline{\lambda}_n}$ à l'itération n. Nous itérons jusqu'à ce que la dernière inégalité soit inférieure à un erreur $\epsilon > 0$. Ainsi, nous aurons $\pi_{\underline{\lambda}_n}$ et $\pi_{\overline{\lambda}_n}$ pour effectuer une combinaison convexe entre elles, en résolvant (A.34) avec $\lambda^* > 0$. Dans ce cas,

$$\mathbb{E}^{\pi_{\overline{\lambda}_n}}_{x_0}[\mathrm{ST}^{\mathcal{G}}_2] \;\;\leq\;\; \nu_2 \;\;\leq\;\; \mathbb{E}^{\pi_{\underline{\lambda}_n}}_{x_0}[\mathrm{ST}^{\mathcal{G}}_2]$$

et, puisque l'espérance sous des stratégies mixtes est continue, il existera une σ^*_α telle que $\sigma^*_\alpha(\pi_{\underline{\lambda}_n}) + \sigma^*_\alpha(\pi_{\overline{\lambda}_n}) = 1$, et

$$\sigma^*_\alpha(\pi_{\underline{\lambda}_n})\, \mathbb{E}^{\pi_{\underline{\lambda}_n}}_{x_0}[\mathrm{ST}^{\mathcal{G}}_2] \;+\; \sigma^*_\alpha(\pi_{\overline{\lambda}_n})\, \mathbb{E}^{\pi_{\overline{\lambda}_n}}_{x_0}[\mathrm{ST}^{\mathcal{G}}_2] \;\;=\;\; \nu_2 \;.$$

Ce dernier est un système de deux équations à deux inconnues ($\sigma^*_\alpha(\pi_{\underline{\lambda}_n})$ et $\sigma^*_\alpha(\pi_{\overline{\lambda}_n})$).

1.5 CONCLUSION

Dans cette thèse, un accent particulier a été placé sur le problème général de planification de la consommation d'énergie. Nous avons essentiellement fourni des modèles formels mathématiques, des solutions algorithmiques et des stratégies appropriées pour garantir l'exactitude et l'optimalité pour ce problème. En particulier, afin d'assurer que les puissances de consommation optimales des consommateurs répondent à leur objectifs et s'adaptent aux contraintes du réseau d'énergie. La conception des stratégies de consommation a été conçues selon une approche centralisée (dans laquelle il n'y a qu'un seul décideur qui contrôle toutes les stratégies des consommateurs) et une approche décentralisée (dans laquelle il y a plusieurs contrôleurs, chacun représentant un consommateur). Nous avons analysé ces deux scénarios en utilisant des méthodes formelles, la théorie des jeux et l'optimisation. Nous avons aussi étudié les cas où le comportement du système du réseau d'énergie est stochastique et déterministe.

Le problème de planification de la consommation d'énergie a inspiré la formulation d'un problème théorique très intéressant et riche en information. Cela a été vu comme un système stochastique avec des multi-objectifs sous contraintes. Ce problème a été modélisé à l'aide des MDPs et des jeux stochastiques, pour lequel nous avons étudié l'existence et la synthèses d'une stratégie optimale. Par conséquent, cette thèse concerne également une contribution aux modèles avec des objectives multicritères, ce qui a permit une conception des stratégies fonctionnellement correctes et robustes aux changements stochastiques de l'environnement du système, et en particulier dans le domaine des réseaux d'énergies. Ceci est principalement confirmé par les applications numériques réalisées dans ce travail.

Dans l'analyse numérique, nous avons mesuré la performance des stratégies fournies dans cette thèse (à savoir, l'impact des opérations de la consommation d'énergie) pour un scénario où les consommateurs étaient représentés comme un pool de VEs dans un quartier résidentielle. Nous avons examiné leur interactions (techniques et économiques) avec le réseau d'énergie, dans lequel l'environnement du système (à savoir, la partie non contrôlable de la consommation totale) représente la consommation hors VEs. Cette dernière était représentée par un modèle de prévision (premièrement, déterministe et, deuxièmement, stochastique) construite à travers une large base de données de la consommation d'un quartier résidentielle. Nous avons synthétisé des stratégies de consommation optimales pour, premièrement, maximiser la durée de vie du transformateur du réseau d'énergie et, deuxièmement, pour minimiser le paiement de la consommation électrique totale. Les résultats numériques montrent d'une part que la méthode de brancher-et-charger a un impact très significatif sur le réseau d'énergie. D'autre part, en analysant les performances sur différents bruits de prévision, lorsque le bruit est proche de zéro (c'est-à-dire, dans le cadre d'une « prévision parfaite ») la méthode de charge dynamique devient optimale. Dans le cas plus réaliste (c'est-à-dire, dans le cadre d'une « prévision imparfaite »), les stratégies construites à l'aide des MDPs étaient presque insensibles au bruit et sont globalement beaucoup plus robustes que les autres méthodes de consommation considérées.

B

MATHEMATICAL PROOFS

In this Appendix B, we provide the main proofs of the Chapter 4 and Chapter 5.

2.1 PROOF OF CHAPTER 4

In this Section B.1, we provide the main proofs of the Chapter 4.

Proof B.1.1: [**Proposition 2.6.6**]

Let $\mathcal{C} = (\mathcal{X}, P, (C_j)_{j=1}^{J})$ be a multi-weighted MC, with $J \in \mathbb{N}$. For a fixed $j \in \{1, \dots, J\}$ and any state $x_0 \in \mathcal{X}$, the following holds:

(i) for $T \in \mathbb{N}$ fixed, $\mathbb{TS}_{j,T}(x_0)$ can be computed iteratively through:

$$\mathbb{TS}_{j,T}(x_0) = C_j(x_0) + \sum_{x_1 \in \mathcal{X}} P(x_0)(x_1)\mathbb{TS}_{j,T-1}(x_1) ,$$

where $\mathbb{TS}_{j,0}(x_0) \coloneqq 0$ for each $x_0 \in \mathcal{X}$.

(ii) if $\mathbb{TS}_{j,\infty}(x_0)$ exists, then it satisfies the system of linear eqs.:

$$\mathbb{TS}_{j,\infty}(x_0) = C_j(x_0) + \sum_{x_1 \in \mathcal{X}} P(x_0)(x_1)\mathbb{TS}_{j,\infty}(x_1) .$$

Proof. First, for any $x_0 \in \mathcal{X}$, we have that:

$$\begin{aligned}
\mathbb{TS}_{j,T}(x_0) &= \mathbb{E}_{x_0}[\,\mathrm{TS}_{j,T}\,] \\
&= \mathbb{E}_{x_0}\left[\sum_{t=1}^{T} C_j(X_{t-1},\, X_t)\right] \\
&\quad \text{(by definition (2.21) of the truncated sum)} \\
&= \sum_{x_1,\ldots,x_T}\left(\sum_{t=1}^{T} C_j(x_{t-1},\, x_t)\right)\prod_{t=1}^{T} P(x_{t-1})(x_t) \\
&\quad \text{(by definition (2.18) of expectation)} \\
&= \sum_{x_1} P(x_0)(x_1)\left(C_j(x_0,\, x_1)\sum_{x_2,\ldots,x_T}\prod_{t=2}^{T} P(x_{t-1})(x_t)\right. \\
&\qquad \left. + \sum_{x_2,\ldots,x_T}\left(\sum_{t=2}^{T} C_j(x_{t-1},\, x_t)\right)\prod_{t=2}^{T} P(x_{t-1})(x_t)\right) \\
&= \sum_{x_1} P(x_0)(x_1)\, C_j(x_0,\, x_1)\sum_{x_2,\ldots,x_T}\prod_{t=2}^{T} P(x_{t-1})(x_t) + \\
&\qquad \sum_{x_1} P(x_0)(x_1)\sum_{x_2,\ldots,x_T}\left(\sum_{t=2}^{T} C_j(x_{t-1},\, x_t)\right)\prod_{t=2}^{T} P(x_{t-1})(x_t)\ .
\end{aligned}$$

Now, by eq. (2.20), it holds $\sum_{x_1} P(x_0)(x_1)\, C_j(x_0,\, x_1) = C_j(x_0)$, and for each $x_1 \in \mathcal{X}$,

$$\sum_{x_2,\ldots,x_T}\prod_{t=2}^{T} P(x_{t-1})(x_t) \;=\; 1\ ,$$

and

$$\begin{aligned}
\sum_{x_2,\ldots,x_T}\left(\sum_{t=2}^{T} C_j(x_{t-1},\, x_t)\right)\prod_{t=2}^{T} P(x_{t-1})(x_t) &= \mathbb{E}_{x_1}\left[\sum_{t=2}^{T} C_j(X_{t-1},\, X_t)\right] \\
&= \mathbb{E}_{x_1}[\,\mathrm{TS}_{j,T-1}\,] \\
&= \mathbb{TS}_{j,T-1}(x_1)\ .
\end{aligned}$$

Thus, it follows that

$$\mathbb{TS}_{j,T}(x_0) \;=\; C_j(x_0) + \sum_{x_1\in\mathcal{X}} P(x_0)(x_1)\mathbb{TS}_{j,T-1}(x_1)\ .$$

Finally, if $\mathbb{TS}_{j,\infty}(x_0)$ exists for each $x_0 \in \mathcal{X}$, then taking limit as $T \to \infty$ in the latter equality, the second statement follows. ■

Proof B.1.2: [Proposition 4.3.2]

The sequence $(\mathbb{J}^*_{\alpha,T})_{T\in\mathbb{N}}$ is nonincreasing for $\alpha = 1$ and nondecreasing for $\alpha = 0$. Moreover, $\mathbb{J}^*_{0,T} \le \mathbb{J}^*_{1,T}$ for each $T \in \mathbb{N}$.

Proof. First, $\mathbb{J}^*_{0,T} \le \mathbb{J}^*_{1,T}$ is clear for each $T \in \mathbb{N}$ from the inequality (4.7). Second, we need to prove that for each $T \in \mathbb{N}$, the inequalities

$\mathbb{J}_{0,T}(\delta) \leq \mathbb{J}_{0,T+1}(\delta)$ and $\mathbb{J}_{1,T+1}(\delta) \leq \mathbb{J}_{1,T}(\delta)$ hold for every $\delta \in \Delta$. From the relation (4.2) of the sets $\{\mathcal{E}_T\}_{T\in\mathbb{N}}$, we have for each $T \in \mathbb{N}$,

$$\mathcal{E}_T \cap \mathrm{TS}_1^{\mathcal{G}} < \nu_1 \quad \subseteq \quad \mathcal{E}_{T+1} \cap \mathrm{TS}_1^{\mathcal{G}} < \nu_1 \; . \tag{B.1}$$

Thus, applying the probability distribution $\mathbb{P}_{x_0}^{\delta}$ above,

$$\mathbb{P}_{x_0}^{\delta}[\, \mathcal{E}_T \cap \mathrm{TS}_1^{\mathcal{G}} < \nu_1 \,] \quad \leq \quad \mathbb{P}_{x_0}^{\delta}[\, \mathcal{E}_{T+1} \cap \mathrm{TS}_1^{\mathcal{G}} < \nu_1 \,] \; .$$

Then $\mathbb{J}_{0,T}(\delta) \leq \mathbb{J}_{0,T+1}(\delta)$. On the other hand, by the same argument as for (B.1), we have that:

$$\mathcal{E}_T \cap \mathrm{TS}_1^{\mathcal{G}} \geq \nu_1 \quad \subseteq \quad \mathcal{E}_{T+1} \cap \mathrm{TS}_1^{\mathcal{G}} \geq \nu_1 \; .$$

Using the complement of sets, see eq. (4.5), it holds:

$$(\mathcal{E}_T \cap \mathrm{TS}_1^{\mathcal{G}} < \nu_1) \cup \overline{\mathcal{E}}_T \quad \supseteq \quad (\mathcal{E}_{T+1} \cap \mathrm{TS}_1^{\mathcal{G}} < \nu_1) \cup \overline{\mathcal{E}}_{T+1} \; .$$

Applying the probability distribution of the $\mathbb{P}_{x_0}^{\delta}$ above,

$$\mathbb{P}_{x_0}^{\delta}[\, (\mathcal{E}_T \cap \mathrm{TS}_1^{\mathcal{G}} < \nu_1) \cup \overline{\mathcal{E}}_T \,] \quad \geq \quad \mathbb{P}_{x_0}^{\delta}[\, (\mathcal{E}_{T+1} \cap \mathrm{TS}_1^{\mathcal{G}} < \nu_1) \cup \overline{\mathcal{E}}_{T+1} \,] \; .$$

Then $\mathbb{J}_{1,T+1}(\delta) \leq \mathbb{J}_{1,T}(\delta)$. Finally, taking the infimum on the inequalities $\mathbb{J}_{0,T}(\delta) \leq \mathbb{J}_{0,T+1}(\delta)$ and $\mathbb{J}_{1,T+1}(\delta) \leq \mathbb{J}_{1,T}(\delta)$, over the strategies δ satisfying the constraint in expectation, see eq. (4.11), we conclude that $\mathbb{J}_{0,T}^{*} \leq \mathbb{J}_{1,T+1}^{*}$ and $\mathbb{J}_{1,T+1}^{*} \leq \mathbb{J}_{1,T}^{*}$, resp. ■

Proof B.1.3: [Theorem 4.3.3]

If $\varepsilon \; < \; \mathbb{J}_{1,\infty}^{*} \Rightarrow$ the problem $\mathbf{Pb}(\varepsilon)$ has no solution.

Proof. By Proposition 4.3.2, we know that the sequence $(\mathbb{J}_{1,T}^{*})_{T\in\mathbb{N}}$ is nonincreasing, hence convergent in $[0,1]$. Thus, we can suppose that $\mathbb{J}_{1,\infty}^{*} \in [0,1]$.

First, it is straightforward to show that if $\mathbb{J}_{1,\infty}^{*} = 0$, then the proof is trivial. We will suppose that $\mathbb{J}_{1,\infty}^{*} > 0$ and $\varepsilon < \mathbb{J}_{1,\infty}^{*}$. Now, by the nonincreasing behavior of the sequence, the following holds for each $T \in \mathbb{N}$:

$$\varepsilon \quad < \quad \mathbb{J}_{1,\infty}^{*} \quad \leq \quad \mathbb{J}_{1,T}^{*} \; . \tag{B.2}$$

Towards a contradiction, suppose that there exists a strategy δ_ε optimal solution of $\mathbf{Pb}(\varepsilon)$ for an ε satisfying (B.2). Thus, from the Definition 4.2.1 of $\mathbf{Pb}(\varepsilon)$, we have that

$$\mathbb{P}_{x_0}^{\delta_\varepsilon}[\, \mathrm{TS}_1^{\mathcal{G}} \geq \nu_1 \,] \quad \geq \quad 1 - \varepsilon \; .$$

In this way, from the Definition (2.6.7) of the truncated sum $\mathrm{TS}_1^{\mathcal{G}}$, we infer that the reachability time (2.23) $T_{\mathcal{G}}$ for which the goal set $\mathcal{G}$ is reached, is such that $\mathbb{P}_{x_0}^{\delta_\varepsilon}[\, T_{\mathcal{G}} < +\infty \,] = 1$. Thus,

$$\begin{aligned}
1 - \varepsilon \quad &\leq \quad \mathbb{P}_{x_0}^{\delta_\varepsilon}[\, \mathrm{TS}_1^{\mathcal{G}} \geq \nu_1 \,] \\
&= \quad \mathbb{P}_{x_0}^{\delta_\varepsilon}[\, \mathrm{TS}_1^{\mathcal{G}} \geq \nu_1 \mid T_{\mathcal{G}} < +\infty \,]\, \mathbb{P}_{x_0}^{\delta_\varepsilon}[\, T_{\mathcal{G}} < +\infty \,] + \\
&\qquad \mathbb{P}_{x_0}^{\delta_\varepsilon}[\, \mathrm{TS}_1^{\mathcal{G}} \geq \nu_1 \mid T_{\mathcal{G}} = +\infty \,]\, \mathbb{P}_{x_0}^{\delta_\varepsilon}[\, T_{\mathcal{G}} = +\infty \,] \\
&\qquad \text{(by splitting w.r.t. the complementary events of } T_{\mathcal{G}} \text{)} \\
&= \quad \mathbb{P}_{x_0}^{\delta_\varepsilon}[\, \mathrm{TS}_1^{\mathcal{G}} \geq \nu_1 \, \cap \, T_{\mathcal{G}} < +\infty \,] \\
&\qquad \text{(since the second term in the precedent equation is zero)} \\
&\leq \quad \mathbb{P}_{x_0}^{\delta_\varepsilon}[\, \mathrm{TS}_1^{\mathcal{G}} \geq \nu_1 \, \cap \, \mathcal{E}_\infty \,]
\end{aligned}$$

where the latter inequality holds by considering the event $\mathcal{E}_\infty$ that represents all paths reaching $\mathcal{G}$ at some time-step ($T \to +\infty$), see eq. (4.1). Thus, the following holds from the latter inequalities by applying complements, see eq. (4.5):

$$\varepsilon \;\geq\; \mathbb{P}_{x_0}^{\delta_\varepsilon}[\,(\mathcal{E}_\infty \;\cap\; \mathrm{TS}_1^{\mathcal{G}} < \nu_1) \;\cup\; \overline{\mathcal{E}}_\infty] \;\geq\; \mathbb{J}_{1,\infty}^* \;>\; \varepsilon\,.$$

which is a contradiction. Therefore, $\mathbf{Pb}(\varepsilon)$ has no solution. ■

Proof B.1.4: [Theorem 4.3.5]

For each $T \in \mathbb{N}$, the problem $\mathbf{Pb}(\varepsilon)$ has a solution $\forall\, \varepsilon > \mathbb{J}_{1,T}^*$, and has no solution $\forall\, \varepsilon < \mathbb{J}_{0,T}^*$.

Proof. Let $T \in \mathbb{N}$ and $\varepsilon > \mathbb{J}_{1,T}^*$. Since $\mathbb{J}_{1,T}^*$ is the optimal value of the optimization problem (4.11) with $\alpha = 1$, let $\gamma > 0$ such that

$$\gamma \;<\; \varepsilon \;-\; \mathbb{J}_{1,T}^*\,.$$

For such a γ, let δ_γ^* be a γ-optimal strategy, see eq. (4.12), for the problem (4.10) with $\alpha = 1$. Thus, the following holds:

$$\mathbb{P}_{x_0}^{\delta_\gamma^*}[\,(\mathcal{E}_T \cap \mathrm{TS}_1^{\mathcal{G}} < \nu_1) \cup \overline{\mathcal{E}}_T\,] \;<\; \varepsilon \quad \text{and} \quad \mathbb{E}_{x_0}^{\delta_\gamma^*}[\,\mathrm{TS}_2^{\mathcal{G}}\,] \;<\; \nu_2\,.$$

Let $\overline{\mathrm{TS}}_2^{\mathrm{sp}}$ a positive upper-bound of the truncated sum with the cost function C_2 when the strategy of the shortest path problem is applied from any state in $\mathcal{M}$ to reach $\mathcal{G}$. The latter strategy is denoted by π^{sp}. For $\tau > T$, we define a non-stationary strategy $\delta_{\gamma,\tau}$ as: play δ_γ^* for the first τ-steps, and if the goal set $\mathcal{G}$ is not reached, then play π^{sp}. We show that we can find τ large enough such that $\delta_{\gamma,\tau}$ is a solution to the problem $\mathbf{Pb}(\varepsilon)$.

The first condition of $\mathbf{Pb}(\varepsilon)$ is satisfied, see Definition 4.2.1, since either the goal set is reached during the first T-steps, i.e., while playing δ_γ^*, or it will be surely reached by playing π^{sp}. Thus, the reachability time (2.23) $T_{\mathcal{G}}$ for which the goal set $\mathcal{G}$ is reached, is such that

$$\mathbb{P}_{x_0}^{\delta_{\gamma,\tau}}[\,T_{\mathcal{G}} < +\infty\,] \;=\; 1\,.$$

Second, since

$$\mathbb{P}_{x_0}^{\delta_{\gamma,\tau}}[\,\mathcal{E}_T \cap \mathrm{TS}_1^{\mathcal{G}} \geq \nu_1\,] \;=\; 1 \;-\; \mathbb{P}_{x_0}^{\delta_{\gamma,\tau}}[\,(\mathcal{E}_T \cap \mathrm{TS}_1^{\mathcal{G}} < \nu_1) \cup \overline{\mathcal{E}}_T\,]\,,$$

and, by the relation of sets (4.2), it holds:

$$\mathbb{P}_{x_0}^{\delta_{\gamma,\tau}}[\,\mathcal{E}_\infty \cap \mathrm{TS}_1^{\mathcal{G}} \geq \nu_1\,] \;\geq\; \mathbb{P}_{x_0}^{\delta_{\gamma,\tau}}[\,\mathcal{E}_T \cap \mathrm{TS}_1^{\mathcal{G}} \geq \nu_1\,]\,.$$

Then, we conclude that:

$$\mathbb{P}_{x_0}^{\delta_{\varepsilon,\tau}}[\,\mathcal{E}_\infty \cap \mathrm{TS}_1^{\mathcal{G}} \geq \nu_1\,] \;\geq\; 1 - \varepsilon\,,$$

which is the second condition of $\mathbf{Pb}(\varepsilon)$, see Definition 4.2.1. Finally, we have that:

$$\begin{aligned}
\mathbb{E}_{x0}^{\delta_{\gamma,\tau}}[\,\mathrm{TS}_2^{\mathcal{G}}\,] = \; & \mathbb{E}_{x0}^{\delta_{\gamma,\tau}}[\,\mathrm{TS}_2^{\mathcal{G}} \mid T_{\mathcal{G}} \le \tau]\,\mathbb{P}_{x0}^{\delta_{\gamma,\tau}}[T_{\mathcal{G}} \le \tau] + \\
& \mathbb{E}_{x0}^{\delta_{\gamma,\tau}}[\,\mathrm{TS}_2^{\mathcal{G}} \mid T_{\mathcal{G}} > \tau]\,\mathbb{P}_{x0}^{\delta_{\gamma,\tau}}[T_{\mathcal{G}} > \tau] \\
& \text{(by splitting w.r.t. the complementary events of } T_{\mathcal{G}}\text{)} \\
= \; & \mathbb{E}_{x0}^{\delta_{\gamma,\tau}}\left[\sum_{t=0}^{\tau} C_2(X_t, A_t, X_{t+1}) \mid T_{\mathcal{G}} \le \tau\right]\mathbb{P}_{x0}^{\delta_{\gamma,\tau}}[T_{\mathcal{G}} \le \tau] + \\
& \mathbb{E}_{x0}^{\delta_{\gamma,\tau}}\left[\sum_{t=0}^{\tau} C_2(X_t, A_t, X_{t+1}) \mid T_{\mathcal{G}} > \tau\right]\mathbb{P}_{x0}^{\delta_{\gamma,\tau}}[T_{\mathcal{G}} > \tau] + \\
& \mathbb{E}_{x0}^{\delta_{\gamma,\tau}}\left[\sum_{t=\tau+1}^{T_{\mathcal{G}}} C_2(X_t, A_t, X_{t+1}) \mid T_{\mathcal{G}} > \tau\right]\mathbb{P}_{x0}^{\delta_{\gamma,\tau}}[T_{\mathcal{G}} > \tau] \\
& \text{(by splitting } \mathrm{TS}_2^{\mathcal{G}} \text{ in the second expectation)} \\
= \; & \mathbb{E}_{x0}^{\delta_{\gamma,\tau}}\left[\sum_{t=0}^{\tau} C_2(X_t, A_t, X_{t+1})\right] + \\
& \mathbb{E}_{x0}^{\delta_{\gamma,\tau}}\left[\sum_{t=\tau+1}^{T_{\mathcal{G}}} C_2(X_t, A_t, X_{t+1}) \mid T_{\mathcal{G}} > \tau\right]\mathbb{P}_{x0}^{\delta_{\gamma,\tau}}[T_{\mathcal{G}} > \tau] \\
& \text{(by combining the first two terms)} \\
\le \; & \mathbb{E}_{x0}^{\delta_{\gamma}^{*}}\left[\sum_{t=0}^{\tau} C_2(X_t, A_t, X_{t+1})\right] + \overline{\mathrm{TS}}_2^{\mathrm{sp}}\,\frac{\mathbb{E}_{x0}^{\delta_{\gamma,\tau}}[\,T_{\mathcal{G}}\,]}{\tau+1}
\end{aligned}$$

where the latter inequality holds since $\delta_{\gamma,\tau}$ plays as δ_{γ}^{*} up to τ and, bounding the second term with $\overline{TS}_2^{\mathrm{sp}}$ and using the Markov's inequality[1]. In summary, it holds from the precedent inequalities:

$$\mathbb{E}_{x0}^{\delta_{\gamma,\tau}}[\,\mathrm{TS}_2^{\mathcal{G}}\,] \;\le\; \mathbb{E}_{x0}^{\delta_{\gamma}^{*}}\left[\sum_{t=0}^{\tau} C_2(X_t, A_t, X_{t+1})\right] + \overline{\mathrm{TS}}_2^{\mathrm{sp}}\,\frac{\mathbb{E}_{x0}^{\delta_{\gamma,\tau}}[\,T_{\mathcal{G}}\,]}{\tau+1}\,.$$

Note that the expectation of $T_{\mathcal{G}}$ is finite, then $\mathbb{E}_{x0}^{\delta_{\gamma,\tau}}[\,\mathrm{TS}_2^{\mathcal{G}}\,]$ is upper-bounded by a finite value. Moreover, the second term on the right side of the precedent inequality is decreasing to zero as $\tau \to +\infty$. Thus, one can choose τ large enough such that:

$$\overline{\mathrm{TS}}_2^{\mathrm{sp}}\,\frac{\mathbb{E}_{x0}^{\delta_{\gamma,\tau}}[T_{\mathcal{G}}]}{\tau+1} \;<\; \frac{\nu_2}{2} \;-\; \mathbb{E}_{x0}^{\delta_{\gamma}^{*}}[\,\mathrm{TS}_2^{\mathcal{G}}\,]\,,$$

and

$$\left|\,\mathbb{E}_{x0}^{\delta_{\gamma}^{*}}\left[\sum_{t=0}^{\tau} C_2(X_t, A_t, X_{t+1})\right] \;-\; \mathbb{E}_{x0}^{\delta_{\gamma}^{*}}[\,\mathrm{TS}_2^{\mathcal{G}}\,]\,\right| \;<\; \frac{\nu_2}{2}\,.$$

In such a way, we conclude that $\mathbb{E}_{x0}^{\delta_{\gamma,\tau}}[\,\mathrm{TS}_2^{\mathcal{G}}\,] < \nu_2$, and then the strategy $\delta_{\gamma,\tau}$ is a solution for the problem $\mathbf{Pb}(\varepsilon)$.

For the second statement of the Theorem 4.3.5 under consideration, suppose that for each $T \in \mathbb{N}$, $\varepsilon < \mathbb{J}_{0,T}^{*}$. We prove that $\mathbf{Pb}(\varepsilon)$ has no

1 That is, for a nonnegative r.v, here $T_{\mathcal{G}}$, and a positive parameter, here $\tau+1$; then the probability that $T_{\mathcal{G}}$ is at least $\tau+1$, is at most the expectation of $T_{\mathcal{G}}$ divided by $\tau+1$.

solution. If $\varepsilon < \mathbb{J}^*_{0,T}$, then by Proposition 4.3.2, we have that $\mathbb{J}^*_{0,T} \leq \mathbb{J}^*_{1,T}$ for each $T \in \mathbb{N}$. Thus, $\varepsilon < \mathbb{J}^*_{1,T}$ for each $T \in \mathbb{N}$ and so $\varepsilon < \mathbb{J}^*_{1,\infty}$. By Theorem 4.3.3, we conclude that $\mathbf{Pb}(\varepsilon)$ has no solution. ■

Proof B.1.5: [Theorem 4.4.1]

If all cycles have positive costs[a] under C_j for some $j \in \{1, 2\}$, then $\mathbb{J}^*_{0,\infty} = \mathbb{J}^*_{1,\infty}$.

[a] When we assume that cycles have positive costs, we mean it for every cycle, except for cycles at $\mathcal{G}$, which we assumed are self-loops with costs equal to zero.

Proof. We prove the theorem for each j independently.

$j = 1$| Suppose that all cycles have positive costs under C_1. In such case, it is straightforward to show that there exists $T_0 \in \mathbb{N}$, such that if

$$\omega \;\in\; \bigcup_{t=T_0}^{+\infty} \overline{\mathcal{E}}_t \,\cup\, \mathcal{E}_\infty \quad\Rightarrow\quad \mathrm{TS}^{\mathcal{G}}_1(\omega) \;\geq\; \nu_1 \,.$$

Let T_0 as before. Since the goal set $\mathcal{G}$ is absorbent, then for any $T \geq T_0$ and every ω of length larger than T, the following holds:

$$\omega \in \mathcal{E}_T \cap \mathrm{TS}^{\mathcal{G}}_1 < \nu_1 \;\Leftrightarrow \omega \in \mathcal{E}_{T_0} \cap \mathrm{TS}^{\mathcal{G}}_1 < \nu_1 \tag{B.3}$$

$$\omega \in \mathcal{E}_T \cap \mathrm{TS}^{\mathcal{G}}_1 \geq \nu_1 \;\Leftrightarrow \omega \in (\mathcal{E}_{T_0} \cap \mathrm{TS}^{\mathcal{G}}_1 \geq \nu_1) \cup \bigcup_{t=T_0+1}^{T} \mathcal{E}_t \cap \overline{\mathcal{E}}_{T_0} \tag{B.4}$$

First, we infer that for every $T \geq T_0$,

$$\begin{aligned}
\mathbb{J}^*_{0,T} \;&=\; \inf_\delta \left\{ \mathbb{J}_{0,T}(\delta) \;\mid\; \mathbb{E}^\delta_{x_0}[\,\mathrm{TS}^{\mathcal{G}}_2\,] \;<\; \nu_2 \right\} \\
&=\; \inf_\delta \left\{ \mathbb{P}^\delta_{x_0}[\,\mathcal{E}_T \cap \mathrm{TS}^{\mathcal{G}}_1 < \nu_1\,] \;\mid\; \mathbb{E}^\delta_{x_0}[\,\mathrm{TS}^{\mathcal{G}}_2\,] \;<\; \nu_2 \right\} \\
&\qquad \text{(by definition of } \mathbb{J}_{0,T}(\delta)\text{)} \\
&=\; \inf_\delta \left\{ \mathbb{P}^\delta_{x_0}[\,\mathcal{E}_{T_0} \cap \mathrm{TS}^{\mathcal{G}}_1 < \nu_1\,] \;\mid\; \mathbb{E}^\delta_{x_0}[\,\mathrm{TS}^{\mathcal{G}}_2\,] \;<\; \nu_2 \right\} \\
&\qquad \text{(by the equivalence (B.3))} \\
&=\; \inf_\delta \left\{ \mathbb{J}_{0,T_0}(\delta) \;\mid\; \mathbb{E}^\delta_{x_0}[\,\mathrm{TS}^{\mathcal{G}}_2\,] \;<\; \nu_2 \right\} \\
&\qquad \text{(by definition of } \mathbb{J}_{0,T_0}(\delta)\text{)} \\
&=\; \mathbb{J}^*_{0,T_0}
\end{aligned}$$

From the latter, taking $T \to +\infty$, we have:

$$\mathbb{J}^*_{0,\infty} \;=\; \mathbb{J}^*_{0,T_0} \,. \tag{B.5}$$

To relax the notation, we write Δ_{ν_2} the set of randomized strategies that satisfy the constraint in expectation, i.e.,

$$\Delta_{\nu_2} \;=\; \{\, \delta \in \Delta \;\mid\; \mathbb{E}^\delta_{x_0}[\,\mathrm{TS}^{\mathcal{G}}_2\,] \;<\; \nu_2 \,\} \,,$$

this set is not empty by Assumption 4.2.3. Now, from the equivalence of the eq. (B.4), we can write for every $T \geq T_0$, the following:

$$\begin{aligned}
\mathbb{J}^*_{1,T} &= \inf_{\delta \in \Delta_{\nu_2}} \mathbb{J}_{1,T}(\delta) \\
&= \inf_{\delta \in \Delta_{\nu_2}} \mathbb{P}^{\delta}_{x_0}[\,(\mathcal{E}_T \cap \mathrm{TS}^{\mathcal{G}}_1 < \nu_1) \cup \overline{\mathcal{E}}_T\,] \\
&\quad \text{(by definition of } \mathbb{J}_{1,T}(\delta)\text{)} \\
&= 1 - \sup_{\delta \in \Delta_{\nu_2}} \mathbb{P}^{\delta}_{x_0}[\,\mathcal{E}_T \cap \mathrm{TS}^{\mathcal{G}}_1 \geq \nu_1\,] \\
&= 1 - \sup_{\delta \in \Delta_{\nu_2}} \mathbb{P}^{\delta}_{x_0}\left[(\mathcal{E}_{T_0} \cap \mathrm{TS}^{\mathcal{G}}_1 \geq \nu_1) \cup \left(\bigcup_{t=T_0+1}^{T} \mathcal{E}_t \cap \overline{\mathcal{E}}_{T_0}\right)\right] \\
&\quad \text{(by the equivalence (B.4))}
\end{aligned}$$

Thus, taking the limit as $T \to +\infty$ on the above, we have that:

$$\mathbb{J}^*_{1,\infty} = 1 - \lim_{T\to+\infty} \sup_{\delta \in \Delta_{\nu_2}} \mathbb{P}^{\delta}_{x_0}\left[(\mathcal{E}_{T_0} \cap \mathrm{TS}^{\mathcal{G}}_1 \geq \nu_1) \cup \left(\bigcup_{t=T_0+1}^{T} \mathcal{E}_t \cap \overline{\mathcal{E}}_{T_0}\right)\right] \tag{B.6}$$

However, we can show that:

$$\begin{aligned}
&\lim_{T\to+\infty} \sup_{\delta \in \Delta_{\nu_2}} \mathbb{P}^{\delta}_{x_0}\left[(\mathcal{E}_{T_0} \cap \mathrm{TS}^{\mathcal{G}}_1 \geq \nu_1) \cup \left(\bigcup_{t=T_0+1}^{T} \mathcal{E}_t \cap \overline{\mathcal{E}}_{T_0}\right)\right] \qquad \text{(B.7)} \\
&\qquad = \sup_{\delta \in \Delta_{\nu_2}} \mathbb{P}^{\delta}_{x_0}\left[(\mathcal{E}_{T_0} \cap \mathrm{TS}^{\mathcal{G}}_1 \geq \nu_1) \cup \left(\bigcup_{t=T_0+1}^{+\infty} \mathcal{E}_t \cap \overline{\mathcal{E}}_{T_0}\right)\right].
\end{aligned}$$

Indeed, from the relation (4.2) of the sets $\{\mathcal{E}_T\}_{T\in\mathbb{N}}$, it holds:

$$\bigcup_{t=T_0+1}^{T} \mathcal{E}_t \cap \overline{\mathcal{E}}_{T_0} \;\subseteq\; \bigcup_{t=T_0+1}^{+\infty} \mathcal{E}_t \cap \overline{\mathcal{E}}_{T_0}\,,$$

then, making the union with the event $\mathcal{E}_{T_0} \cap \mathrm{TS}^{\mathcal{G}}_1 \geq \nu_1$ and computing the probability distribution, the following holds:

$$\begin{aligned}
&\mathbb{P}^{\delta}_{x_0}\left[(\mathcal{E}_{T_0} \cap \mathrm{TS}^{\mathcal{G}}_1 \geq \nu_1) \cup \left(\bigcup_{t=T_0+1}^{T} \mathcal{E}_t \cap \overline{\mathcal{E}}_{T_0}\right)\right] \\
&\qquad \leq \mathbb{P}^{\delta}_{x_0}\left[(\mathcal{E}_{T_0} \cap \mathrm{TS}^{\mathcal{G}}_1 \geq \nu_1) \cup \left(\bigcup_{t=T_0+1}^{+\infty} \mathcal{E}_t \cap \overline{\mathcal{E}}_{T_0}\right)\right],
\end{aligned}$$

and then, taking now the supremum on all strategies $\delta \in \Delta_{\nu_2}$, and after the limit as $T \to +\infty$, it holds:

$$\begin{aligned}
&\lim_{T\to+\infty} \sup_{\delta \in \Delta_{\nu_2}} \mathbb{P}^{\delta}_{x_0}\left[(\mathcal{E}_{T_0} \cap \mathrm{TS}^{\mathcal{G}}_1 \geq \nu_1) \cup \left(\bigcup_{t=T_0+1}^{T} \mathcal{E}_t \cap \overline{\mathcal{E}}_{T_0}\right)\right] \qquad \text{(B.8)} \\
&\qquad \leq \sup_{\delta \in \Delta_{\nu_2}} \mathbb{P}^{\delta}_{x_0}\left[(\mathcal{E}_{T_0} \cap \mathrm{TS}^{\mathcal{G}}_1 \geq \nu_1) \cup \left(\bigcup_{t=T_0+1}^{+\infty} \mathcal{E}_t \cap \overline{\mathcal{E}}_{T_0}\right)\right].
\end{aligned}$$

Now, we prove the converse inequality to conclude (B.7). By definition of supremum, for every $\vartheta > 0$ there is $\delta_\vartheta \in \Delta_{\nu_2}$, such that:

$$\sup_{\delta\in\Delta_{\nu_2}} \mathbb{P}^{\delta}_{x_0}\Big[(\mathcal{E}_{T_0}\cap \mathrm{TS}^{\mathcal{G}}_1 \geq \nu_1)\cup\Big(\bigcup_{t=T_0+1}^{+\infty}\mathcal{E}_t\cap\overline{\mathcal{E}}_{T_0}\Big)\Big] - \vartheta$$
$$< \mathbb{P}^{\delta_\vartheta}_{x_0}\Big[(\mathcal{E}_{T_0}\cap \mathrm{TS}^{\mathcal{G}}_1 \geq \nu_1)\cup\Big(\bigcup_{t=T_0+1}^{+\infty}\mathcal{E}_t\cap\overline{\mathcal{E}}_{T_0}\Big)\Big].$$

Thus, we fix $\vartheta > 0$ and then a strategy $\delta_\vartheta \in \Delta_{\nu_2}$, we have that the goal set $\mathcal{G}$ is reached by using δ_ϑ (because it belongs to Δ_{ν_2} and by definition of the truncated sum with the cost function C_2, see Definition. 2.6.7), there exists $T_\vartheta \in \mathbb{N}$, such that for each $T \geq T_\vartheta$, it holds:

$$\sup_{\delta\in\Delta_{\nu_2}} \mathbb{P}^{\delta}_{x_0}\Big[(\mathcal{E}_{T_0}\cap \mathrm{TS}^{\mathcal{G}}_1 \geq \nu_1)\cup\Big(\bigcup_{t=T_0+1}^{+\infty}\mathcal{E}_t\cap\overline{\mathcal{E}}_{T_0}\Big)\Big] - \vartheta$$
$$< \mathbb{P}^{\delta_\vartheta}_{x_0}\Big[(\mathcal{E}_{T_0}\cap \mathrm{TS}^{\mathcal{G}}_1 \geq \nu_1)\cup\Big(\bigcup_{t=T_0+1}^{T}\mathcal{E}_t\cap\overline{\mathcal{E}}_{T_0}\Big)\Big].$$

Thus, taking the supremum on all strategies $\delta_\vartheta \in \Delta_{\nu_2}$, and after the limit at $T \to +\infty$, the following holds:

$$\sup_{\delta\in\Delta_{\nu_2}} \mathbb{P}^{\delta}_{x_0}\Big[(\mathcal{E}_{T_0}\cap \mathrm{TS}^{\mathcal{G}}_1 \geq \nu_1)\cup\Big(\bigcup_{t=T_0+1}^{+\infty}\mathcal{E}_t\cap\overline{\mathcal{E}}_{T_0}\Big)\Big] - \vartheta$$
$$\leq \lim_{T\to+\infty}\sup_{\delta_\vartheta\in\Delta_{\nu_2}} \mathbb{P}^{\delta_\vartheta}_{x_0}\Big[(\mathcal{E}_{T_0}\cap \mathrm{TS}^{\mathcal{G}}_1 \geq \nu_1)\cup\Big(\bigcup_{t=T_0+1}^{T}\mathcal{E}_t\cap\overline{\mathcal{E}}_{T_0}\Big)\Big].$$

As the latter inequality holds for every $\vartheta > 0$, the inequality of Eq. (B.8) holds and thus, the Eq. (B.7) as well. Coming back to the equality (B.6), we have:

$$\begin{aligned}
\mathbb{J}^*_{1,\infty} &= 1 - \lim_{T\to+\infty}\sup_{\delta\in\Delta_{\nu_2}} \mathbb{P}^{\delta}_{x_0}\Big[(\mathcal{E}_{T_0}\cap \mathrm{TS}^{\mathcal{G}}_1 \geq \nu_1)\cup\Big(\bigcup_{t=T_0+1}^{T}\mathcal{E}_t\cap\overline{\mathcal{E}}_{T_0}\Big)\Big]\\
&= 1 - \sup_{\delta\in\Delta_{\nu_2}} \mathbb{P}^{\delta}_{x_0}\Big[(\mathcal{E}_{T_0}\cap \mathrm{TS}^{\mathcal{G}}_1 \geq \nu_1)\cup\Big(\bigcup_{t=T_0+1}^{+\infty}\mathcal{E}_t\cap\overline{\mathcal{E}}_{T_0}\Big)\Big]\\
&\qquad \text{(by using the eq. (B.7))}\\
&= \inf_{\delta\in\Delta_{\nu_2}} \mathbb{P}^{\delta}_{x_0}[\,(\mathcal{E}_{T_0}\cap \mathrm{TS}^{\mathcal{G}}_1 < \nu_1)\cup\overline{\mathcal{E}}_\infty\,] \qquad\qquad (B.9)
\end{aligned}$$

Note that if a path

$$\omega \notin \Big(\mathcal{E}_{T_0}\cap \mathrm{TS}^{\mathcal{G}}_1 \geq \nu_1\Big)\cup\Big(\bigcup_{t=T_0+1}^{+\infty}\mathcal{E}_t\cap\overline{\mathcal{E}}_{T_0}\Big)$$

then $\omega \in (\mathcal{E}_{T_0}\cap \mathrm{TS}^{\mathcal{G}}_1 < \nu_1)\cup\overline{\mathcal{E}}_\infty$. Thus, coming back to the eq. (B.9), the following holds:

$$\begin{aligned}
\mathbb{J}^*_{1,\infty} &= \inf_{\delta\in\Delta_{\nu_2}} \mathbb{P}^{\delta}_{x_0}[\,(\mathcal{E}_{T_0}\cap \mathrm{TS}^{\mathcal{G}}_1 < \nu_1)\cup\overline{\mathcal{E}}_\infty\,]\\
&= \inf_{\delta\in\Delta_{\nu_2}} \mathbb{P}^{\delta}_{x_0}[\,\mathcal{E}_{T_0}\cap \mathrm{TS}^{\mathcal{G}}_1 < \nu_1\,]\\
&\qquad \text{(since } \mathbb{P}^{\delta}_{x_0}[\,\overline{\mathcal{E}}_\infty\,] = 0 \text{ if } \delta\in\Delta_{\nu_2}\text{, because } \mathbb{E}^{\delta}_{x_0}[\mathrm{TS}^{\mathcal{G}}_2] < \nu_2\text{)}\\
&= \mathbb{J}^*_{0,T_0}
\end{aligned}$$

Thus, $\mathbb{J}^*_{1,\infty} = \mathbb{J}^*_{0,T_0}$. Finally, by the equality (B.5), we conclude that $\mathbb{J}^*_{1,\infty} = \mathbb{J}^*_{0,\infty}$, where both limits are reached after finitely many steps.

$j = 2$| First, by Proposition 4.3.2, we deduce that for each $T \in \mathbb{N}$, it holds:

$$0 \;\leq\; \mathbb{J}^*_{1,T} \;-\; \mathbb{J}^*_{0,T} \,. \tag{B.10}$$

In the following, we will find an sequence, upper-bound for the eq. (B.10) and converging to zero. First, the following holds for any strategy $\delta \in \Delta$ such that $\mathbb{E}^{\delta}_{x_0}[\mathrm{TS}^{\mathcal{G}}_2] < \nu_2$:

$$\begin{aligned}
\mathbb{J}^*_{1,T} &= \inf_{\delta} \left\{ \mathbb{J}_{1,T}(\delta) \;\middle|\; \mathbb{E}^{\sigma}_{x_0}[\,\mathrm{TS}^{\mathcal{G}}_2\,] \;<\; \nu_2 \right\} \\
&\leq \mathbb{J}_{1,T}(\delta) \\
&= \mathbb{P}^{\delta}_{x_0}[\,(\mathcal{E}_T \cap \mathrm{TS}^{\mathcal{G}}_1 < \nu_1) \cup \overline{\mathcal{E}}_T\,] \\
&\quad \text{(by definition of } \mathbb{J}_{1,T}(\delta)\text{)} \\
&= \mathbb{P}^{\delta}_{x_0}[\,\mathcal{E}_T \cap \mathrm{TS}^{\mathcal{G}}_1 < \nu_1\,] \;+\; \mathbb{P}^{\delta}_{x_0}[\,\overline{\mathcal{E}}_T\,]
\end{aligned}$$

Thus, for any strategy δ such that $\mathbb{E}^{\delta}_{x_0}[\mathrm{TS}^{\mathcal{G}}_2] < \nu_2$, it holds:

$$\mathbb{J}^*_{1,T} \;-\; \mathbb{P}^{\delta}_{x_0}[\,\mathcal{E}_T \cap \mathrm{TS}^{\mathcal{G}}_1 < \nu_1\,] \;\;\leq\;\; \mathbb{P}^{\delta}_{x_0}[\,\overline{\mathcal{E}}_T\,] \,. \tag{B.11}$$

Now, we will find an upper-bound of $\mathbb{P}^{\delta}_{x_0}[\,\overline{\mathcal{E}}_T\,]$, because otherwise, the impact of all paths belonging to $\overline{\mathcal{E}}_T$ would be too large for the constraint of the expectation on $\mathrm{TS}^{\mathcal{G}}_2$. Applying the law of total expectation, we can write for every $T >|\,\mathcal{X}\,|$, where $\mathcal{X}$ is the set of the states of the MDP $\mathcal{M}$, and for any strategy δ satisfying $\mathbb{E}^{\delta}_{x_0}[\mathrm{TS}^{\mathcal{G}}_2] < \nu_2$, the following equality:

$$\begin{aligned}
\mathbb{E}^{\delta}_{x_0}[\,\mathrm{TS}^{\mathcal{G}}_2\,] \;=\; &\mathbb{E}^{\delta}_{x_0}[\,\mathrm{TS}^{\mathcal{G}}_2 \mid \mathcal{E}_T]\,\mathbb{P}^{\delta}_{x_0}[\mathcal{E}_T] \;+ \\
&\mathbb{E}^{\delta}_{x_0}[\,\mathrm{TS}^{\mathcal{G}}_2 \mid \overline{\mathcal{E}}_T]\,\mathbb{P}^{\delta}_{x_0}[\,\overline{\mathcal{E}}_T]
\end{aligned} \tag{B.12}$$

Let $C^{\min}_2$ the minimal (possibly negative) cost appearing in the transitions of $\mathcal{M}$ due to the cost function C_2, and let $\underline{C}^{\min}_2$ the minimal (positive by hypothesis) cost of cycles in $\mathcal{M}$ under C_2. Noticing that along any path, at most $|\,\mathcal{X}\,|$ edges may be outside any cycle, we get:

$$\mathbb{E}^{\delta}_{x_0}[\,\mathrm{TS}^{\mathcal{G}}_2 \mid \mathcal{E}_T] \;\;\geq\; |\,\mathcal{X}\,|\;C^{\min}_2 \,, \tag{B.13}$$

and

$$\mathbb{E}^{\delta}_{x_0}[\,\mathrm{TS}^{\mathcal{G}}_2 \mid \overline{\mathcal{E}}_T] \;\;\geq\; |\,\mathcal{X}\,|\;C^{\min}_2 \;+\; \frac{T-|\,\mathcal{X}\,|}{|\,\mathcal{X}\,|}\,\underline{C}^{\min}_2 \,. \tag{B.14}$$

Thus, coming back to the eq. (B.12), the following holds:

$$\begin{aligned}
\mathbb{E}^{\delta}_{x_0}[\mathrm{TS}^{\mathcal{G}}_2] &= \mathbb{E}^{\delta}_{x_0}[\mathrm{TS}^{\mathcal{G}}_2 \mid \mathcal{E}_T]\,\mathbb{P}^{\delta}_{x_0}[\mathcal{E}_T] + \mathbb{E}^{\delta}_{x_0}[\mathrm{TS}^{\mathcal{G}}_2 \mid \overline{\mathcal{E}}_T]\,\mathbb{P}^{\delta}_{x_0}[\overline{\mathcal{E}}_T] \\
&\geq \mid \mathcal{X} \mid C_2^{\min}\,\mathbb{P}^{\delta}_{x_0}[\mathcal{E}_T] + \\
&\quad \left(\mid \mathcal{X} \mid C_2^{\min} + \frac{T - \mid \mathcal{X} \mid}{\mid \mathcal{X} \mid}\, \underline{C}_2^{\min} \right) \mathbb{P}^{\delta}_{x_0}[\overline{\mathcal{E}}_T] \\
&\quad \text{(by using the eqs. (B.13) and (B.14))} \\
&= \mid \mathcal{X} \mid C_2^{\min} \left(\mathbb{P}^{\delta}_{x_0}[\mathcal{E}_T] + \mathbb{P}^{\delta}_{x_0}[\overline{\mathcal{E}}_T] \right) + \\
&\quad \frac{T - \mid \mathcal{X} \mid}{\mid \mathcal{X} \mid}\, \underline{C}_2^{\min}\,\mathbb{P}^{\delta}_{x_0}[\overline{\mathcal{E}}_T] \\
&\quad \text{(by distributing the probability into the parenthesis)} \\
&= \mid \mathcal{X} \mid C_2^{\min} + \frac{T - \mid \mathcal{X} \mid}{\mid \mathcal{X} \mid}\, \underline{C}_2^{\min}\,\mathbb{P}^{\delta}_{x_0}[\overline{\mathcal{E}}_T] \\
&\quad \text{(since } \mathbb{P}^{\delta}_{x_0}[\mathcal{E}_T] + \mathbb{P}^{\delta}_{x_0}[\overline{\mathcal{E}}_T] = 1 \text{)}
\end{aligned}$$

Thus, from the latter, it holds in summary:

$$\mathbb{E}^{\delta}_{x_0}[\mathrm{TS}^{\mathcal{G}}_2] \;\geq\; \mid \mathcal{X} \mid C_2^{\min} + \frac{T - \mid \mathcal{X} \mid}{\mid \mathcal{X} \mid}\, \underline{C}_2^{\min}\,\mathbb{P}^{\delta}_{x_0}[\overline{\mathcal{E}}_T]$$

Then,

$$\underline{C}_2^{\min}\,\mathbb{P}^{\delta}_{x_0}[\overline{\mathcal{E}}_T] \;\leq\; \left(\mathbb{E}^{\delta}_{x_0}[\mathrm{TS}^{\mathcal{G}}_2] - \mid \mathcal{X} \mid C_2^{\min} \right) \frac{\mid \mathcal{X} \mid}{T - \mid \mathcal{X} \mid}\,.$$

In addition, because by hypothesis δ is such that $\mathbb{E}^{\delta}_{x_0}[\mathrm{TS}^{\mathcal{G}}_2] < \nu_2$, we have that for every $T > \mid \mathcal{X} \mid$,

$$\mathbb{P}^{\delta}_{x_0}[\overline{\mathcal{E}}_T] \;<\; \left\lceil \frac{\nu_2 - \mid \mathcal{X} \mid C_2^{\min}}{\underline{C}_2^{\min}} \right\rceil \frac{\mid \mathcal{X} \mid}{T - \mid \mathcal{X} \mid}\,.$$

Thus, coming back to the eq. (B.11), the following holds:

$$\mathbb{J}^*_{1,T} - \mathbb{P}^{\delta}_{x_0}[\,\mathcal{E}_T \cap \mathrm{TS}^{\mathcal{G}}_1 < \nu_1\,] \;<\; \left\lceil \frac{\nu_2 - \mid \mathcal{X} \mid C_2^{\min}}{\underline{C}_2^{\min}} \right\rceil \frac{\mid \mathcal{X} \mid}{T - \mid \mathcal{X} \mid}\,.$$

Taking above the minimum on all the strategies δ such that $\mathbb{E}^{\delta}_{s_0}[\mathrm{TS}^{\mathcal{G}}_2] < \nu_2$ and noticing that the right side is converging to zero as $T \to +\infty$, we conclude that $\mathbb{J}^*_{1,\infty} = \mathbb{J}^*_{0,\infty}$.

■

Proof B.1.6: [Theorem 4.5.3]

There exists a solution to $\mathbf{Pb}(0)$ if, and only if, there is a strategy δ^* in the unfolding $\mathcal{M}_{T^{\max}}$, such that each $\omega \in \Omega^{\delta^*}_{s_0}$ reaches $\mathcal{G}_{\nu_1}$ and $\mathbb{E}^{\delta^*}_{s_0}[\,\mathrm{TS}^{\mathcal{G}_{\nu_1}}\,] < \nu_2$.

Proof. $\underline{\Rightarrow}|$ Assume that there exists a strategy δ^*_0 solution of the problem $\mathbf{Pb}(0)$. We will construct a strategy in $\mathcal{M}_{T^{\max}}$ satisfying the reachability condition and the constraint in the expectation of the theorem.

Let $\overline{\omega}_t = s_0 a_0 s_1 \ldots s_t$ a finite path of length $t \in \mathbb{N}_0$ in the unfolding $\mathcal{M}_{T^{\max}}$. By definition, there exists a unique correspondence of $\overline{\omega}_t$ with a path in $\mathcal{M}$. We express the latter path as:

$$\omega_t \quad = \quad \mathrm{proj}_1(s_0)\, \mathrm{id}_{\mathcal{A}}(a_0)\, \mathrm{proj}_1(s_1) \ldots \mathrm{proj}_1(s_t)\ .$$

Let now $x_{t+1} \in \mathcal{X}$ be a state in $\mathcal{M}$, such that a transition from the state $\mathrm{proj}_1(s_t)$ to x_{t+1} takes place using the strategy δ_0^*, i.e., from the definition of eq. (2.28),

$$0 \quad < \quad \sum_{a_t \in \mathcal{A}} \delta_0^*(\omega_t)(a_t) P(\mathrm{proj}_1(s_t), a_t)(x_{t+1})\ . \tag{B.15}$$

For each possible action $a_t \in \mathcal{A}$, that can be selected in the current state, we define the strategy δ^* in $\mathcal{M}_{T^{\max}}$ as following:

$$\delta^*(\overline{\omega}_t)(a_t) \quad := \quad \delta_0^*(\omega_t)(a_t)$$

Note that the strategy δ^* is well-defined as δ_0^* is. In the following, we show that the strategy δ^* satisfies the reachability condition and the constraint in the expectation of the theorem.

Let $\overline{\omega} \in \Omega_{s_0}^{\delta^*}$ be an infinite path in $\mathcal{M}_{T^{\max}}$. Towards a contradiction, suppose that $\overline{\omega}$ is not reaching the goal set $\mathcal{G}_{\nu_1}$. Thus, writing such a path by $\overline{\omega} = s_0 a_0 s_1 \ldots$, it follows that each state s_t belonging to $\overline{\omega}$ is such that $s_t \notin \mathcal{G}_{\nu_1}$ for each $t \in \mathbb{N}_0$. From the definition of $\mathcal{G}_{\nu_1}$, see eq. (4.14), it is the same as saying for each $t \in \mathbb{N}_0$,

$$\mathrm{proj}_1(s_t) \notin \mathcal{G} \qquad \text{or} \qquad \mathrm{proj}_2(s_t) < \nu_1\ . \tag{B.16}$$

On the other hand, by the definition of δ^* and the correspondence between paths of $\mathcal{M}$ and $\mathcal{M}_{T^{\max}}$, the path

$$\omega \quad = \quad \mathrm{proj}_1(s_0)\, \mathrm{id}_{\mathcal{A}}(a_0)\, \mathrm{proj}_1(s_1) \ldots \tag{B.17}$$

is such that $\omega \in \Omega_{x_0}^{\delta_0^*}$, where $x_0 = \mathrm{proj}_1(s_0)$. On the other hand, since by hypothesis δ_0^* is a solution of the problem $\mathbf{Pb}(0)$ in $\mathcal{M}$, then each path reaches $\mathcal{G}$,

$$\mathbb{P}_{x_0}^{\delta_0^*}[\,\mathrm{TS}_1^{\mathcal{G}} \geq \nu_1\,] \;\geq\; 1 \qquad \text{and} \qquad \mathbb{E}_{x_0}^{\delta_0^*}[\,\mathrm{TS}_2^{\mathcal{G}}\,] \;<\; \nu_2\ ,$$

but it holds for each path $\Omega_{x_0}^{\delta_0^*}$, in particular for the path ω of (B.17). But, it is an immediate contradiction with (B.16). Thus, we conclude that each $\overline{\omega} \in \Omega_{s_0}^{\delta^*}$ reaches $\mathcal{G}_{\nu_1}$.

The constraint in expectation is verified analogously, because the expectation of a path ω in $\mathcal{M}$ is the same (by definition) as the one of $\overline{\omega}$ in the unfolding $\mathcal{M}_{T^{\max}}$.

$\Leftarrow|$ Assume that there exists a strategy δ^* in $\mathcal{M}_{T^{\max}}$, such that each path $\overline{\omega} \in \Omega_{s_0}^{\delta^*}$ reaches the goal set $\mathcal{G}_{\nu_1}$ and $\mathbb{E}_{s_0}^{\delta^*}[\mathrm{TS}^{\mathcal{G}_{\nu_1}}] < \nu_2$. We will construct a strategy in $\mathcal{M}$ satisfying the reachability condition to $\mathcal{G}$, and the constraints on the thresholds ν_1 and ν_2 for the problem $\mathbf{Pb}(0)$.

Let $\omega_t = x_0 a_0 x_1 \ldots x_t$ be a finite path of length $t \in \mathbb{N}$ in $\mathcal{M}$. By definition, there exists a unique correspondence of ω_t with a path in $\mathcal{M}_{T^{\max}}$. We express the latter path as:

$$\overline{\omega}_t \quad = \quad s_0\, a_0\, a_1 \ldots s_t$$

where $s_\tau = (x_\tau, y_\tau, \tau)$ for each $\tau = 0, ..., t$; whit $y_0 = 0$ and for $\tau = 1, ..., t$,

$$y_\tau = \min\left\{\nu_1,\ y_{\tau-1} + C_1(x_{\tau-1}, a_{\tau-1}, x_\tau)\right\} .$$

Let now $s_{t+1} \in \mathcal{S}_{T^{\max}}$ a state in $\mathcal{M}_{T^{\max}}$ such that a transition from the state s_t to s_{t+1} takes place using the strategy δ^*, i.e., from the definition of eq. (2.28),

$$0 < \sum_{a_t \in \mathcal{A}} \delta^*(\overline{\omega}_t)(a_t) P(s_t, a_t)(s_{t+1}) . \tag{B.18}$$

For each possible action $a_t \in \mathcal{A}$, that can be selected in the current state, we define the strategy δ_0^* in $\mathcal{M}$ as following:

$$\delta_0^*(\omega_t)(a_t) := \delta^*(\overline{\omega}_t)(a_t) .$$

Note that the strategy δ_0^* is well-defined as δ^* is. In the following, we show that the strategy δ_0^* satisfies the reachability condition to $\mathcal{G}$, and the constraints on the thresholds ν_1 and ν_2 for the problem $\mathbf{Pb}(0)$.

Let $\omega \in \Omega_{x_0}^{\delta_0^*}$ an infinite path in $\mathcal{M}$. Towards a contradiction, suppose that ω is not reaching the goal set $\mathcal{G}$. Thus, writing such a path by $\omega = x_0 a_0 x_1 ...$, it follows that each state x_t belonging to the path ω is such that

$$x_t \notin \mathcal{G} , \tag{B.19}$$

for each $t \in \mathbb{N}_0$. On the other hand, by the definition of δ_0^* and the correspondence between paths of $\mathcal{M}_{T^{\max}}$ and $\mathcal{M}$, the path

$$\overline{\omega} = s_0\, a_0\, a_1 \dots \tag{B.20}$$

where $s_t = (x_t, y_t, t)$ for each $t \in \mathbb{N}_0$; whit $y_0 = 0$ and,

$$y_t = \min\left\{\nu_1,\ y_{t-1} + C_1(x_{t-1}, a_{t-1}, x_t)\right\} ;$$

is such that $\overline{\omega} \in \Omega_{s_0}^{\delta^*}$. On the other hand, since by hypothesis δ^* is such that each path in $\Omega_{s_0}^{\delta^*}$ reaches the goal set $\mathcal{G}_{\nu_1}$ and $\mathbb{E}_{s_0}^{\delta^*}[\mathrm{TS}^{\mathcal{G}_{\nu_1}}] < \nu_2$, then in particular for the path (B.20), the reachability time $T_\mathcal{G}$ is finite, see eq. (2.23), for which the path $\overline{\omega}$ reaches some $x \in \mathcal{G}$. Thus, for such time-step, $T = T_\mathcal{G}(\overline{\omega})$, it follows that the state $s_T \in \mathcal{G}_{\nu_1}$. But, the latter is the same as saying that

$$x_T \in \mathcal{G} \quad \text{and} \quad y_T = \nu_1 . \tag{B.21}$$

But first, it is a contradiction with (B.19). Then, ω reaches $\mathcal{G}$. Second, because the the second component of the states of $\overline{\omega}$ can be written for each $t \in \mathbb{N}$:

$$y_t = \sum_{t=0}^{t} C_1(x_{t-1}, a_{t-1}, x_t) ,$$

thus we conclude, by the second equation in (B.21), that $\mathbb{P}_{x_0}^{\delta_0^*}[\mathrm{TS}_1^\mathcal{G} \geq \nu_1] \geq 1$. Finally, the constraint in expectation is verified analogously, because the expectation of a path ω in $\mathcal{M}$ is the same (by definition) as the one of $\overline{\omega}$ in the unfolding $\mathcal{M}_{T^{\max}}$, and here $\mathbb{E}_{s_0}^{\delta^*}[\mathrm{TS}^{\mathcal{G}_{\nu_1}}] < \nu_2$. ■

2.2 PROOF OF CHAPTER 5

In this Section B.2, we provide the main proofs of the Chapter 5.

Proof B.2.1: [Proposition 5.3.1]

The expectation $\mathbb{E}_{s_0}^{\sigma}[\mathrm{TS}^{\mathcal{G}}]$ and $\mathbb{J}_{\alpha,T}(\sigma)$ are continuous for each $\sigma \in \Delta[\underline{\Pi}_\gamma]$.

Proof. Let $\sigma \in \Delta[\underline{\Pi}_\gamma]$ a mixed strategy. By eqs. (5.18) and (5.21), the expectation $\mathbb{E}_{s_0}^{\sigma}[\mathrm{TS}^{\mathcal{G}}]$ can be written as a convex combination of the expectation of pure strategies, i.e., as:

$$\mathbb{E}_{s_0}^{\sigma}[\mathrm{TS}^{\mathcal{G}}] = \sum_{k=1}^{K} \sigma(\overline{\pi}_k)\, \mathbb{E}_{s_0}^{\overline{\pi}_k}[\mathrm{TS}^{\mathcal{G}}]$$

Then, the expectation is clearly linear in σ, and therefore continuous in any one. For $\mathbb{J}_{\alpha,T}(\sigma)$, it is similar. This can be written as:

$$\mathbb{J}_{\alpha,T}(\sigma) = \sum_{k=1}^{K} \sigma(\overline{\pi}_k)\, \mathbb{J}_{\alpha,T}(\overline{\pi}_k)$$

and the reasoning is the same. ■

Proof B.2.2: [Proposition 5.3.2]

Mixed strategies always provide optimal values that are at least as good as the values obtained with pure strategies, i.e.,

$$\mathbb{J}_{\alpha,T}^{\mathrm{m}*} \leq \mathbb{J}_{\alpha,T}^{\mathrm{p}*}$$

Proof. The proof follows from the following inequality:

$$\begin{aligned} \mathbb{J}_{\alpha,T}^{\mathrm{m}*} &= \inf_{\sigma} \left\{ \mathbb{J}_{\alpha,T}(\sigma) \;\middle|\; \mathbb{E}_{s_0}^{\sigma}[\mathrm{TS}^{\mathcal{G}}] < \nu \right\} \\ &\leq \mathbb{J}_{\alpha,T}(\sigma) \end{aligned}$$

But the latter is true for every mixed strategy σ such that the constraint $\mathbb{E}_{s_0}^{\sigma}[\mathrm{TS}^{\mathcal{G}}] < \nu$ is satisfied, in particular for the strategies assigning probability one to pure strategies $\overline{\pi} \in \underline{\Pi}_\gamma$. In this way, we can write:

$$\mathbb{J}_{\alpha,T}^{\mathrm{m}*} \leq \mathbb{J}_{\alpha,T}(\overline{\pi}) ,$$

such that $\mathbb{E}_{s_0}^{\overline{\pi}}[\mathrm{TS}^{\mathcal{G}}] < \nu$. Taking now the minimum on all pure strategies that satisfy the latter constraint in expectation, it holds:

$$\begin{aligned} \mathbb{J}_{\alpha,T}^{\mathrm{m}*} &\leq \min_{\overline{\pi}} \left\{ \mathbb{J}_{\alpha,T}(\overline{\pi}) \;\middle|\; \mathbb{E}_{s_0}^{\overline{\pi}}[\mathrm{TS}^{\mathcal{G}}] < \nu \right\} \\ &= \mathbb{J}_{\alpha,T}^{\mathrm{p}*} \end{aligned}$$

■

Proof B.2.3: [**Lemma 5.4.1**]

Suppose that we can obtain $\overline{\pi}^*_{\alpha,\lambda} \in \arg\min_{\overline{\pi}} \mathbb{L}_{\alpha,T}(\overline{\pi},\lambda)$ for each $\lambda \geq 0$. Then, $\lambda \mapsto \mathbb{E}^{\overline{\pi}^*_{\alpha,\lambda}}_{s_0}[\mathrm{TS}^{\mathcal{G}}]$ is nonincreasing.

Proof. Using ideas from [27], it follows that for $\lambda_1 \geq 0$ and $\lambda_2 > \lambda_1$,

$$\mathbb{L}_{\alpha,T}(\overline{\pi}^*_{\alpha,\lambda_1+\lambda_2}, \lambda_1+\lambda_2) \;=\; \min_{\overline{\pi}} \mathbb{L}_{\alpha,T}(\overline{\pi}, \lambda_1+\lambda_2) \;\leq\; \mathbb{L}_{\alpha,T}(\overline{\pi}, \lambda_1+\lambda_2)$$

where the last inequality holds $\forall \overline{\pi} \in \underline{\Pi}_\gamma$, in particular for $\overline{\pi}^*_{\alpha,\lambda_1}$, then

$$\mathbb{L}_{\alpha,T}(\overline{\pi}^*_{\alpha,\lambda_1+\lambda_2}, \lambda_1+\lambda_2) \;\leq\; \mathbb{L}_{\alpha,T}(\overline{\pi}^*_{\alpha,\lambda_1}, \lambda_1+\lambda_2)\;. \tag{B.22}$$

On the other hand, it follows that:

$$\mathbb{L}_{\alpha,T}(\overline{\pi}^*_{\alpha,\lambda_1}, \lambda_1) \;=\; \min_{\overline{\pi}} \mathbb{L}_{\alpha,T}(\overline{\pi}, \lambda_1) \;\leq\; \mathbb{L}_{\alpha,T}(\overline{\pi}, \lambda_1)\;.$$

Again, the last inequality holds $\forall \overline{\pi} \in \underline{\Pi}_\gamma$, in particular for $\overline{\pi}^*_{\alpha,\lambda_1+\lambda_2}$, then:

$$\mathbb{L}_{\alpha,T}(\overline{\pi}^*_{\alpha,\lambda_1}, \lambda_1) \;\leq\; \mathbb{L}_{\alpha,T}(\overline{\pi}^*_{\alpha,\lambda_1+\lambda_2}, \lambda_1)\;. \tag{B.23}$$

Thus, we have by taking difference of (B.22) and (B.23) that:

$$\begin{aligned}\mathbb{L}_{\alpha,T}(\overline{\pi}^*_{\alpha,\lambda_1}, \lambda_1+\lambda_2) \;-\; \mathbb{L}_{\alpha,T}(\overline{\pi}^*_{\alpha,\lambda_1}, \lambda_1)&\\ \geq\; \mathbb{L}_{\alpha,T}(\overline{\pi}^*_{\alpha,\lambda_1+\lambda_2}, \lambda_1+\lambda_2) \;-\; \mathbb{L}_{\alpha,T}(\overline{\pi}^*_{\alpha,\lambda_1+\lambda_2}, \lambda_1)\;.&\end{aligned}$$

Now, by the definition of the Lagrange function, see eq. (5.22), it is easy to see that the above inequalities are the same as:

$$\lambda_2\Big(\mathbb{E}^{\overline{\pi}^*_{\alpha,\lambda_1}}_{s_0}[\,\mathrm{TS}^{\mathcal{G}}] - \nu\Big) \;\geq\; \lambda_2\Big(\mathbb{E}^{\overline{\pi}^*_{\alpha,\lambda_1+\lambda_2}}_{s_0}[\,\mathrm{TS}^{\mathcal{G}}] - \nu\Big)\;.$$

Hence,

$$\mathbb{E}^{\overline{\pi}^*_{\alpha,\lambda_1}}_{s_0}[\,\mathrm{TS}^{\mathcal{G}}] \;\geq\; \mathbb{E}^{\overline{\pi}^*_{\alpha,\lambda_1+\lambda_2}}_{s_0}[\,\mathrm{TS}^{\mathcal{G}}]$$

for any $\lambda_2 > \lambda_1 \geq 0$. Thus, $\lambda \mapsto \mathbb{E}^{\overline{\pi}^*_{\alpha,\lambda}}_{s_0}[\,\mathrm{TS}^{\mathcal{G}}]$ is nonincreasing. ∎

Proof B.2.4: [**Proposition 5.4.2**]

The Lagrange functions $\lambda \mapsto \mathbb{L}_{\alpha,T}(\overline{\pi},\lambda)$ and $\lambda \mapsto \mathbb{L}_{\alpha,T}(\sigma,\lambda)$ are linear on $\lambda \geq 0$. Moreover, the Lagrange dual functions $\lambda \mapsto \mathbb{L}^{\mathrm{p}}_{\alpha,T}(\lambda)$ and $\lambda \mapsto \mathbb{L}^{\mathrm{m}}_{\alpha,T}(\lambda)$ are concave.

Proof. The proof is similar for the two cases. So, we prove the proposition for mixed strategies. Let $\lambda_1, \lambda_2 \geq 0$, $\theta \in [0,1]$ and $\sigma \in \Delta[\,\underline{\Pi}_\gamma\,]$. The following holds:

$$\begin{aligned}\mathbb{L}_{\alpha,T}(\sigma, \theta\lambda_1 + (1-\theta)\lambda_2) &= \mathbb{J}_{\alpha,T}(\sigma) + (\theta\lambda_1 + (1-\theta)\lambda_2)(\mathbb{E}^{\sigma}_{s_0}[\mathrm{TS}^{\mathcal{G}}] - \nu)\\ &\quad \text{(it holds by definition, see eq. (5.22))}\\ &= \theta\Big(\mathbb{J}_{\alpha,T}(\sigma) + \lambda_1(\mathbb{E}^{\sigma}_{s_0}[\mathrm{TS}^{\mathcal{G}}] - \nu)\Big)\\ &\quad + (1-\theta)\Big(\mathbb{J}_{\alpha,T}(\sigma) + \lambda_2(\mathbb{E}^{\sigma}_{s_0}[\mathrm{TS}^{\mathcal{G}}] - \nu)\Big)\\ &\quad \text{(by adding a convenient zero)}\\ &= \theta\,\mathbb{L}_{\alpha,T}(\sigma,\lambda_1) + (1-\theta)\,\mathbb{L}_{\alpha,T}(\sigma,\lambda_2)\\ &\quad \text{(again by definition of } \mathbb{L}_{\alpha,T}(\sigma,\lambda)\text{)}\end{aligned}$$

Thus, in summary, it holds:

$$\mathbb{L}_{\alpha,T}(\sigma, \theta\lambda_1 + (1-\theta)\lambda_2) \; = \; \theta\, \mathbb{L}_{\alpha,T}(\sigma, \lambda_1) + (1-\theta)\, \mathbb{L}_{\alpha,T}(\sigma, \lambda_2)\;. \tag{B.24}$$

Based on the latter, the Lagrange function is therefore affine on $\lambda \geq 0$ and then, concave and convex on $\lambda \geq 0$. Now, applying the infimum on (B.24) over all mixed strategies, we have that:

$$\begin{aligned}
\inf_{\sigma} \mathbb{L}_{\alpha,T}(\sigma, \theta\lambda_1 + (1-\theta)\lambda_2) \; &= \inf_{\sigma} \left\{ \theta\, \mathbb{L}_{\alpha,T}(\sigma, \lambda_1) \; + \; (1-\theta)\, \mathbb{L}_{\alpha,T}(\sigma, \lambda_2) \right\} \\
&\geq \theta \inf_{\sigma} \mathbb{L}_{\alpha,T}(\sigma, \lambda_1) \; + \; (1-\theta) \inf_{\sigma} \mathbb{L}_{\alpha,T}(\sigma, \lambda_2) \\
&\quad \text{(by a property of the infimum)} \\
&= \theta\, \mathbb{L}^{\mathrm{m}}_{\alpha,T}(\lambda_1) \; + \; (1-\theta)\, \mathbb{L}^{\mathrm{m}}_{\alpha,T}(\lambda_2) \\
&\quad \text{(by definition, see eq. (5.24))}
\end{aligned}$$

Noticing that the left side above is the definition of the Lagrange dual function (5.24) for mixed strategies with $\lambda = \theta\lambda_1 + (1-\theta)\lambda_2$, we conclude that:

$$\mathbb{L}^{\mathrm{m}}_{\alpha,T}(\theta\lambda_1 + (1-\theta)\lambda_2) \;\; \geq \;\; \theta\, \mathbb{L}^{\mathrm{m}}_{\alpha,T}(\lambda_1) \; + \; (1-\theta)\, \mathbb{L}^{\mathrm{m}}_{\alpha,T}(\lambda_2)$$

Thus, the Lagrange dual function is concave. Note that this is also true for pure strategies. ■

Proof B.2.5: [Lemma 5.4.3]

Let $\mathbb{J}^{\mathrm{p}*}_{\alpha,T}$ and $\mathbb{J}^{\mathrm{m}*}_{\alpha,T}$ resp. the optimal values of the P-PS and P-MS problem, see eq. (5.14) and eq. (5.20). Then,

$$\mathbb{L}^{\mathrm{p}*}_{\alpha,T} \;\leq\; \mathbb{J}^{\mathrm{p}*}_{\alpha,T} \qquad \text{and} \qquad \mathbb{L}^{\mathrm{m}*}_{\alpha,T} \;\leq\; \mathbb{J}^{\mathrm{m}*}_{\alpha,T} \tag{5.27}$$

Proof. From the definition of the Lagrange dual functions (5.24), we have that for every $\lambda \geq 0$,

$$\begin{aligned}
\mathbb{L}^{\mathrm{p}}_{\alpha,T}(\lambda) \; &= \min_{\overline{\pi}} \;\; \mathbb{L}_{\alpha,T}(\overline{\pi}, \lambda) \\
&= \min_{\overline{\pi}} \left\{ \mathbb{J}_{\alpha,T}(\overline{\pi}) \; + \; \lambda(\mathbb{E}^{\overline{\pi}}_{s_0}[\mathrm{TS}^{\mathcal{G}}] \; - \; \nu) \right\} \\
&\quad \text{(by definition (5.22) of the Lagrange function)} \\
&\leq \min_{\overline{\pi}} \left\{ \mathbb{J}_{\alpha,T}(\overline{\pi}) \; + \; \lambda(\mathbb{E}^{\overline{\pi}}_{s_0}[\mathrm{TS}^{\mathcal{G}}] \; - \; \nu) \;\; \middle| \;\; \mathbb{E}^{\overline{\pi}}_{s_0}[\mathrm{TS}^{\mathcal{G}}] < \nu \right\} \\
&\leq \min_{\overline{\pi}} \left\{ \mathbb{J}_{\alpha,T}(\overline{\pi}) \;\; \middle| \;\; \mathbb{E}^{\overline{\pi}}_{s_0}[\mathrm{TS}^{\mathcal{G}}] < \nu \right\} \\
&\quad \text{(because } \lambda(\mathbb{E}^{\overline{\pi}}_{s_0}[\mathrm{TS}^{\mathcal{G}}] - \nu) \leq 0 \text{)} \\
&= \;\; \mathbb{J}^{\mathrm{p}*}_{\alpha,T} \\
&\quad \text{(by definition (5.14) of the pure strategy optimal value)}
\end{aligned}$$

applying now the supremum over all $\lambda \geq 0$, we have the first inequality, i.e., $\mathbb{L}^{\mathrm{p}*}_{\alpha,T} \leq \mathbb{J}^{\mathrm{p}*}_{\alpha,T}$. The proof for $\mathbb{L}^{\mathrm{m}*}_{\alpha,T} \leq \mathbb{J}^{\mathrm{m}*}_{\alpha,T}$ is completely analogous. ■

Proof B.2.6: [**Proposition 5.4.4**]

There is strong duality for mixed strategy problems D-MS and P-MS, i.e.,

$$\mathbb{L}^{\mathrm{m}*}_{\alpha,T} \quad = \quad \mathbb{J}^{\mathrm{m}*}_{\alpha,T} \; . \tag{5.28}$$

Moreover, for $\lambda \geq 0$ fixed, there exists $\sigma^*_{\alpha,\lambda} \in \Delta[\,\underline{\Pi}_\gamma]$ such that:

$$\mathbb{L}_{\alpha,T}(\sigma^*_{\alpha,\lambda}, \lambda) \quad = \quad \mathbb{L}^{\mathrm{m}}_{\alpha,T}(\lambda) \; . \tag{5.29}$$

Proof. To prove this proposition, we use the following theorem [9]. First, we prove the eq. (5.28).

Remark 2.2.7: MinMax Theorem

Let X and Y be convex subsets of linear topological spaces, with X compact. Let f a function, such that $f : X \times Y \longrightarrow \mathbb{R}$ and

- $\forall\, y \in Y, \quad x \mapsto f(x, y)$ is convex and lower semi-continuous.
- $\forall\, x \in X, \quad y \mapsto f(x, y)$ is concave.

Then, $\exists\, x^* \in X$ such that:

$$\inf_{x \in X} \sup_{y \in Y} f(x, y) \; = \; \sup_{y \in Y} f(x^*, y) \; = \; \sup_{y \in Y} \inf_{x \in X} f(x, y)$$

Considering the theorem for the problems D-MS and P-MS of mixed strategies, we have:

- The set X stands for the set $\Delta[\,\underline{\Pi}_\gamma]$, which is convex and compact, see eq. (5.16).
- The set Y stands for the convex set $\mathbb{R}^+_0$.
- f stands for the Lagrange function $\mathbb{L}_{\alpha,T} : \Delta[\,\underline{\Pi}_\gamma\,] \times \mathbb{R}^+_0 \to \mathbb{R}$ for mixed strategies, see eq. (5.22).
- $\forall\, \lambda \in \mathbb{R}^+_0, \quad \sigma \mapsto \mathbb{L}_{\alpha,T}(\sigma, \lambda)$ is continuous and linear in σ, because by Proposition 5.3.1 and the P-MS problem (5.21), the expectation and $\mathbb{J}_{\alpha,T}(\sigma)$ are continuous and linear in σ. Thus, $\sigma \mapsto \mathbb{L}_{\alpha,T}(\sigma, \lambda)$ is also convex on σ.
- $\forall\, \sigma \in \Delta[\,\underline{\Pi}_\gamma],\quad \lambda \mapsto \mathbb{L}_{\alpha,T}(\sigma, \lambda)$ is affine in λ and then concave in $\lambda \in \mathbb{R}^+_0$, see Proposition 5.4.2.

Thus, applying the MinMax Theorem in such a context, there exists a mixed strategy $\sigma^*_{\alpha,\lambda} \in \Delta[\underline{\Pi}_\gamma]$, such that the following holds:

$$
\begin{aligned}
\mathbb{L}^{m*}_{\alpha,T} &= \sup_{\lambda \geq 0} \mathbb{L}^{m}_{\alpha,T}(\lambda) \\
&\quad \text{(by definition (5.26) of the dual optimal value)} \\
&= \sup_{\lambda \geq 0} \inf_{\sigma} \mathbb{L}_{\alpha,T}(\sigma,\lambda) \\
&\quad \text{(by definition (5.24) of the Lagrange dual function)} \\
&= \sup_{\lambda \geq 0} \mathbb{L}_{\alpha,T}(\sigma^*_{\alpha,\lambda},\lambda) \\
&\quad \text{(by the existence of } \sigma^*_{\alpha,\lambda} \text{ from the MinMax Theorem)} \\
&= \sup_{\lambda \geq 0} \left\{ \mathbb{J}_{\alpha,T}(\sigma^*_{\alpha,\lambda}) + \lambda\left(\mathbb{E}^{\sigma^*_{\alpha,\lambda}}_{s_0}[\mathrm{TS}^{\mathcal{G}}] - \nu \right) \right\} \\
&\quad \text{(by definition (5.22) of the Lagrange function)} \\
&= \begin{cases} \mathbb{J}_{\alpha,T}(\sigma^*_{\alpha,\lambda}) & \text{if} \quad \mathbb{E}^{\sigma^*_{\alpha,\lambda}}_{s_0}[\mathrm{TS}^{\mathcal{G}}] < \nu \\ +\infty & \text{otherwise} \end{cases} \\
&\geq \mathbb{J}_{\alpha,T}(\sigma^*_{\alpha,\lambda}) \\
&\geq \mathbb{J}^{m*}_{\alpha,T} \\
&\quad \text{(by definition (5.20) of the mixed strategy optimal value)}
\end{aligned}
$$

In this way, we have that $\mathbb{L}^{m*}_{\alpha,T} \geq \mathbb{J}^{m*}_{\alpha,T}$. On the other hand, since the weak duality holds, see Lemma 5.4.3, we conclude that $\mathbb{L}^{m*}_{\alpha,T} = \mathbb{J}^{m*}_{\alpha,T}$.

Second, we prove the eq. (5.29) of the proposition. Let $\lambda \geq 0$ and $(\sigma_n)_{n\in\mathbb{N}}$ a sequence of mixed strategies in $\Delta[\underline{\Pi}_\gamma]$, such that:

$$
\lim_{n\to+\infty} \mathbb{L}_{\alpha,T}(\sigma_n,\lambda) = \mathbb{L}^{m}_{\alpha,T}(\lambda) \,. \tag{B.25}
$$

Now, because $\Delta[\underline{\Pi}_\gamma]$ is compact, there is a subsequence $(\sigma_{n_j})_{j\in\mathbb{N}}$ of $(\sigma_n)_{n\in\mathbb{N}}$ converging to some $\sigma_0 \in \Delta[\underline{\Pi}_\gamma]$. Since $\sigma \mapsto \mathbb{L}_{\alpha,T}(\sigma,\lambda)$ is lower semi-continuous, in particular for σ_0, the following holds:

$$
\mathbb{L}_{\alpha,T}(\sigma_0,\lambda) \leq \lim_{j\to+\infty} \mathbb{L}_{\alpha,T}(\sigma_{n_j},\lambda) \,.
$$

By eq. (B.25), we also have that:

$$
\lim_{j\to+\infty} \mathbb{L}_{\alpha,T}(\sigma_{n_j},\lambda) = \mathbb{L}^{m}_{\alpha,T}(\lambda) \,.
$$

Thus,

$$
\mathbb{L}_{\alpha,T}(\sigma_0,\lambda) \leq \mathbb{L}^{m}_{\alpha,T}(\lambda) \,.
$$

But by definition, $\mathbb{L}^{m}_{\alpha,T}(\lambda) = \inf_{\sigma} \mathbb{L}_{\alpha,T}(\sigma,\lambda)$. Hence,

$$
\mathbb{L}_{\alpha,T}(\sigma_0,\lambda) = \inf_{\sigma} \mathbb{L}_{\alpha,T}(\sigma,\lambda) \,,
$$

which proves the existence of a mixed strategy for the eq. (5.29). ∎

Proof B.2.8: [Proposition 5.4.5]

Consider the dual problems D-PS and D-MS of (5.25). Then,

$$\mathbb{L}^{p*}_{\alpha,T} \quad = \quad \mathbb{L}^{m*}_{\alpha,T}$$

Proof. Since mixed strategies extend the solution set, as we have seen in Proposition 5.3.2, then we have that for $\lambda \geq 0$ fixed:

$$\inf_{\sigma} \ \mathbb{L}_{\alpha,T}(\sigma,\lambda) \quad \leq \quad \min_{\overline{\pi}} \ \mathbb{L}_{\alpha,T}(\overline{\pi},\lambda) \ .$$

Noticing that each side above corresponds to the Lagrange dual function of eq. (5.24) and so, applying supremum on $\lambda \geq 0$, it follows that:

$$\sup_{\lambda\geq 0} \ \mathbb{L}^{m}_{\alpha,T}(\lambda) \quad \leq \quad \sup_{\lambda\geq 0} \ \mathbb{L}^{p}_{\alpha,T}(\lambda) \ . \tag{B.26}$$

On the other hand, for $\sigma \in \Delta[\,\underline{\Pi}_{\gamma}]$ and $\lambda \geq 0$ fixed, the Lagrange function $\mathbb{L}_{\alpha,T}(\sigma,\lambda)$ can be rewritten based on eq. (5.18) as:

$$\begin{aligned}
\mathbb{L}_{\alpha,T}(\sigma,\lambda) \quad &= \quad \mathbb{J}_{\alpha,T}(\sigma) + \lambda(\,\mathbb{E}^{\sigma}_{s_0}[\mathrm{TS}^{\mathcal{G}}] \,-\, \nu\,) \\
&= \quad \sum_{k=1}^{K} \sigma(\overline{\pi}_k)\Big(\mathbb{J}_{\alpha,T}(\overline{\pi}_k) + \lambda\Big(\,\mathbb{E}^{\overline{\pi}_k}_{s_0}[\mathrm{TS}^{\mathcal{G}}] \,-\, \nu\,\Big)\Big) \\
&= \quad \sum_{k=1}^{K} \sigma(\overline{\pi}_k)\ \mathbb{L}_{\alpha,T}(\overline{\pi}_k,\lambda)
\end{aligned} \tag{B.27}$$

Since $\sigma(\overline{\pi}_1),\ldots,\sigma(\overline{\pi}_K)$ are marginal probabilities of the pure strategies $\overline{\pi}_1,\ldots,\overline{\pi}_K$ resp., the above equality allows us to infer that:

$$\mathbb{L}_{\alpha,T}(\sigma,\lambda) \quad \geq \quad \min_{\overline{\pi}} \ \mathbb{L}_{\alpha,T}(\overline{\pi},\lambda) \quad = \quad \mathbb{L}^{p}_{\alpha,T}(\lambda) \ . \tag{B.28}$$

Applying now the infimum on the strategies $\sigma \in \Delta[\,\underline{\Pi}_{\gamma}\,]$, we obtain $\mathbb{L}^{m}_{\alpha,T}(\lambda)$ on the left side of the above inequality, and then applying supremum on $\lambda \geq 0$, we have that:

$$\sup_{\lambda\geq 0} \ \mathbb{L}^{m}_{\alpha,T}(\lambda) \quad \geq \quad \sup_{\lambda\geq 0} \ \mathbb{L}^{p}_{\alpha,T}(\lambda) \ . \tag{B.29}$$

Thus, by eqs. (B.26) and (B.29), the equality $\mathbb{L}^{p*}_{\alpha,T} \;=\; \mathbb{L}^{m*}_{\alpha,T}$ holds. ■

Proof B.2.9: [Proposition 5.4.6]

Let $\sigma^*_{\alpha} \in \Delta[\,\underline{\Pi}_{\gamma}]$, $\lambda^* \geq 0$ and $\nu' \leq \nu$ a threshold. The following statements are equivalent:

(i) σ^*_{α} is an optimal mixed strategy solution of the $[\text{P-MS}]_{\leq\nu'}$ problem and λ^* is an optimal dual solution of the underlying $[\text{D-MS}]_{\nu'}$ problem.

(ii) $\sigma^*_{\alpha} \in \arg\min_{\sigma} \ \mathbb{L}_{\alpha,T}(\sigma,\lambda^*)$, $\mathbb{E}^{\sigma^*_{\alpha}}_{s_0}[\,\mathrm{TS}^{\mathcal{G}}] \leq \nu'$, and

$$\lambda^*\Big(\,\mathbb{E}^{\sigma^*_{\alpha}}_{s_0}[\,\mathrm{TS}^{\mathcal{G}}] \,-\, \nu'\,\Big) \quad = \quad 0 \ . \tag{5.31}$$

Proof. $(i) \Rightarrow (ii)$| It follows that:

$$\begin{aligned}
\mathbb{J}^{\mathrm{m}*}_{\alpha,T} &= \mathbb{L}^{\mathrm{m}*}_{\alpha,T} \\
&\quad \text{(by Proposition 5.4.4, which is also true in this context)} \\
&= \mathbb{L}^{\mathrm{m}}_{\alpha,T}(\lambda^*) \\
&\quad \text{(because } \lambda^* \text{ is an optimal dual solution)} \\
&= \inf_{\sigma} \ \mathbb{L}_{\alpha,T}(\sigma, \lambda^*) \\
&\quad \text{(by definition (5.24) of the Lagrange dual function)} \\
&\leq \mathbb{L}_{\alpha,T}(\sigma^*_\alpha, \lambda^*) \\
&\quad \text{(by optimality and using the strategy } \sigma^*_\alpha \text{)} \\
&= \mathbb{J}_{\alpha,T}(\sigma^*_\alpha) \ + \ \lambda^*(\,\mathbb{E}^{\sigma^*_\alpha}_{s_0}[\,\mathrm{TS}^{\mathcal{G}}] \ - \ \nu'\,) \\
&\quad \text{(by definition (5.22) of the Lagrange function)} \\
&\leq \mathbb{J}_{\alpha,T}(\sigma^*_\alpha) \\
&\quad \text{(because } \lambda^*(\,\mathbb{E}^{\sigma^*_\alpha}_{s_0}[\,\mathrm{TS}^{\mathcal{G}}] \ - \ \nu'\,) \leq 0 \text{)}
\end{aligned}$$

More precisely, the latter inequality follows because σ^*_α is an optimal solution of the $[\text{P-MS}]_{\leq\nu'}$ problem by hypothesis, then $\mathbb{E}^{\sigma^*_\alpha}_{s_0}[\mathrm{TS}^{\mathcal{G}}] \leq \nu'$. Since σ^*_α is an optimal solution of the problem $[\text{P-MS}]_{\leq\nu'}$, the equality holds in the above inequalities and then, it follows that:

$$\sigma^*_\alpha \in \arg\min_{\sigma} \ \mathbb{L}_{\alpha,T}(\sigma, \lambda^*) \ ,$$

and also, $\lambda^*(\,\mathbb{E}^{\sigma^*_\alpha}_{s_0}[\,\mathrm{TS}^{\mathcal{G}}] \ - \ \nu'\,) = 0$ holds.

$(ii) \Rightarrow (i)$| Conversely, we need to show that σ^*_α and λ^* are resp. an optimal solution of $[\text{P-MS}]_{\leq\nu'}$ and $[\text{D-MS}]_{\nu'}$. It follows that:

$$\begin{aligned}
\mathbb{J}_{\alpha,T}(\sigma^*_\alpha) &= \mathbb{J}_{\alpha,T}(\sigma^*_\alpha) \ + \ \lambda^*(\,\mathbb{E}^{\sigma^*_\alpha}_{s_0}[\,\mathrm{TS}^{\mathcal{G}}] \ - \ \nu'\,) \\
&\quad \text{(since the second term is zero by hypothesis)} \\
&= \mathbb{L}_{\alpha,T}(\sigma^*_\alpha, \lambda^*) \\
&\quad \text{(by definition (5.22) of the Lagrange function)} \\
&= \inf_{\sigma} \ \mathbb{L}_{\alpha,T}(\sigma, \lambda^*) \\
&\quad \text{(by the hypotheses over } \sigma^*_\alpha \text{)} \\
&= \mathbb{L}^{\mathrm{m}}_{\alpha,T}(\lambda^*) \\
&\quad \text{(by definition (5.24) of the Lagrange dual function)} \\
&\leq \mathbb{L}^{\mathrm{m}*}_{\alpha,T} \\
&\quad \text{(by optimality of the optimal dual value (5.26))} \\
&= \mathbb{J}^{\mathrm{m}*}_{\alpha,T} \\
&\quad \text{(by Proposition 5.4.4, which is also true in this context)} \\
&\leq \mathbb{J}_{\alpha,T}(\sigma^*_\alpha) \\
&\quad \text{(since } \sigma^*_\alpha \text{ is feasible by hypothesis)}
\end{aligned}$$

Thus, the equality holds throughout above and then, σ^*_α and λ^* are resp. optimal solutions of the $[\text{P-MS}]_{\leq\nu'}$ and $[\text{D-MS}]_{\nu'}$ problems. Moreover,

$$\mathbb{L}^{\mathrm{m}*} \ = \ \mathbb{L}^{\mathrm{m}}_{\alpha,T}(\lambda^*) \ = \ \mathbb{J}^{\mathrm{m}*}_{\alpha,T} \ = \ \mathbb{J}^{\mathrm{m}}_{\alpha,T}(\sigma^*_\alpha) \ .$$

■

Proof B.2.10: [**Corollary** 5.4.7]

Let $\lambda^* > 0$ and $\nu' = \nu - \varsigma$ a threshold, where ς satisfies the eq. (5.32). Suppose that there is $\sigma^*_\alpha \in \arg\min_\sigma \mathbb{L}_{\alpha,T}(\sigma, \lambda^*)$ such that:

$$\mathbb{E}^{\sigma^*_\alpha}_{s_0}[\,\mathrm{TS}^{\mathcal{G}}\,] \;=\; \nu' \,.$$

Then, σ^*_α is an optimal mixed strategy solution of the $[\text{P-MS}]_{\leq\nu'}$ problem, λ^* is an optimal dual solution of the underlying $[\text{D-MS}]_{\nu'}$ problem, and the following holds:

$$\mathbb{L}^{\mathrm{m}*}_{\alpha,T} = \mathbb{L}^{\mathrm{m}}_{\alpha,T}(\lambda^*) = \mathbb{L}^{\mathrm{p}*}_{\alpha,T} = \mathbb{L}^{\mathrm{p}}_{\alpha,T}(\lambda^*) = \mathbb{J}^{\mathrm{m}*}_{\alpha,T} = \mathbb{J}_{\alpha,T}(\sigma^*_\alpha) \leq \mathbb{J}^{\mathrm{p}*}_{\alpha,T}$$

Proof. The proof follows from the particular case in the Proposition 5.4.6, with $\lambda^* > 0$. The equalities hold resp. because λ^* is an optimal dual solution for mixed strategies, i.e., for the problem $[\text{D-MS}]_{\nu'}$, then

$$\mathbb{L}^{\mathrm{m}*}_{\alpha,T} \;=\; \mathbb{L}^{\mathrm{m}}_{\alpha,T}(\lambda^*) \,.$$

Next, by Proposition 5.4.5, we have that

$$\mathbb{L}^{\mathrm{m}*}_{\alpha,T} \;=\; \mathbb{L}^{\mathrm{p}*}_{\alpha,T}$$

and then, λ^* is also the optimal dual solution for pure strategies, i.e., for the problem $[\text{D-PS}]_{\nu'}$, thus

$$\mathbb{L}^{\mathrm{p}*}_{\alpha,T} \;=\; \mathbb{L}^{\mathrm{p}}_{\alpha,T}(\lambda^*) \,.$$

The equality

$$\mathbb{J}^{\mathrm{m}*}_{\alpha,T} \;=\; \mathbb{L}^{\mathrm{m}*}_{\alpha,T}$$

holds by Proposition 5.4.4. In addition, since σ^*_α is an optimal mixed strategy for the problem $[\text{P-MS}]_{\leq\nu'}$, we have that:

$$\mathbb{J}^{\mathrm{m}*}_{\alpha,T} \;=\; \mathbb{J}^{\mathrm{m}}_{\alpha,T}(\sigma^*_\alpha) \,.$$

Finally, the inequality

$$\mathbb{J}_{\alpha,T}(\sigma^*_\alpha) \;\leq\; \mathbb{J}^{\mathrm{p}*}_{\alpha,T}$$

follows by Proposition 5.3.2. ∎

Proof B.2.11: [**Proposition** 5.4.8]

There exist two pure strategies $\overline{\pi}'_\alpha,\, \overline{\pi}''_\alpha \in \arg\min_{\overline{\pi}} \mathbb{L}_{\alpha,T}(\overline{\pi}, \lambda^*)$ for a dual variable $\lambda^* > 0$, defining a mixed strategy $\sigma^*_\alpha \in \Delta[\,\{\overline{\pi}'_\alpha,\, \overline{\pi}''_\alpha\}\,]$ to be a solution of the $[\text{P-MS}]_{\leq\nu}$ problem.

Proof. We are interested here in solving the $[\text{P-MS}]_{\leq\nu}$ problem by using the Corollary 5.4.7. For that, we need to define a dual variable $\lambda^* > 0$ and a mixed strategy $\sigma^*_\alpha \in \arg\min_\sigma \mathbb{L}_{\alpha,T}(\sigma, \lambda^*)$ such that $\mathbb{E}^{\sigma^*_\alpha}_{s_0}[\,\mathrm{TS}^{\mathcal{G}}\,] = \nu$.

First, we focus in feasible pure strategies, i.e., strategies satisfying the constraint in expectation, to define a finite dual variable λ^*. Such a variable must be consistent with the hypothesis of the Corollary 5.4.7 and the optimal dual value (5.26). So, since there is at least one feasible pure strategy by Assumption 4.2.3, more precisely the strategy π_γ of the expected SSP-problem (5.6), we have the following:

$$\begin{aligned}
\mathbb{L}^{\mathrm{p}*}_{\alpha,T} \;&\leq\; \mathbb{J}^{\mathrm{p}*}_{\alpha,T} \\
&\quad \text{(by Lemma 5.4.3)} \\
&\leq\; \mathbb{J}_{\alpha,T}(\pi_\gamma) \\
&\quad \text{(by the existence of the feasible pure strategy } \pi_\gamma \text{)} \\
&\leq\; 1 \\
&\quad \text{(since the objective function (5.12) is a probability)}
\end{aligned}$$

In this way, the optimal dual value $\mathbb{L}^{\mathrm{p}*}_{\alpha,T}$ is finite for feasible pure strategies. Now, by definition (5.26) of $\mathbb{L}^{\mathrm{p}*}_{\alpha,T}$, we can focus in the latter restricted to be:

$$\sup_{\lambda\geq 0}\left\{ \mathbb{L}^{\mathrm{p}}_{\alpha,T}(\lambda) \;\middle|\; \mathbb{E}^{\overline{\pi}^*_{\alpha,\lambda}}_{s_0}[\,\mathrm{TS}^{\mathcal{G}}\,] \leq \nu \;,\;\; \overline{\pi}^*_{\alpha,\lambda} \in \arg\min_{\overline{\pi}}\; \mathbb{L}_{\alpha,T}(\overline{\pi},\lambda) \right\} .$$

Based on this, the dual variable that we are interested is the next one:

$$\lambda^* \;=\; \sup\left\{ \lambda \geq 0 \;\middle|\; \mathbb{E}^{\overline{\pi}^*_{\alpha,\lambda}}_{s_0}[\,\mathrm{TS}^{\mathcal{G}}\,] \leq \nu \;,\; \overline{\pi}^*_{\alpha,\lambda} \in \arg\min_{\overline{\pi}}\; \mathbb{L}_{\alpha,T}(\overline{\pi},\lambda) \right\} .$$

Note that for the moment, we have not restricted λ^* to be strictly positive. Moreover, such a dual variable is finite, as shown below.

Towards a contradiction, suppose that $\lambda^* = +\infty$. Then, we can deduce that $\mathbb{E}^{\overline{\pi}^*_{\alpha,\lambda}}_{s_0}[\,\mathrm{TS}^{\mathcal{G}}\,] > \nu$ for any λ, where $\overline{\pi}^*_{\alpha,\lambda} \in \arg\min_{\overline{\pi}} \mathbb{L}_{\alpha,T}(\overline{\pi},\lambda)$. In addition, because there exists π_γ satisfying the (strict) inequality in the constraint in expectation for ν, i.e., $\mathbb{E}^{\pi_\gamma}_{s_0}[\,\mathrm{TS}^{\mathcal{G}}\,] < \nu$, then it is clear that the following holds for any $\lambda > 0$:

$$\begin{aligned}
0 \;&\leq\; \mathbb{J}_{\alpha,T}(\overline{\pi}^*_{\alpha,\lambda}) \;+\; \lambda\Big(\,\mathbb{E}^{\overline{\pi}^*_{\alpha,\lambda}}_{s_0}[\,\mathrm{TS}^{\mathcal{G}}\,] \;-\; \nu\,\Big) \\
&\quad \text{(because } \mathbb{E}^{\overline{\pi}^*_{\alpha,\lambda}}_{s_0}[\,\mathrm{TS}^{\mathcal{G}}\,] > \nu \text{ by hypothesis)} \\
&\leq\; \mathbb{J}_{\alpha,T}(\pi_\gamma) \;+\; \lambda\Big(\,\mathbb{E}^{\pi_\gamma}_{s_0}[\,\mathrm{TS}^{\mathcal{G}}\,] \;-\; \nu\,\Big) \\
&\quad \text{(because } \overline{\pi}^*_{\alpha,\lambda} \text{ gives the minimum, and by using } \pi_\gamma \text{)} \\
&\leq\; 1 \\
&\quad \text{(since } \lambda(\mathbb{E}^{\pi_\gamma}_{s_0}[\,\mathrm{TS}^{\mathcal{G}}\,] - \nu) \leq 0 \text{ and } \mathbb{J}_{\alpha,T}(\pi_\gamma) \leq 1 \text{)}
\end{aligned}$$

In this way, for any $\lambda > 0$, it holds:

$$0 \;\leq\; \mathbb{J}_{\alpha,T}(\pi_\gamma) \;+\; \lambda\Big(\,\mathbb{E}^{\pi_\gamma}_{s_0}[\,\mathrm{TS}^{\mathcal{G}}\,] \;-\; \nu\,\Big) \;\;\leq\;\; 1\;,$$

which is a contradiction. For example, consider for any $J > 0$,

$$\lambda \;=\; \frac{J + \mathbb{J}_{\alpha,T}(\pi_\gamma)}{\nu - \mathbb{E}^{\pi_\gamma}_{s_0}[\,\mathrm{TS}^{\mathcal{G}}\,]}\;.$$

Now, the idea is to show that two pure strategies are sufficient so that by combining them we can construct an optimal mixed strategy solving the [P-MS]$_{\leq \nu}$ problem. First, we let two values $\underline{\lambda}_0 = 0$ and $\overline{\lambda}_0 = \lambda^+$, where $\lambda^+ > 0$ is taken large enough until to find a strategy:

$$\overline{\pi}^*_{\alpha,\lambda^+} \in \arg\min_{\overline{\pi}} \ \mathbb{L}_{\alpha,T}(\overline{\pi}, \lambda^+) \ ,$$

such that the constraint in expectation is satisfied with the strict inequality for ν, in which there is at least one by Assumption 4.2.3, i.e., the strategy π_γ. On the other hand, $\underline{\lambda}_0 = 0$ is used to find an optimal pure strategy in the unconstrained problem:

$$\overline{\pi}^*_{\alpha,0} \in \arg\min_{\overline{\pi}} \ \mathbb{L}_{\alpha,T}(\overline{\pi}, 0) \quad = \quad \arg\min_{\overline{\pi}} \ \mathbb{J}_{\alpha,T}(\overline{\pi}) \ . \tag{B.30}$$

Note that if this strategy is such that the constraint in expectation for ν is satisfied, then such a constraint is not relevant for the (primal mixed strategy) problem, because we only need to minimize $\mathbb{J}_{\alpha,T}(\overline{\pi})$ under pure strategies and then a solution will be the strategy $\overline{\pi}^*_{\alpha,0}$. Otherwise, then we assume that the constraint is not satisfied for such a strategy. Thus, it is clear that $\underline{\lambda}_0 \leq \lambda^* < \overline{\lambda}_0$ and

$$\mathbb{E}^{\overline{\pi}^*_{\alpha,\overline{\lambda}_0}}_{s_0}[\mathrm{TS}^{\mathcal{G}}] \ < \ \nu \ < \ \mathbb{E}^{\overline{\pi}^*_{\alpha,\underline{\lambda}_0}}_{s_0}[\mathrm{TS}^{\mathcal{G}}] \ .$$

Since λ^* is finite and non-negative, we can consider two monotone nonnegative sequences to approach this value, a nondecreasing one $(\underline{\lambda}_n)_{n\in\mathbb{N}_0}$ and another nonincreasing $(\overline{\lambda}_n)_{n\in\mathbb{N}_0}$, such that:

$$\underline{\lambda}_n \nearrow \lambda^* \qquad \text{and} \qquad \overline{\lambda}_n \searrow \lambda^* \qquad \text{as} \qquad n \to +\infty \ .$$

This sequences can be constructed, e.g., by the so-called bisection method. In this way, the following holds for each $n \in \mathbb{N}_0$:

$$\lambda^* \ \in \ \bigcap_{m=0}^{n} \left[\underline{\lambda}_m \, , \, \overline{\lambda}_m\right] \ = \ \left[\underline{\lambda}_n \, , \, \overline{\lambda}_n\right] . \tag{B.31}$$

Note that $(\underline{\lambda}_n)_{n\in\mathbb{N}_0}$ may be a zero sequence (this may occur when $\lambda^* = 0$, which is not yet a case analysed for the moment). Now, since for each $n \in \mathbb{N}$ fixed, the values $\underline{\lambda}_n$ and $\overline{\lambda}_n$ are fixed as well, we can obtain resp. two strategies:

$$\overline{\pi}^*_{\alpha,\underline{\lambda}_n} \in \arg\min_{\overline{\pi}} \ \mathbb{L}_{\alpha,T}(\overline{\pi}, \underline{\lambda}_n) \qquad \text{and} \qquad \overline{\pi}^*_{\alpha,\overline{\lambda}_n} \in \arg\min_{\overline{\pi}} \ \mathbb{L}_{\alpha,T}(\overline{\pi}, \overline{\lambda}_n) \ .$$

In addition, because the sequences $(\underline{\lambda}_n)_{n\in\mathbb{N}_0}$ and $(\overline{\lambda}_n)_{n\in\mathbb{N}_0}$ are monotones, we have by Lemma 5.4.1 that:

$$\underline{\lambda}_n \ \mapsto \ \mathbb{E}^{\overline{\pi}^*_{\alpha,\underline{\lambda}_n}}_{s_0}[\mathrm{TS}^{\mathcal{G}}] \qquad \text{and} \qquad \overline{\lambda}_n \ \mapsto \ \mathbb{E}^{\overline{\pi}^*_{\alpha,\overline{\lambda}_n}}_{s_0}[\mathrm{TS}^{\mathcal{G}}]$$

are resp. nonincreasing and nondecreasing. Based on this, the following inequality holds for each $n \in \mathbb{N}$:

$$\mathbb{E}^{\overline{\pi}^*_{\alpha,\overline{\lambda}_n}}_{s_0}[\mathrm{TS}^{\mathcal{G}}] \ \leq \ \nu \ \leq \ \mathbb{E}^{\overline{\pi}^*_{\alpha,\underline{\lambda}_n}}_{s_0}[\mathrm{TS}^{\mathcal{G}}] \ . \tag{B.32}$$

On the other hand, since we assume that there are a finite number of pure strategies, then the set $\underline{\Pi}_\gamma$ is compact. We can therefore consider two monotone convergent subsequences of pure strategies, one for each of the following sequences:

$$\left(\overline{\pi}^*_{\alpha,\underline{\lambda}_n}\right)_{n\in\mathbb{N}_0} \quad \text{and} \quad \left(\overline{\pi}^*_{\alpha,\overline{\lambda}_n}\right)_{n\in\mathbb{N}_0} .$$

More precisely, we can assume resp. that

$$\left(\overline{\pi}^*_{\alpha,\underline{\lambda}_{n_j}}\right)_{j\in\mathbb{N}_0} \quad \text{and} \quad \left(\overline{\pi}^*_{\alpha,\overline{\lambda}_{n_j}}\right)_{j\in\mathbb{N}_0} \tag{B.33}$$

are infinitely constant from some index and then convergent in $\underline{\Pi}_\gamma$. It is clear that under the above subsequences, the inequality (B.32) also holds. We fix $\underline{\lambda}_{n_0} = \underline{\lambda}_0$ and $\overline{\lambda}_{n_0} = \overline{\lambda}_0$ as the initial values of the subsequences. In the following, we pay particular attention to the subsequence on the right side in (B.33), because at this point, we can define a pure strategy giving the strict inequality in the constraint in expectation, i.e., less than ν. Without loss of generality, the following holds for each $j \in \mathbb{N}$:

$$\mathbb{E}_{s_0}^{\overline{\pi}^*_{\alpha,\overline{\lambda}_0}}[\mathrm{TS}^{\mathcal{G}}] \;<\; \mathbb{E}_{s_0}^{\overline{\pi}^*_{\alpha,\overline{\lambda}_{n_j}}}[\mathrm{TS}^{\mathcal{G}}] \;\leq\; \nu \,,$$

where $\overline{\lambda}_{n_j} \mapsto \mathbb{E}_{s_0}^{\overline{\pi}^*_{\alpha,\overline{\lambda}_{n_j}}}[\mathrm{TS}^{\mathcal{G}}]$ is nondecreasing. Then, since the subsequences in (B.33) are infinitely constant from some index, we infer that there must be $j^* \in \mathbb{N}_0$ for which the previous inequality on the right side is strict, i.e.,

$$\mathbb{E}_{s_0}^{\overline{\pi}^*_{\alpha,\overline{\lambda}_{n_{j^*}}}}[\mathrm{TS}^{\mathcal{G}}] \;<\; \nu$$

Taking into account such an index, we can redefine (with a small abuse of notation) the subsequence on the right side in (B.33) as:

$$\overline{\pi}^*_{\alpha,\overline{\lambda}_{n_j}} \;:=\; \begin{cases} \overline{\pi}^*_{\alpha,\overline{\lambda}_{n_j}} & \text{if } j \leq j^* \\ \overline{\pi}^*_{\alpha,\overline{\lambda}_{n_{j^*}}} & \text{otherwise} \end{cases}$$

As we have said above, the subsequences are convergent in $\underline{\Pi}_\gamma$, in particular the subsequence redefined above. Then, there exist two pure strategies, let say π'_α, $\pi''_\alpha \in \underline{\Pi}_\gamma$, such that:

$$\overline{\pi}^*_{\alpha,\underline{\lambda}_{n_j}} \to \pi'_\alpha \quad \text{and} \quad \overline{\pi}^*_{\alpha,\overline{\lambda}_{n_j}} \to \pi''_\alpha \quad \text{as} \quad j \to +\infty \,.$$

Note that π''_α corresponds to the strategy $\overline{\pi}^*_{\alpha,\overline{\lambda}_{n_{j^*}}}$. In addition, the following inequality holds:

$$\mathbb{E}_{s_0}^{\pi''_\alpha}[\,\mathrm{TS}^{\mathcal{G}}] \;<\; \nu \;\leq\; \mathbb{E}_{s_0}^{\pi'_\alpha}[\,\mathrm{TS}^{\mathcal{G}}] \,. \tag{B.34}$$

We notice in particular that if $\lambda^* = 0$, then the sequence $(\underline{\lambda}_n)_{n\in\mathbb{N}_0}$ will be the zero sequence and then, also the convergent subsequence $(\underline{\lambda}_{n_j})_{j\in\mathbb{N}_0}$. In such a case, the strategy π'_α is thus the strategy $\overline{\pi}^*_{\alpha,0}$, see eq. (B.30), that is assumed not satisfying the constraint (otherwise, the constraint is not relevant for the problem). Then, we only use the sequence $(\overline{\lambda}_n)_{n\in\mathbb{N}_0}$ to find the strategy $\overline{\pi}''_\alpha$, which will be a solution of the problem $[\text{P-MS}]_{\leq\nu}$. Indeed, for any pure strategy $\overline{\pi}$,

$$\mathbb{L}_{\alpha,T}\left(\overline{\pi}^*_{\alpha,\overline{\lambda}_{n_j}}, \overline{\lambda}_{n_j}\right) \;\leq\; \mathbb{L}_{\alpha,T}\left(\overline{\pi}, \overline{\lambda}_{n_j}\right)\,,$$

in particular for the strategies satisfying the constraint in expectation. Then, from the definition (5.22) of the Lagrange function, it holds:

$$\mathbb{J}_{\alpha,T}\left(\overline{\pi}^*_{\alpha,\overline{\lambda}_{n_j}}\right) \,-\, \mathbb{J}_{\alpha,T}(\overline{\pi}) \;\leq\; \overline{\lambda}_{n_j}\left(\mathbb{E}^{\overline{\pi}}_{s_0}[\mathrm{TS}^{\mathcal{G}}] \,-\, \mathbb{E}^{\overline{\pi}^*_{\alpha,\overline{\lambda}_{n_j}}}_{s_0}[\mathrm{TS}^{\mathcal{G}}]\right)\,.$$

Taking $j\to+\infty$, we have that $\mathbb{J}_{\alpha,T}(\overline{\pi}''_\alpha)\leq\mathbb{J}_{\alpha,T}(\overline{\pi})$ and it satisfies the constraint in expectation, see eq (B.34). We conclude that $\overline{\pi}''_\alpha$ is optimal for the problem $[\text{P-MS}]_{\leq\nu}$, and mixed strategies are not necessary for this case. From now, we consider that λ^* is restricted to be positive.

As we have said in the beginning, to solve the $[\text{P-MS}]_{<\nu}$ problem, we solve the $[\text{P-MS}]_{\leq\nu}$ problem by using the Corollary 5.4.7. The idea is to combine the strategies $\overline{\pi}'_\alpha$ and $\overline{\pi}''_\alpha$. First, we show that these strategies are such that:

$$\overline{\pi}'_\alpha,\,\overline{\pi}''_\alpha \;\in\; \arg\min_{\overline{\pi}}\; \mathbb{L}_{\alpha,T}(\overline{\pi},\lambda^*)\,. \tag{B.35}$$

Indeed,

$$\begin{aligned}
\mathbb{L}_{\alpha,T}\left(\overline{\pi}^*_{\alpha,\underline{\lambda}_{n_j}}, \underline{\lambda}_{n_j}\right) &\;=\; \min_{\overline{\pi}}\;\mathbb{L}_{\alpha,T}(\overline{\pi},\underline{\lambda}_{n_j}) \;\leq\; \mathbb{L}_{\alpha,T}(\overline{\pi},\underline{\lambda}_{n_j})\;, \quad \forall\overline{\pi}\\
\mathbb{L}_{\alpha,T}\left(\overline{\pi}^*_{\alpha,\overline{\lambda}_{n_j}}, \overline{\lambda}_{n_j}\right) &\;=\; \min_{\overline{\pi}}\;\mathbb{L}_{\alpha,T}(\overline{\pi},\overline{\lambda}_{n_j}) \;\leq\; \mathbb{L}_{\alpha,T}(\overline{\pi},\overline{\lambda}_{n_j})\;, \quad \forall\overline{\pi}
\end{aligned}$$

Thus, taking $j\to+\infty$ on the two inequalities, the following holds:

$$\begin{aligned}
\mathbb{L}_{\alpha,T}(\overline{\pi}'_\alpha,\lambda^*) &\;\leq\; \mathbb{L}_{\alpha,T}(\overline{\pi},\lambda^*) \;, \quad \forall\overline{\pi}\\
\mathbb{L}_{\alpha,T}(\overline{\pi}''_\alpha,\lambda^*) &\;\leq\; \mathbb{L}_{\alpha,T}(\overline{\pi},\lambda^*) \;, \quad \forall\overline{\pi}
\end{aligned}$$

which shows that $\overline{\pi}'_\alpha$ and $\overline{\pi}''_\alpha$ satisfy (B.35). In the following, the proof continues based mainly on eqs. (B.34) and (B.35). On the other hand, we notice from the eq. (B.34) that $\mathbb{E}^{\overline{\pi}'_\alpha}_{s_0}[\,\mathrm{TS}^{\mathcal{G}}]$ could be equal to ν. In such a case, since the strategy $\overline{\pi}'_\alpha$ satisfies (B.35), we can infer that $\overline{\pi}'_\alpha$ is a solution of the $[\text{P-MS}]_{\leq\nu}$ problem by applying the Corollary 5.4.7. Thus, mixed strategies are not necessary to solve the $[\text{P-MS}]_{\leq\nu}$ problem, in other words, $\overline{\pi}'_\alpha$ is chosen with probability one. On the contrary, we can in fact combine $\overline{\pi}'_\alpha$ and $\overline{\pi}''_\alpha$, as it is shown below.

Let $\Delta[\,\{\overline{\pi}'_\alpha,\,\overline{\pi}''_\alpha\}\,] \;\subseteq\; \Delta[\,\underline{\Pi}_\gamma\,]$ be the set of mixed strategies combining convexly only over $\overline{\pi}'_\alpha$ and $\overline{\pi}''_\alpha$, i.e., a mixed strategy $\sigma\in\Delta[\,\{\overline{\pi}'_\alpha,\overline{\pi}''_\alpha\}\,]$ is such that:

$$\sigma(\overline{\pi}'_\alpha) \,+\, \sigma(\overline{\pi}''_\alpha) \;=\; 1\,. \tag{B.36}$$

Now, for any of these mixed strategies, we have by Proposition 5.3.1 that $\mathbb{E}^{\sigma}_{s_0}[\,\mathrm{TS}^{\mathcal{G}}]$ is continuous for any σ, in particular for the strategies in $\Delta[\,\{\overline{\pi}'_\alpha,\overline{\pi}''_\alpha\}\,]$. Thus, from the inequality (B.34), we use the eq. (5.18) and the so-called intermediate value theorem, to conclude that there exists a value within $[0,1]$, let say $\sigma^*_\alpha(\overline{\pi}'_\alpha)$ for $\overline{\pi}'_\alpha$ and $\sigma^*_\alpha(\overline{\pi}''_\alpha)$ for $\overline{\pi}''_\alpha$, such that:

$$\sigma^*_\alpha(\overline{\pi}'_\alpha)\,\mathbb{E}^{\overline{\pi}'_\alpha}_{s_0}[\,\mathrm{TS}^{\mathcal{G}}] \,+\, \sigma^*_\alpha(\overline{\pi}''_\alpha)\,\mathbb{E}^{\overline{\pi}''_\alpha}_{s_0}[\,\mathrm{TS}^{\mathcal{G}}] \;=\; \nu \tag{B.37}$$

which proves the existence of a mixed strategy σ^*_α constructed by combining at most two pure strategies. Moreover, this strategy is explicitly defined from the eqs. (B.36) and (B.37) by:

$$\begin{aligned}\sigma^*_\alpha(\overline{\pi}'_\alpha) \quad &:= \quad \frac{\nu \quad - \quad \mathbb{E}^{\overline{\pi}''_\alpha}_{s_0}[\,\mathrm{TS}^{\mathcal{G}}]}{\mathbb{E}^{\overline{\pi}'_\alpha}_{s_0}[\,\mathrm{TS}^{\mathcal{G}}] \quad - \quad \mathbb{E}^{\overline{\pi}''_\alpha}_{s_0}[\,\mathrm{TS}^{\mathcal{G}}]} \\ \sigma^*_\alpha(\overline{\pi}''_\alpha) \quad &:= \quad \frac{\mathbb{E}^{\overline{\pi}'_\alpha}_{s_0}[\,\mathrm{TS}^{\mathcal{G}}] \quad - \quad \nu}{\mathbb{E}^{\overline{\pi}'_\alpha}_{s_0}[\,\mathrm{TS}^{\mathcal{G}}] \quad - \quad \mathbb{E}^{\overline{\pi}''_\alpha}_{s_0}[\,\mathrm{TS}^{\mathcal{G}}]}\end{aligned} \tag{B.38}$$

To apply the Corollary 5.4.7 and conclude that σ^*_α is an optimal mixed strategy solution of the $[\text{P-MS}]_{\leq\nu}$ problem, we show that:

$$\sigma^*_\alpha \;\in\; \arg\min_\sigma \; \mathbb{L}_{\alpha,T}(\sigma,\lambda^*)\;. \tag{B.39}$$

First, if $\sigma^*_\alpha(\overline{\pi}'_\alpha) = 1$, so the strategy to play is $\overline{\pi}'_\alpha$, and this one is such that (B.35) is satisfied, i.e., satisfying (B.39) above. In the following, we suppose that $\sigma^*_\alpha(\overline{\pi}'_\alpha) \neq 1$. Note also that $\sigma^*_\alpha(\overline{\pi}'_\alpha) \neq 0$, because on the contrary, $\sigma^*_\alpha(\overline{\pi}''_\alpha) = 1$ and $\overline{\pi}''_\alpha$ is such that the inequality in the constraint in expectation is strict, see eq. (B.34). Thus, we consider that $\sigma^*_\alpha(\overline{\pi}'_\alpha), \sigma^*_\alpha(\overline{\pi}''_\alpha) \in (0,1)$. Now, since $\sigma \mapsto \mathbb{L}_{\alpha,T}(\sigma,\lambda^*)$ is linear over mixed strategies σ, and each one assigns marginal probabilities over pure strategies, the following holds for any σ:

$$\mathbb{L}_{\alpha,T}(\sigma,\lambda^*) \quad \geq \quad \min_{\overline{\pi}} \; \mathbb{L}_{\alpha,T}(\overline{\pi},\lambda^*)\;.$$

In addition, since the strategies $\overline{\pi}'_\alpha$ and $\overline{\pi}''_\alpha$ are such that (B.35) holds, we have that for any σ,

$$\mathbb{L}_{\alpha,T}(\sigma,\lambda^*) \;\geq\; \mathbb{L}_{\alpha,T}(\overline{\pi}'_\alpha,\lambda^*) \quad \text{and} \quad \mathbb{L}_{\alpha,T}(\sigma,\lambda^*) \;\geq\; \mathbb{L}_{\alpha,T}(\overline{\pi}''_\alpha,\lambda^*)\;.$$

Multiplying now by $\sigma^*_\alpha(\overline{\pi}'_\alpha)$ the first inequality and by $\sigma^*_\alpha(\overline{\pi}''_\alpha)$ the second one, and adding after, it holds:

$$\begin{aligned}\mathbb{L}_{\alpha,T}(\sigma,\lambda^*) \quad &= \;\sigma^*_\alpha(\overline{\pi}'_\alpha)\,\mathbb{L}_{\alpha,T}(\sigma,\lambda^*) \;+\; \sigma^*_\alpha(\overline{\pi}''_\alpha)\,\mathbb{L}_{\alpha,T}(\sigma,\lambda^*) \\ &\qquad (\text{because} \;\; \sigma^*_\alpha(\overline{\pi}'_\alpha) \;+\; \sigma^*_\alpha(\overline{\pi}''_\alpha) = 1\;) \\ &\geq \;\sigma^*_\alpha(\overline{\pi}'_\alpha)\,\mathbb{L}_{\alpha,T}(\overline{\pi}'_\alpha,\lambda^*) \;+\; \sigma^*_\alpha(\overline{\pi}''_\alpha)\,\mathbb{L}_{\alpha,T}(\overline{\pi}''_\alpha,\lambda^*) \\ &= \;\mathbb{L}_{\alpha,T}(\sigma^*_\alpha,\lambda^*)\end{aligned}$$

Thus, σ^*_α is such that (B.39) holds. We conclude that by Corollary 5.4.7, σ^*_α is solution of the $[\text{P-MS}]_{\leq\nu}$ problem. ■

Proof B.2.12: [Proposition 5.4.9]

For any $\kappa > 0$, the mixed strategy $\sigma^\star_{\alpha,\varsigma}$ defined in (5.36) is κ-optimal for the $[\text{P-MS}]_{<\nu}$ problem, i.e.,

$$\mathbb{J}_{\alpha,T}(\sigma^\star_{\alpha,\varsigma}) \;\;\leq\;\; \mathbb{J}^{\mathrm{m}*}_{\alpha,T} \;+\; \kappa\;.$$

Proof. From Proposition 5.4.8, there exist $\overline{\pi}'_\alpha, \overline{\pi}''_\alpha \in \arg\min_{\overline{\pi}} \mathbb{L}_{\alpha,T}(\overline{\pi},\lambda^*)$ for a dual variable $\lambda^* > 0$, defining the mixed strategy $\sigma^\star_{\alpha,\varsigma}$ in (5.36). On the other hand, the following holds:

$$
\begin{aligned}
0 \;&\le\; \mathbb{J}_{\alpha,T}(\sigma^\star_{\alpha,\varsigma}) \;-\; \mathbb{J}^{m*}_{\alpha,T} \\
&=\; \sigma^\star_{\alpha,\varsigma}(\overline{\pi}'_\alpha)\,\mathbb{J}_{\alpha,T}(\overline{\pi}'_\alpha) \;+\; \sigma^\star_{\alpha,\varsigma}(\overline{\pi}''_\alpha)\,\mathbb{J}_{\alpha,T}(\overline{\pi}''_\alpha) \;-\; \mathbb{J}^{m*}_{\alpha,T} \\
&\quad \text{(because } \sigma \mapsto \mathbb{J}_{\alpha,T}(\sigma) \text{ is linear on mixed strategies)} \\
&=\; \sigma^\star_{\alpha,\varsigma}(\overline{\pi}'_\alpha)\,(\mathbb{J}_{\alpha,T}(\overline{\pi}'_\alpha) - \mathbb{J}^{m*}_{\alpha,T}) \;+\; \sigma^\star_{\alpha,\varsigma}(\overline{\pi}''_\alpha)\,(\mathbb{J}_{\alpha,T}(\overline{\pi}''_\alpha) - \mathbb{J}^{m*}_{\alpha,T}) \\
&\quad \text{(arranging terms and because } \sigma^\star_{\alpha,\varsigma}(\overline{\pi}'_\alpha) + \sigma^\star_{\alpha,\varsigma}(\overline{\pi}''_\alpha) = 1) \\
&=\; \sigma^*_\alpha(\overline{\pi}'_\alpha)\,(\mathbb{J}_{\alpha,T}(\overline{\pi}'_\alpha) - \mathbb{J}^{m*}_{\alpha,T}) \;+\; \sigma^*_\alpha(\overline{\pi}''_\alpha)\,(\mathbb{J}_{\alpha,T}(\overline{\pi}''_\alpha) - \mathbb{J}^{m*}_{\alpha,T}) \\
&\quad +\; \varsigma\,\frac{\mathbb{J}_{\alpha,T}(\overline{\pi}''_\alpha) \;-\; \mathbb{J}_{\alpha,T}(\overline{\pi}'_\alpha)}{\mathbb{E}^{\overline{\pi}'_\alpha}_{s_0}[\,\mathrm{TS}^{\mathcal{G}}\,] \;-\; \mathbb{E}^{\overline{\pi}''_\alpha}_{s_0}[\,\mathrm{TS}^{\mathcal{G}}\,]} \\
&\quad \text{(by using the definition (5.36) of the strategy } \sigma^\star_{\alpha,\varsigma}\text{)} \\
&\le\; \sigma^*_\alpha(\overline{\pi}'_\alpha)\,\lambda^*\,(\nu - \mathbb{E}^{\overline{\pi}'_\alpha}_{s_0}[\,\mathrm{TS}^{\mathcal{G}}]) \;+\; \sigma^*_\alpha(\overline{\pi}''_\alpha)\,\lambda^*\,(\nu - \mathbb{E}^{\overline{\pi}''_\alpha}_{s_0}[\,\mathrm{TS}^{\mathcal{G}}]) \\
&\quad +\; \varsigma\,\frac{\mathbb{J}_{\alpha,T}(\overline{\pi}''_\alpha) \;-\; \mathbb{J}_{\alpha,T}(\overline{\pi}'_\alpha)}{\mathbb{E}^{\overline{\pi}'_\alpha}_{s_0}[\,\mathrm{TS}^{\mathcal{G}}\,] \;-\; \mathbb{E}^{\overline{\pi}''_\alpha}_{s_0}[\,\mathrm{TS}^{\mathcal{G}}\,]} \\
&\quad \text{(since } \mathbb{L}_{\alpha,T}(\overline{\pi}'_\alpha,\lambda^*)\,,\,\mathbb{L}_{\alpha,T}(\overline{\pi}''_\alpha,\lambda^*) \;\le\; \mathbb{J}^{m*}_{\alpha,T} \text{ and using (5.22))} \\
&=\; \varsigma\,\frac{\mathbb{J}_{\alpha,T}(\overline{\pi}''_\alpha) \;-\; \mathbb{J}_{\alpha,T}(\overline{\pi}'_\alpha)}{\mathbb{E}^{\overline{\pi}'_\alpha}_{s_0}[\,\mathrm{TS}^{\mathcal{G}}\,] \;-\; \mathbb{E}^{\overline{\pi}''_\alpha}_{s_0}[\,\mathrm{TS}^{\mathcal{G}}\,]} \\
&\quad \text{(by using the definition (5.35) of the strategy } \sigma^*_\alpha\text{)}
\end{aligned}
$$

Then, because $\varsigma > 0$ can be small enough,

$$
\kappa \;:=\; \varsigma\,\frac{\mathbb{J}_{\alpha,T}(\overline{\pi}''_\alpha) \;-\; \mathbb{J}_{\alpha,T}(\overline{\pi}'_\alpha)}{\mathbb{E}^{\overline{\pi}'_\alpha}_{s_0}[\,\mathrm{TS}^{\mathcal{G}}\,] \;-\; \mathbb{E}^{\overline{\pi}''_\alpha}_{s_0}[\,\mathrm{TS}^{\mathcal{G}}\,]} \;>\; 0
$$

can be small enough as well. Thus, $\mathbb{J}_{\alpha,T}(\sigma^\star_{\alpha,\varsigma}) \;\le\; \mathbb{J}^{m*}_{\alpha,T} \;+\; \kappa$. ■

Proof B.2.13: [Proposition 5.4.10]

For $\lambda \ge 0$ fixed, a pure strategy $\overline{\pi}^*_{\alpha,\lambda} \in \arg\min_{\overline{\pi}} \mathbb{L}_{\alpha,T}(\overline{\pi},\lambda)$ can be computed by the classical Bellman backward recursion.

Proof. First, from the unfolding $\mathcal{M}_T$, defined in eq. (5.2),

$$
\mathcal{M}_T \;=\; \Big(\mathcal{S}_T,\, s_0,\, \mathcal{A},\, P_T,\, C\Big)\,, \tag{5.2}
$$

wherein each leaf has a copy of the original MDP $\mathcal{M}$, we "collapse" each copy in a fresh state, denoted as s_{T+1}, and we connect each leaf s_T with each corresponding fresh state s_{T+1} and add the expected shortest path cost to the goal set $\mathcal{G}$. More precisely, from each s_T at level T in $\mathcal{M}_T$ (corresponding to the state $\mathrm{proj}_1(s_T)$ in $\mathcal{M}$), we add a single edge to the respective fresh state s_{T+1} and we label such a transition by the SSP-value (5.5) computed from $\mathrm{proj}_1(s_T)$ to $\mathcal{G}$ by using the strategy π_γ, i.e., the value $\mathbb{SP}^{\pi_\gamma}_{s_T}$. We will assume that each fresh states is to be zero-cost with a self-loop. We denote such a new structure as $\underline{\mathcal{M}}_T$. Note that the latter is as the unfolding $\mathcal{M}_T$ up to the level T, with the addition of a state after each leaf in $\mathcal{M}_T$. We say so that $\underline{\mathcal{M}}_T$ has a depth $T+1$.

On the other hand, from definition (5.22) of the Lagrange function,

$$
\mathbb{L}_{\alpha,T}(\overline{\pi},\lambda) \;=\; \mathbb{J}_{\alpha,T}(\overline{\pi}) \;+\; \lambda\,(\,\mathbb{E}^{\overline{\pi}}_{s_0}[\mathrm{TS}^{\mathcal{G}}] \;-\; \nu\,)\,. \tag{5.22}
$$

The idea is then to compute a pure strategy $\overline{\pi}$ giving the minimum on $\mathbb{L}_{\alpha,T}(\overline{\pi},\lambda)$ for a fixed dual variable $\lambda \geq 0$ in the structure $\underline{\mathcal{M}}_T$. It this context, the Lagrange function can be written as:

$$\mathbb{L}_{\alpha,T}(\overline{\pi},\lambda) \;=\; \mathbb{J}_{\alpha,T}(\overline{\pi}) \;+\; \lambda\,(\,\mathbb{E}^{\overline{\pi}}_{s_0}[\mathrm{TS}_{T+1}] \;-\; \nu\,)\,, \qquad \text{(B.40)}$$

where $\mathbb{E}^{\overline{\pi}}_{s_0}[\mathrm{TS}_{T+1}]$ is the expected truncated sum (2.22) up to $T+1$. Now, we prove that the objective function for pure strategies $\mathbb{J}_{\alpha,T}(\overline{\pi})$, see eq. (5.12), can be written as the following expectation:

$$\mathbb{J}_{\alpha,T}(\pi) = \mathbb{E}^{\overline{\pi}}_{s_0}\left[\sum_{t=1}^{T}\prod_{\tau=0}^{t-1}(1-\mathbb{1}_{\mathcal{G}_T}(s_\tau))\mathbb{1}_{\mathcal{G}_T}(s_t) + \alpha\prod_{t=1}^{T}\mathbb{1}_{\overline{\mathcal{G}}}(\mathrm{proj}_1(s_t))\right]\,.$$

Indeed, $\mathbb{J}_{\alpha,T}(\overline{\pi}) = \mathbb{P}^{\overline{\pi}}_{s_0}[\,\boldsymbol{\mathcal{E}}_T \cup \overline{\boldsymbol{\mathcal{E}}}_{\alpha,T}\,]$ by definition (5.12). Now, based on the definitions of $\boldsymbol{\mathcal{E}}_T$ and $\overline{\boldsymbol{\mathcal{E}}}_{\alpha,T}$ in eq. (5.8) and eq. (5.10) resp., we prove separately that the probability of each event can be written as a expected function. Let $\omega \in \Omega_{s_0}$ a path. We have thatthe following is true by considering $\alpha = 1$:

$$\begin{aligned}\mathbb{P}^{\overline{\pi}}_{s_0}[\,\omega \in \overline{\boldsymbol{\mathcal{E}}}_{1,T}\,] \;&=\; 1\cdot\mathbb{P}^{\overline{\pi}}_{s_0}[\,\omega \in \overline{\boldsymbol{\mathcal{E}}}_{1,T}\,] \;+\; 0\cdot\mathbb{P}^{\overline{\pi}}_{s_0}[\,\omega \in \boldsymbol{\mathcal{E}}_T\,]\\ &=\; \mathbb{E}^{\overline{\pi}}_{s_0}[\,\mathbb{1}_{\overline{\boldsymbol{\mathcal{E}}}_{1,T}}(\omega)\,]\end{aligned}$$

But, from the definition (5.10) of $\overline{\boldsymbol{\mathcal{E}}}_{\alpha,T}$ for $\alpha = 1$, it holds:

$$\omega \in \overline{\boldsymbol{\mathcal{E}}}_{1,T} \quad\Leftrightarrow\quad \bigwedge_{t=0}^{T}\,(\mathrm{proj}_1 \circ \mathrm{proj}^{\mathcal{S}_T}_t\,)(\omega) \notin \mathcal{G}$$

Thus, we can write:

$$\mathbb{P}^{\overline{\pi}}_{s_0}[\,\omega \in \overline{\boldsymbol{\mathcal{E}}}_{\alpha,T}] \;=\; \alpha\,\mathbb{E}^{\overline{\pi}}_{s_0}\left[\prod_{t=0}^{T}\mathbb{1}_{\overline{\mathcal{G}}}((\mathrm{proj}_1 \circ \mathrm{proj}^{\mathcal{S}_T}_t\,)(\omega))\right]\,. \qquad \text{(B.41)}$$

On the other hand, from the definition $\boldsymbol{\mathcal{E}}_T$ in eq. (5.8), it holds:

$$\omega \in \boldsymbol{\mathcal{E}}_T \;\Leftrightarrow\; \left\{\mathrm{proj}^{\mathcal{S}_T}_0(\omega) \in \mathcal{G}_T\right\} \bigvee_{t=1}^{T} \left\{\mathrm{proj}^{\mathcal{S}_T}_t(\omega) \in \mathcal{G}_T \bigwedge_{\tau=0}^{t-1} \mathrm{proj}^{\mathcal{S}_T}_\tau(\omega) \notin \mathcal{G}_T\right\}$$

Then, assuming that $\mathrm{proj}^{\mathcal{S}_T}_0(\omega) \notin \mathcal{G}_T$, we have the following:

$$\mathbb{P}^{\overline{\pi}}_{s_0}[\,\omega \in \boldsymbol{\mathcal{E}}_T] \;=\; \mathbb{E}^{\overline{\pi}}_{s_0}\left[\sum_{t=1}^{T}\prod_{\tau=0}^{t-1}\Big(1-\mathbb{1}_{\mathcal{G}_T}(\mathrm{proj}^{\mathcal{S}_T}_\tau(\omega))\Big)\mathbb{1}_{\mathcal{G}_T}(\mathrm{proj}^{\mathcal{S}_T}_t(\omega))\right]$$

Because $\mathbb{P}^{\overline{\pi}}_{s_0}[\,\boldsymbol{\mathcal{E}}_T \cup \overline{\boldsymbol{\mathcal{E}}}_{\alpha,T}] = \mathbb{P}^{\overline{\pi}}_{s_0}[\,\boldsymbol{\mathcal{E}}_T] + \mathbb{P}^{\overline{\pi}}_{s_0}[\,\overline{\boldsymbol{\mathcal{E}}}_{\alpha,T}]$, so $\mathbb{J}_{\alpha,T}(\overline{\pi})$ can be written as an expectation as we wanted.

Coming back to (B.40), we expand the space of states in $\underline{\mathcal{M}}_T$ with two new components. With a small abuse of notation, we write

$$\mathcal{S}_{T+1} \;=\; \mathcal{X} \times \Big([T\,C^{\min}_1,\, T\,C^{\max}_1] \cap \mathbb{Q}\Big) \times \{0,1,\ldots,T+1\} \times \{0,1\} \times \{0,1\}$$

so that, a state $s_t \in \mathcal{S}_{T+1}$ is such that $\mathrm{proj}_4(s_t) = \mathrm{proj}_5(s_t) = 1$ for $t = 0$, and for $t = 1,\ldots,T$ this is such that:

$$\begin{aligned}\mathrm{proj}_4(s_t) \;&:=\; (1 \;-\; \mathbb{1}_{\mathcal{G}_T}(s_{t-1}))\,\mathrm{proj}_4(s_{t-1})\\ \mathrm{proj}_5(s_t) \;&:=\; \mathbb{1}_{\overline{\mathcal{G}}}(\mathrm{proj}_1(s_{t-1}))\,\mathrm{proj}_5(s_{t-1})\end{aligned} \qquad \text{(B.42)}$$

Without loss of generality, because at level $T+1$ there is a fresh state s_{T+1} for each leaf s_T, we can consider that such a state is of the form $s_{T+1} = s_T$, where the cost function between the transition from s_T to s_{T+1} is $C'(s_T, s_{T+1}) = \mathbb{SP}^{\pi_\gamma}_{s_T}$, i.e., the SSP-value (5.5) computed from $\mathrm{proj}_1(s_T)$ to $\mathcal{G}$ by using the strategy π_γ. Associated with such an expanded space of states, we add another cost function in $\underline{\mathcal{M}}_T$, defined for each $t = 0, ..., T-1$ by:

$$C'(s_t) \quad := \quad \mathbb{1}_{\mathcal{G}_T}(s_t)\,\mathrm{proj}_4(s_t)\ ,$$

and at level $t = T$, by:

$$C'(s_T) \quad := \quad \mathbb{1}_{\mathcal{G}_T}(s_T)\,\mathrm{proj}_4(s_T)\ +\ \alpha\,\mathbb{1}_{\overline{\mathcal{G}}}(\mathrm{proj}_1(s_T))\,\mathrm{proj}_5(s_T)\ .$$

Based on the fact that the objective function $\mathbb{J}_{\alpha,T}(\overline{\pi})$ can be written as an expectation, $\mathbb{L}_{\alpha,T}(\overline{\pi}, \lambda)$ can be so rewritten as:

$$\begin{aligned} \mathbb{L}_{\alpha,T}(\overline{\pi}, \lambda) &= \mathbb{E}^{\overline{\pi}}_{s_0}\left[\sum_{t=0}^{T} C'(s_t)\right] + \lambda\Big(\mathbb{E}^{\overline{\pi}}_{s_0}[\mathrm{TS}_{T+1}] - \nu\Big) \\ &= \mathbb{E}^{\overline{\pi}}_{s_0}\left[\sum_{t=0}^{T} C'(s_t)\ +\ \lambda C(s_t, a_t)\ +\ \lambda\Big(C'(s_T, s_{T+1}) - \nu\Big)\right] \end{aligned}$$

Thus, from each state s_{T+1}, we can compute recursively the minimum value at each step for the Lagrange function by the classical Bellman dynamic equation [21]. ■

Proof B.2.14: [Proposition 5.4.11]

Let $n \in \mathbb{N}$. For $0 < \underline{\lambda}_n \leq \overline{\lambda}_n$ fixed, let $\sigma^*_{\alpha,n} \in \Delta[\{\overline{\pi}^*_{\alpha,\underline{\lambda}_n}, \overline{\pi}^*_{\alpha,\overline{\lambda}_n}\}]$ a mixed strategy combining convexly between the pure strategies:

$$\overline{\pi}^*_{\alpha,\underline{\lambda}_n} \in \arg\min_{\overline{\pi}}\ \mathbb{L}_{\alpha,T}(\overline{\pi}, \underline{\lambda}_n) \quad \text{and} \quad \overline{\pi}^*_{\alpha,\overline{\lambda}_n} \in \arg\min_{\overline{\pi}}\ \mathbb{L}_{\alpha,T}(\overline{\pi}, \overline{\lambda}_n)\ ,$$

and defined as:

$$\begin{aligned} \sigma^*_{\alpha,n}(\overline{\pi}^*_{\alpha,\underline{\lambda}_n}) \quad &= \quad \frac{\nu\ -\ \mathbb{E}^{\overline{\pi}^*_{\alpha,\overline{\lambda}_n}}_{s_0}[\mathrm{TS}^{\mathcal{G}}]}{\mathbb{E}^{\overline{\pi}^*_{\alpha,\underline{\lambda}_n}}_{s_0}[\mathrm{TS}^{\mathcal{G}}]\ -\ \mathbb{E}^{\overline{\pi}^*_{\alpha,\overline{\lambda}_n}}_{s_0}[\mathrm{TS}^{\mathcal{G}}]} \\ \sigma^*_{\alpha,n}(\overline{\pi}^*_{\alpha,\overline{\lambda}_n}) \quad &= \quad \frac{\mathbb{E}^{\overline{\pi}^*_{\alpha,\underline{\lambda}_n}}_{s_0}[\mathrm{TS}^{\mathcal{G}}]\ -\ \nu}{\mathbb{E}^{\overline{\pi}^*_{\alpha,\underline{\lambda}_n}}_{s_0}[\mathrm{TS}^{\mathcal{G}}]\ -\ \mathbb{E}^{\overline{\pi}^*_{\alpha,\overline{\lambda}_n}}_{s_0}[\mathrm{TS}^{\mathcal{G}}]} \end{aligned} \tag{5.37}$$

Then, $\mathbb{J}_{\alpha,T}(\sigma^*_{\alpha,n}) \longrightarrow \mathbb{J}^{m*}_{\alpha,T}$ as $n \to +\infty$. Moreover, there exists a constant $E \in \mathbb{R}^+$, such that the number of iterations needed $n \in \mathbb{N}$ to achieve a given tolerance $\epsilon > 0$, is such that:

$$n \ \geq\ \log_2\left(E\,\frac{(\overline{\lambda}_0 - \underline{\lambda}_0)}{\epsilon}\right)$$

Proof. First, we have the following:

$$
\begin{aligned}
0 \;&\leq \mathbb{J}_{\alpha,T}(\sigma^*_{\alpha,n}) \;-\; \mathbb{J}^{\mathrm{m}*}_{\alpha,T} \qquad\qquad (\mathrm{B}.43)\\
&\quad \text{(by the definition (5.20) of the mixed strategy optimal value)}\\
&= \sigma^*_{\alpha,n}(\overline{\pi}^*_{\alpha,\underline{\lambda}_n})\,\mathbb{J}_{\alpha,T}(\overline{\pi}^*_{\alpha,\underline{\lambda}_n}) \;+\; \sigma^*_{\alpha,n}(\overline{\pi}^*_{\alpha,\overline{\lambda}_n})\,\mathbb{J}_{\alpha,T}(\overline{\pi}^*_{\alpha,\overline{\lambda}_n}) \;-\; \mathbb{J}^{\mathrm{m}*}_{\alpha,T}\\
&\quad \text{(because } \sigma \mapsto \mathbb{J}_{\alpha,T}(\sigma) \text{ is linear on mixed strategies)}\\
&= \sigma^*_{\alpha,n}(\overline{\pi}^*_{\alpha,\underline{\lambda}_n})(\mathbb{J}_{\alpha,T}(\overline{\pi}^*_{\alpha,\underline{\lambda}_n}) - \mathbb{J}^{\mathrm{m}*}_{\alpha,T}) + \sigma^*_{\alpha,n}(\overline{\pi}^*_{\alpha,\overline{\lambda}_n})(\mathbb{J}_{\alpha,T}(\overline{\pi}^*_{\alpha,\overline{\lambda}_n}) - \mathbb{J}^{\mathrm{m}*}_{\alpha,T})\\
&\quad \text{(arranging terms and because } \sigma^*_{\alpha,n}(\overline{\pi}^*_{\alpha,\underline{\lambda}_n}) + \sigma^*_{\alpha,n}(\overline{\pi}^*_{\alpha,\overline{\lambda}_n}) = 1)
\end{aligned}
$$

On the other hand, it also holds:

$$
\begin{aligned}
\mathbb{J}^{\mathrm{m}*}_{\alpha,T} \;&=\; \mathbb{L}^{\mathrm{p}*}_{\alpha,T}\\
&\quad \text{(by Propositions 5.4.4 and 5.4.5)}\\
&= \sup_{\lambda\geq 0}\; \mathbb{L}^{\mathrm{p}}_{\alpha,T}(\lambda)\\
&\quad \text{(by definition (5.26) of the optimal dual value)}\\
&\geq \mathbb{L}^{\mathrm{p}}_{\alpha,T}(\lambda)\\
&\quad \text{(for any } \lambda \geq 0\text{)}\\
&= \mathbb{J}_{\alpha,T}\left(\overline{\pi}^*_{\alpha,\lambda}\right) \;+\; \lambda\left(\mathbb{E}^{\overline{\pi}^*_{\alpha,\lambda}}_{s_0}[\,\mathrm{TS}^{\mathcal{G}}] \;-\; \nu\right)\\
&\quad \text{(by definition (5.24) of the Lagrange dual function)}
\end{aligned}
$$

The latter inequality is true $\forall\, \lambda \geq 0$, in particular for the dual variables $\underline{\lambda}_n$ and $\overline{\lambda}_n$. Thus, the two following inequalities hold:

$$
\begin{aligned}
\mathbb{J}_{\alpha,T}(\overline{\pi}^*_{\alpha,\underline{\lambda}_n}) \;-\; \mathbb{J}^{\mathrm{m}*}_{\alpha,T} \;&\leq\; \underline{\lambda}_n\left(\nu \;-\; \mathbb{E}^{\overline{\pi}^*_{\alpha,\underline{\lambda}_n}}_{s_0}[\,\mathrm{TS}^{\mathcal{G}}]\right)\\
\mathbb{J}_{\alpha,T}(\overline{\pi}^*_{\alpha,\overline{\lambda}_n}) \;-\; \mathbb{J}^{\mathrm{m}*}_{\alpha,T} \;&\leq\; \overline{\lambda}_n\left(\nu \;-\; \mathbb{E}^{\overline{\pi}^*_{\alpha,\overline{\lambda}_n}}_{s_0}[\,\mathrm{TS}^{\mathcal{G}}]\right)
\end{aligned}
\qquad (\mathrm{B}.44)
$$

Using the last two inequalities, it follows that:

$$
\begin{aligned}
\mathbb{J}^{\mathrm{m}}_{\alpha,T}(\sigma^*_{\alpha,n}) \;-\; \mathbb{J}^{\mathrm{m}*}_{\alpha,T} &= \sigma^*_{\alpha,n}(\overline{\pi}^*_{\alpha,\underline{\lambda}_n})(\mathbb{J}_{\alpha,T}(\overline{\pi}^*_{\alpha,\underline{\lambda}_n}) \;-\; \mathbb{J}^{\mathrm{m}*}_{\alpha,T}) \;+\\
&\quad \sigma^*_{\alpha,n}\left(\overline{\pi}^*_{\alpha,\overline{\lambda}_n}\right)(\mathbb{J}_{\alpha,T}\left(\overline{\pi}^*_{\alpha,\overline{\lambda}_n}\right) \;-\; \mathbb{J}^{\mathrm{m}*}_{\alpha,T})\\
&\quad \text{(by equality (B.43))}\\
&\leq \sigma^*_{\alpha,n}(\overline{\pi}^*_{\alpha,\underline{\lambda}_n})\,\underline{\lambda}_n\left(\nu - \mathbb{E}^{\overline{\pi}^*_{\alpha,\underline{\lambda}_n}}_{s_0}[\mathrm{TS}^{\mathcal{G}}]\right) \;+\\
&\quad \sigma^*_{\alpha,n}(\overline{\pi}^*_{\alpha,\overline{\lambda}_n})\,\overline{\lambda}_n\left(\nu - \mathbb{E}^{\overline{\pi}^*_{\alpha,\overline{\lambda}_n}}_{s_0}[\mathrm{TS}^{\mathcal{G}}]\right) \qquad (\mathrm{B}.45)\\
&\quad \text{(by using the inequalities (B.44))}\\
&= \frac{\left(\mathbb{E}^{\overline{\pi}^*_{\alpha,\underline{\lambda}_n}}_{s_0}[\mathrm{TS}^{\mathcal{G}}] - \nu\right)\left(\nu - \mathbb{E}^{\overline{\pi}^*_{\alpha,\overline{\lambda}_n}}_{s_0}[\mathrm{TS}^{\mathcal{G}}]\right)}{\mathbb{E}^{\overline{\pi}^*_{\alpha,\underline{\lambda}_n}}_{s_0}[\mathrm{TS}^{\mathcal{G}}] - \mathbb{E}^{\overline{\pi}^*_{\alpha,\overline{\lambda}_n}}_{s_0}[\mathrm{TS}^{\mathcal{G}}]}(\overline{\lambda}_n - \underline{\lambda}_n)\\
&\quad \text{(by using the definition (5.37) of } \sigma^*_{\alpha,n}\text{)}
\end{aligned}
$$

Now, by using the so-called bisection method, we can consider the midpoint of the interval $[\,\underline{\lambda}_n,\, \overline{\lambda}_n\,] \ni \lambda^*$ at the iteration $n+1$, i.e., the point:

$$
\lambda_{n+1} \;:=\; \frac{\overline{\lambda}_n + \underline{\lambda}_n}{2}
$$

and thus, the difference between λ_n and λ^* is bounded by:

$$\mid \lambda_{n+1} - \lambda^* \mid \ \leq \ \frac{\overline{\lambda}_0 - \underline{\lambda}_0}{2^{n+1}} \ . \tag{B.46}$$

On the other hand, we define

$$E_n \ := \ \frac{\left(\mathbb{E}_{s_0}^{\overline{\pi}^*_{\alpha,\underline{\lambda}_n}}[\mathrm{TS}^{\mathcal{G}}] - \nu\right)\left(\nu_2 - \mathbb{E}_{s_0}^{\overline{\pi}^*_{\alpha,\overline{\lambda}_n}}[\mathrm{TS}^{\mathcal{G}}]\right)}{\mathbb{E}_{s_0}^{\overline{\pi}^*_{\alpha,\underline{\lambda}_n}}[\mathrm{TS}^{\mathcal{G}}] - \mathbb{E}_{s_0}^{\overline{\pi}^*_{\alpha,\overline{\lambda}_n}}[\mathrm{TS}^{\mathcal{G}_T}]}$$

to show that the sequence $(E_n)_{n\in\mathbb{N}}$ is nondecreasing. First, because

$$\mathbb{E}_{s_0}^{\overline{\pi}^*_{\alpha,\overline{\lambda}_n}}[\mathrm{TS}^{\mathcal{G}}] \ \leq \ \nu \ \leq \ \mathbb{E}_{s_0}^{\overline{\pi}^*_{\alpha,\underline{\lambda}_n}}[\mathrm{TS}^{\mathcal{G}_T}]$$

for each $n \in \mathbb{N}$ and second, since the sequences $(\underline{\lambda}_n)_{n\in\mathbb{N}_0}$ and $(\overline{\lambda}_n)_{n\in\mathbb{N}_0}$ are resp. nondecreasing and nonincreasing, then by Lemma 5.4.1,

$$\underline{\lambda}_n \ \mapsto \ \mathbb{E}_{s_0}^{\overline{\pi}^*_{\alpha,\underline{\lambda}_n}}[\mathrm{TS}^{\mathcal{G}}] \qquad \text{and} \qquad \overline{\lambda}_n \ \mapsto \ \mathbb{E}_{s_0}^{\overline{\pi}^*_{\alpha,\overline{\lambda}_n}}[\mathrm{TS}^{\mathcal{G}}]$$

are resp. nonincreasing and nondecreasing. Thus, we infer that:

$$\begin{aligned}&\left(\mathbb{E}_{s_0}^{\overline{\pi}^*_{\alpha,\underline{\lambda}_{n+1}}}[\mathrm{TS}^{\mathcal{G}}] - \nu\right)\left(\nu - \mathbb{E}_{s_0}^{\overline{\pi}^*_{\alpha,\overline{\lambda}_{n+1}}}[\mathrm{TS}^{\mathcal{G}}]\right) \\ &\qquad\qquad \geq \ \left(\mathbb{E}_{s_0}^{\overline{\pi}^*_{\alpha,\underline{\lambda}_n}}[\mathrm{TS}^{\mathcal{G}}] - \nu\right)\left(\nu - \mathbb{E}_{s_0}^{\overline{\pi}^*_{\alpha,\overline{\lambda}_n}}[\mathrm{TS}^{\mathcal{G}}]\right)\end{aligned}$$

and that:

$$\frac{1}{\mathbb{E}_{s_0}^{\overline{\pi}^*_{\alpha,\underline{\lambda}_{n+1}}}[\mathrm{TS}^{\mathcal{G}}] - \mathbb{E}_{s_0}^{\overline{\pi}^*_{\alpha,\overline{\lambda}_{n+1}}}[\mathrm{TS}^{\mathcal{G}}]} \ \geq \ \frac{1}{\mathbb{E}_{s_0}^{\overline{\pi}^*_{\alpha,\underline{\lambda}_n}}[\mathrm{TS}^{\mathcal{G}}] - \mathbb{E}_{s_0}^{\overline{\pi}^*_{\alpha,\overline{\lambda}_n}}[\mathrm{TS}^{\mathcal{G}}]} \ .$$

Thus, it holds:

$$\begin{aligned}E_{n+1} \ &= \ \frac{\left(\mathbb{E}_{s_0}^{\overline{\pi}^*_{\alpha,\underline{\lambda}_{n+1}}}[\mathrm{TS}^{\mathcal{G}}] - \nu\right)\left(\nu - \mathbb{E}_{s_0}^{\overline{\pi}^*_{\alpha,\overline{\lambda}_{n+1}}}[\mathrm{TS}^{\mathcal{G}}]\right)}{\mathbb{E}_{s_0}^{\overline{\pi}^*_{\alpha,\underline{\lambda}_{n+1}}}[\mathrm{TS}^{\mathcal{G}}] - \mathbb{E}_{s_0}^{\overline{\pi}^*_{\alpha,\overline{\lambda}_{n+1}}}[\mathrm{TS}^{\mathcal{G}}]} \\ &\geq \ \frac{\left(\mathbb{E}_{s_0}^{\overline{\pi}^*_{\alpha,\underline{\lambda}_n}}[\mathrm{TS}^{\mathcal{G}}] - \nu\right)\left(\nu - \mathbb{E}_{s_0}^{\overline{\pi}^*_{\alpha,\overline{\lambda}_n}}[\mathrm{TS}^{\mathcal{G}}]\right)}{\mathbb{E}_{s_0}^{\overline{\pi}^*_{\alpha,\underline{\lambda}_n}}[\mathrm{TS}^{\mathcal{G}}] - \mathbb{E}_{s_0}^{\overline{\pi}^*_{\alpha,\overline{\lambda}_n}}[\mathrm{TS}^{\mathcal{G}}]} \ = \ E_n \ .\end{aligned}$$

Then, the sequence $(E_n)_{n\in\mathbb{N}}$ is nondecreasing. In addition, because there is a finite number of pure strategies, there exists $E \in \mathbb{R}^+$, such that:

$$E_n \ \leq \ E \qquad , \forall\, n \in \mathbb{N} \tag{B.47}$$

Coming back to (B.45), the following holds:

$$\begin{aligned}\mathbb{J}^{\mathrm{m}}_{\alpha,T}(\sigma^*_{\alpha,n}) - \mathbb{J}^{\mathrm{m}*}_{\alpha,T} \ &\leq \frac{\left(\mathbb{E}_{s_0}^{\overline{\pi}^*_{\alpha,\underline{\lambda}_n}}[\mathrm{TS}^{\mathcal{G}}] - \nu\right)\left(\nu - \mathbb{E}_{s_0}^{\overline{\pi}^*_{\alpha,\overline{\lambda}_n}}[\mathrm{TS}^{\mathcal{G}}]\right)}{\mathbb{E}_{s_0}^{\overline{\pi}^*_{\alpha,\underline{\lambda}_n}}[\mathrm{TS}^{\mathcal{G}}] - \mathbb{E}_{s_0}^{\overline{\pi}^*_{\alpha,\overline{\lambda}_n}}[\mathrm{TS}^{\mathcal{G}}]}(\overline{\lambda}_n - \underline{\lambda}_n) \\ &\qquad \text{(by inequality (B.45))} \\ &\leq E\,\frac{\left(\overline{\lambda}_0 - \underline{\lambda}_0\right)}{2^n} \\ &\qquad \text{(by using the inequality (B.46))}\end{aligned}$$

Thus, taking $n \to +\infty$ above, we have the first affirmation of the Proposition. For the second one, let $\epsilon > 0$ such that for a $n \in \mathbb{N}$ large enough,

$$E\,\frac{\left(\overline{\lambda}_0 - \underline{\lambda}_0\right)}{2^n} \quad \leq \quad \epsilon\ .$$

Applying $\log_2$ on both sides on the previous inequality, it holds:

$$n \quad \geq \quad \log_2\left(E\,\frac{(\overline{\lambda}_0 - \underline{\lambda}_0)}{\epsilon}\right)$$

$\blacksquare$

BIBLIOGRAPHY

[1] URL: www.energy.gov.

[2] URL: www.statista.com.

[3] URL: www.ausgrid.com.au.

[4] URL: www.prismmodelchecker.org.

[5] URL: www.bom.gov.au/climate/data-services.

[6] URL: www.aemo.com.au.

[7] AIE. *Global Electric Vehicles Outlook*. 2018. URL: https://www.connaissancedesenergies.org/sites/default/files/pdf-actualites/globalevoutlook2018.pdf.

[8] E. Allen and M. Ilic. *Price-based commitment decisions in the electricity market*. Springer Science & Business Media, 2012.

[9] E. Altman. *Constrained Markov decision processes*. Vol. 7. CRC Press, 1999.

[10] L. Andrieu, G. Cohen, and F. Vázquez-Abad. “Stochastic programming with probability.” In: *arXiv preprint arXiv:0708.0281* (2007).

[11] K. J. Åström. *Introduction to stochastic control theory*. Courier Corporation, 2012.

[12] R. J. Aumann. *Mixed and behavior strategies in infinite extensive games*. Tech. rep. PRINCETON UNIV NJ, 1961.

[13] S. Bahrami, V. Wong, and J. Huang. “An online learning algorithm for demand response in smart grid.” In: *IEEE Transactions on Smart Grid* 9.5 (2017), pp. 4712–4725.

[14] C. Baier, N. Bertrand, C. Dubslaff, D. Gburek, and O. Sankur. “Stochastic Shortest Paths and Weight-Bounded Properties in Markov Decision Processes.” In: *Proceedings of the 33rd Annual ACM/IEEE Symposium on Logic in Computer Science*. ACM. 2018, pp. 86–94.

[15] C. Baier, C. Dubslaff, and S. Klüppelholz. “Trade-off analysis meets probabilistic model checking.” In: *Proceedings of the Joint Meeting of the Twenty-Third EACSL Annual Conference on Computer Science Logic (CSL) and the Twenty-Ninth Annual ACM/IEEE Symposium on Logic in Computer Science (LICS)*. ACM. 2014, p. 1.

[16] C. Baier and J.-P. Katoen. *Principles of model checking*. MIT press, 2008.

[17] O. Beaude. “Modélisation et optimisation de l’interaction entre véhicules électriques et réseaux d’électricité: apport de la théorie des jeux.” PhD thesis. Université Paris-Saclay, 2015.

[18] O. Beaude, S. Lasaulce, and M. Hennebel. "Charging games in networks of electrical vehicles." In: *2012 6th International Conference on Network Games, Control and Optimization (NetGCooP)*. IEEE. 2012, pp. 96–103.

[19] O. Beaude, S. Lasaulce, M. Hennebel, and J. Daafouz. "Minimizing the impact of EV charging on the electricity distribution network." In: *2015 European Control Conference (ECC)*. IEEE. 2015, pp. 648–653.

[20] O. Beaude, S. Lasaulce, Hennebel M., and I. Mohand-Kaci. "Reducing the impact of EV charging operations on the distribution network." In: *IEEE Trans. on Smart Grid* 7.6 (2016), pp. 2666–2679.

[21] R. Bellman. "Dynamic programming." In: *Science* 153.3731 (1966), pp. 34–37.

[22] D. S. Bernstein, R. Givan, N. Immerman, and S. Zilberstein. "The complexity of decentralized control of Markov decision processes." In: *Mathematics of operations research* 27.4 (2002), pp. 819–840.

[23] D. P. Bertsekas. *Nonlinear programming*. Athena scientific Belmont, 1999.

[24] D. P. Bertsekas. *Convex optimization theory*. Athena Scientific Belmont, 2009.

[25] D. P. Bertsekas and J. N. Tsitsiklis. "An analysis of stochastic shortest path problems." In: *Mathematics of Operations Research* 16.3 (1991), pp. 580–595.

[26] H. Bessembinder and M. L. Lemmon. "Equilibrium pricing and optimal hedging in electricity forward markets." In: *the Journal of Finance* 57.3 (2002), pp. 1347–1382.

[27] F. J. Beutler and K. W. Ross. "Optimal policies for controlled Markov chains with a constraint." In: (1985).

[28] E. Bitar, R. Rajagopal, P. Khargonekar, and K. Poolla. "The role of co-located storage for wind power producers in conventional electricity markets." In: *Proceedings of the 2011 American Control Conference*. IEEE. 2011, pp. 3886–3891.

[29] P. Bouyer, N. Markey, J. Olschewski, and M. Ummels. "Measuring permissiveness in parity games: Mean-payoff parity games revisited." In: *International Symposium on Automated Technology for Verification and Analysis*. Springer. 2011, pp. 135–149.

[30] S. Boyd and L. Vandenberghe. *Convex optimization*. Cambridge university press, 2004.

[31] T. Brázdil, K. Chatterjee, V. Forejt, and A. Kučera. "Trading performance for stability in Markov decision processes." In: *Journal of Computer and System Sciences* 84 (2017), pp. 144–170.

[32] T. Brihaye, A. K. Dhar, G. Geeraerts, A. Haddad, and B. Monmege. "Efficient Energy Distribution in a Smart Grid using Multi-Player Games." In: *arXiv preprint arXiv:1608.00652* (2016).

[33] G. W. Brown. “Iterative solution of games by fictitious play.” In: *Activity analysis of production and allocation* 13.1 (1951), pp. 374–376.

[34] V. Bruyere, E. Filiot, M. Randour, and J.-F. Raskin. “Meet your expectations with guarantees: Beyond worst-case synthesis in quantitative games.” In: *Information and Computation* 254 (2017), pp. 259–295.

[35] L. Buşoniu, R. Babuška, and B. De Schutter. “Multi-agent reinforcement learning: An overview.” In: *Innovations in multi-agent systems and applications-1*. Springer, 2010, pp. 183–221.

[36] K. Chatterjee, L. Doyen, T. A Henzinger, and J.-F. Raskin. “Generalized mean-payoff and energy games.” In: *arXiv preprint arXiv:1007.1669* (2010).

[37] K. Chatterjee, Z. Komarkova, and J. Kretínskỳ. “Unifying two views on multiple mean-payoff objectives in Markov decision processes.” In: *2015 30th Annual ACM/IEEE Symposium on Logic in Computer Science*. IEEE. 2015, pp. 244–256.

[38] B. V. Cherkassky, A. V. Goldberg, and T. Radzik. “Shortest paths algorithms: Theory and experimental evaluation.” In: *Mathematical programming* 73.2 (1996), pp. 129–174.

[39] A. Chiş, J. Lundén, and V. Koivunen. “Reinforcement learning-based plug-in electric vehicle charging with forecasted price.” In: *IEEE Transactions on Vehicular Technology* 66.5 (2016), pp. 3674–3684.

[40] S. L. Chiu. “Fuzzy model identification based on cluster estimation.” In: *Journal of Intelligent & fuzzy systems* 2.3 (1994), pp. 267–278.

[41] E. M. Clarke Jr, O. Grumberg, D. Kroening, D. Peled, and H. Veith. *Model checking*. MIT press, 2018.

[42] E. M. Clarke and E. A. Emerson. “Design and synthesis of synchronization skeletons using branching time temporal logic.” In: *Workshop on Logic of Programs*. Springer. 1981, pp. 52–71.

[43] E. M Clarke and J. M. Wing. “Formal methods: State of the art and future directions.” In: *ACM Computing Surveys (CSUR)* 28.4 (1996), pp. 626–643.

[44] K. Clement-Nyns, E. Haesen, and J. Driesen. “The impact of charging plug-in hybrid electric vehicles on a residential distribution grid.” In: *IEEE Transactions on power systems* 25.1 (2009), pp. 371–380.

[45] K. Clement, E. Haesen, and J. Driesen. “Coordinated charging of multiple plug-in hybrid electric vehicles in residential distribution grids.” In: *Power Systems Conference and Exposition (PSCE’09). IEEE/PES*. IEEE. 2009, pp. 1–7.

[46] C. Daskalakis, P. W. Goldberg, and C. H. Papadimitriou. “The complexity of computing a Nash equilibrium.” In: *SIAM Journal on Computing* 39.1 (2009), pp. 195–259.

[47] L. De Alfaro. "Computing minimum and maximum reachability times in probabilistic systems." In: *International Conference on Concurrency Theory*. Springer. 1999, pp. 66–81.

[48] A. J. Dix. *Formal methods for interactive systems*. Vol. 16. Academic Press London, 1991.

[49] J. Donadee and M. D. Ilić. "Stochastic optimization of grid to vehicle frequency regulation capacity bids." In: *IEEE Transactions on Smart Grid* 5.2 (2014), pp. 1061–1069.

[50] K. Etessami, M. Kwiatkowska, M. Y Vardi, and M. Yannakakis. "Multi-objective model checking of Markov decision processes." In: *International Conference on Tools and Algorithms for the Construction and Analysis of Systems*. Springer. 2007, pp. 50–65.

[51] A. Fabrikant, C. Papadimitriou, and K. Talwar. "The complexity of pure Nash equilibria." In: *Proceedings of the thirty-sixth annual ACM symposium on Theory of computing*. ACM. 2004, pp. 604–612.

[52] W. Feller. *An introduction to probability theory and its applications*. Vol. 2. John Wiley & Sons, 2008.

[53] V. Forejt, M. Kwiatkowska, G. Norman, and D. Parker. "Automated verification techniques for probabilistic systems." In: *International School on Formal Methods for the Design of Computer, Communication and Software Systems*. Springer. 2011, pp. 53–113.

[54] N. Forouzandehmehr, M. Esmalifalak, H. Mohsenian-Rad, and Z. Han. "Autonomous demand response using stochastic differential games." In: *IEEE Transactions on Smart Grid* 6.1 (2014), pp. 291–300.

[55] N. Fulton and A. Platzer. "Safe reinforcement learning via formal methods: Toward safe control through proof and learning." In: *Thirty-Second AAAI Conference on Artificial Intelligence*. 2018.

[56] L. Gan, U. Topcu, and S. H. Low. "Optimal decentralized protocol for electric vehicle charging." In: *IEEE Transactions on Power Systems* 28.2 (2012), pp. 940–951.

[57] R. Garcia-Valle and J. A. Lopes. *Electric vehicle integration into modern power networks*. Springer Science & Business Media, 2012.

[58] A. V Goldberg, É. Tardos, and R. Tarjan. *Network flow algorithms*. Tech. rep. PRINCETON UNIV NJ DEPT OF COMPUTER SCIENCE, 1989.

[59] J. C. Gómez and M. M. Morcos. "Impact of EV battery chargers on the power quality of distribution systems." In: *IEEE Transactions on Power Delivery* 18.3 (2003), pp. 975–981.

[60] A. Gouberman and M. Siegle. "Markov reward models and Markov decision processes in discrete and continuous time: Performance evaluation and optimization." In: *International Autumn School on Rigorous Dependability Analysis Using Model Checking Techniques for Stochastic Systems*. Springer. 2012, pp. 156–241.

[61] E. Gradel and W. Thomas. *Automata, logics, and infinite games: a guide to current research*. Vol. 2500. Springer Science & Business Media, 2002.

[62] C. Haase and S. Kiefer. "The odds of staying on budget." In: *International Colloquium on Automata, Languages, and Programming*. Springer. 2015, pp. 234–246.

[63] D. Harel. "Statecharts: A visual formalism for complex systems." In: *Science of computer programming* 8.3 (1987), pp. 231–274.

[64] D. Harel and A. Pnueli. "On the development of reactive systems." In: *Logics and models of concurrent systems*. Springer, 1985, pp. 477–498.

[65] C. M. Holloway. "Why engineers should consider formal methods." In: *16th DASC. AIAA/IEEE Digital Avionics Systems Conference. Reflections to the Future. Proceedings*. Vol. 1. IEEE. 1997, pp. 1–3.

[66] J. Hu, M. P. Wellman, et al. "Multiagent reinforcement learning: theoretical framework and an algorithm." In: *ICML*. Vol. 98. Citeseer. 1998, pp. 242–250.

[67] Q. Huang, Q.-S. Jia, Z. Qiu, X. Guan, and G. Deconinck. "Matching EV charging load with uncertain wind: A simulation-based policy improvement approach." In: *IEEE Transactions on Smart Grid* 6.3 (2015), pp. 1425–1433.

[68] IEEE. "Guide for loading mineral-oil-immersed transformers." In: *IEEE Std, C57.91-1995* (1995), pp. 860–865.

[69] A. Ipakchi and F. Albuyeh. "Grid of the future." In: *IEEE power and energy magazine* 7.2 (2009), pp. 52–62.

[70] L. Jauregui-Rivera and D. Tylavsky. "Acceptability of four transformer top-oil thermal models: Part I: Defining metrics." In: *IEEE Transactions on Power Delivery* 23.2 (2008), pp. 860–865.

[71] P. Kall and S. Wallace. *Stochastic programming*. Springer, 1994.

[72] L. Kelly, A. Rowe, and P. Wild. "Analyzing the impacts of plug-in electric vehicles on distribution networks in British Columbia." In: *2009 IEEE Electrical Power & Energy Conference (EPEC)*. IEEE. 2009, pp. 1–6.

[73] M. Kezunovic, V. Vittal, S. Meliopoulos, and T. Mount. "The big picture: Smart research for large-scale integrated smart grid solutions." In: *IEEE Power and Energy Magazine* 10.4 (2012), pp. 22–34.

[74] T. T. Kim and H. V. Poor. "Scheduling power consumption with price uncertainty." In: *IEEE Transactions on Smart Grid* 2.3 (2011), pp. 519–527.

[75] M. J. Kochenderfer. *Decision making under uncertainty: theory and application*. MIT press, 2015.

[76] J. Křetínskỳ and T. Meggendorfer. "Conditional value-at-risk for reachability and mean payoff in Markov decision processes." In: *Proceedings of the 33rd Annual ACM/IEEE Symposium on Logic in Computer Science*. ACM. 2018, pp. 609–618.

[77] H. Kuhn. "Extensive games and the problem of information." In: *InH. KuhnandA. Tucker, editors, Contributions to the Theory of Games* (1953), pp. 193–216.

[78] M. Kwiatkowska, G. Norman, and D. Parker. "PRISM 4.0: Verification of probabilistic real-time systems." In: *International conference on computer aided verification*. Springer. 2011, pp. 585–591.

[79] S. Lasaulce and H. Tembine. *Game theory and learning for wireless networks: fundamentals and applications*. Academic Press, 2011.

[80] W. Leterme, F. Ruelens, B. Claessens, and R. Belmans. "A flexible stochastic optimization method for wind power balancing with PHEVs." In: *IEEE Transactions on Smart Grid* 5.3 (2014), pp. 1238–1245.

[81] M. L. Littman. "Markov games as a framework for multi-agent reinforcement learning." In: *Machine learning proceedings 1994*. Elsevier, 1994, pp. 157–163.

[82] Z. Ma, D. S Callaway, and I. A. Hiskens. "Decentralized charging control of large populations of plug-in electric vehicles." In: *IEEE Transactions on Control Systems Technology* 21.1 (2011), pp. 67–78.

[83] S. Mannor and J. Tsitsiklis. "Mean-variance optimization in Markov decision processes." In: *arXiv preprint arXiv:1104.5601* (2011).

[84] M. Maschler, E. Solan, and S. Zamir. *Game theory*. 2013.

[85] F. S Melo and M. Veloso. "Decentralized MDPs with sparse interactions." In: *Artificial Intelligence* 175.11 (2011), pp. 1757–1789.

[86] H. Mohsenian-Rad and A. Leon-Garcia. "Optimal residential load control with price prediction in real-time electricity pricing environments." In: *IEEE Transaction on Smart Grid* 1.2 (2010), pp. 120–133.

[87] H. Mohsenian-Rad et al. "Optimal charging of electric vehicles with uncertain departure times: A closed-form solution." In: *IEEE Transactions on Smart Grid* 6.2 (2014), pp. 940–942.

[88] J. A. Momoh. *Smart grid: fundamentals of design and analysis*. Vol. 63. John Wiley & Sons, 2012.

[89] D. Monderer and L. S. Shapley. "Potential games." In: *Games and economic behavior* 14.1 (1996), pp. 124–143.

[90] N. Nisan, T. Roughgarden, E. Tardos, and V. V. Vazirani. *Algorithmic game theory*. Cambridge university press, 2007.

[91] D. O'Neill, M. Levorato, A. Goldsmith, and U. Mitra. "Residential demand response using reinforcement learning." In: *2010 First IEEE International Conference on Smart Grid Communications*. IEEE. 2010, pp. 409–414.

[92] C. Pang, P. Dutta, and M. Kezunovic. "BEVs/PHEVs as dispersed energy storage for V2B uses in the smart grid." In: *IEEE Transactions on Smart Grid* 3.1 (2011), pp. 473–482.

[93] C. H. Papadimitriou and J. N. Tsitsiklis. "The complexity of Markov decision processes." In: *Mathematics of operations research* 12.3 (1987), pp. 441–450.

[94] D. Pérez Palomar and J. Rodríguez Fonollosa. "Practical algorithms for a family of waterfilling solutions." In: (2005).

[95] A. Pnueli and R. Rosner. "On the synthesis of a reactive module." In: *Proceedings of the 16th ACM SIGPLAN-SIGACT symposium on Principles of programming languages*. ACM. 1989, pp. 179–190.

[96] M. L. Puterman. *Markov decision processes: discrete stochastic dynamic programming*. John Wiley & Sons, 2014.

[97] D. V. Pynadath and M. Tambe. "Multiagent teamwork: Analyzing the optimality and complexity of key theories and models." In: *Proceedings of the first international joint conference on Autonomous agents and multiagent systems: part 2*. ACM. 2002, pp. 873–880.

[98] F. Rahimi and A. Ipakchi. "Demand response as a market resource under the smart grid paradigm." In: *IEEE Transactions on smart grid* 1.1 (2010), pp. 82–88.

[99] S. Rajakaruna, F. Shahnia, and A. Ghosh. *Plug in electric vehicles in smart grids: charging strategies*. Springer, 2014.

[100] P. J. Ramadge and W. M. Wonham. "Supervisory control of a class of discrete event processes." In: *SIAM journal on control and optimization* 25.1 (1987), pp. 206–230.

[101] M. Randour, J.-F. Raskin, and O. Sankur. "Variations on the stochastic shortest path problem." In: *International Workshop on Verification, Model Checking, and Abstract Interpretation*. Springer. 2015, pp. 1–18.

[102] M. Randour, J.-F. Raskin, and O. Sankur. "Percentile queries in multi-dimensional Markov decision processes." In: *Formal Methods in System Design* 50.2-3 (2017), pp. 207–248.

[103] P. Samadi, H. Mohsenian-Rad, V. W. Wong, and R. Schober. "Tackling the load uncertainty challenges for energy consumption scheduling in smart grid." In: *IEEE Transactions on Smart Grid* 4.2 (2013), pp. 1007–1016.

[104] P. Samadi, H. Mohsenian-Rad, V. Wong, and R. Schober. "Adaptive energy consumption scheduling with load uncertainty for the smart grid." In: *2013 IEEE International Conference on Communications (ICC)*. IEEE. 2013, pp. 4244–4249.

[105] A. Sheikhi, Sh. Bahrami, AM. Ranjbar, and H. Oraee. "Strategic charging method for plugged in hybrid electric vehicles in smart grids; a game theoretic approach." In: *International Journal of Electrical Power & Energy Systems* 53 (2013), pp. 499–506.

[106] M. Shinwari, A. Youssef, and W. Hamouda. "A water-filling based scheduling algorithm for the smart grid." In: *IEEE Transactions on Smart Grid* 3.2 (2012), pp. 710–719.

[107] A. Subramanian, M. Garcia, A. Domínguez-García, D. Callaway, K. Poolla, and P. Varaiya. "Real-time scheduling of deferrable electric loads." In: *2012 American Control Conference (ACC)*. IEEE. 2012, pp. 3643–3650.

[108] R. S. Sutton and A. G. Barto. *Reinforcement learning: An introduction*. MIT press, 2018.

[109] M. Y. Vardi and P. Wolper. "An automata-theoretic approach to automatic program verification." In: *Proceedings of the First Symposium on Logic in Computer Science*. IEEE Computer Society. 1986, pp. 322–331.

[110] C. Villani. *Optimal transport: old and new*. Vol. 338. Springer Science & Business Media, 2008.

[111] J. Von Neumann and O. Morgenstern. *Theory of games and economic behavior (commemorative edition)*. Princeton university press, 2007.

[112] T. Wu, Q. Yang, Z. Bao, and W. Yan. "Coordinated energy dispatching in microgrid with wind power generation and plug-in electric vehicles." In: *IEEE Transactions on Smart Grid* 4.3 (2013), pp. 1453–1463.

[113] X. Xi and R. Sioshansi. "Using price-based signals to control plug-in electric vehicle fleet charging." In: *IEEE Transactions on Smart Grid* 5.3 (2014), pp. 1451–1464.

[114] W. Yu, G. Ginis, and J. M. Cioffi. "Distributed multiuser power control for digital subscriber lines." In: *IEEE Journal on Selected areas in Communications* 20.5 (2002), pp. 1105–1115.

Titre : Jeux Stochastiques sur des Graphes avec des Applications à l'Optimisation des Smart-Grids
Mots-clés : Méthodes formelles, Processus de décision de Markov, Analyse quantitative, Théorie des jeux, Optimisation stochastique, Smart grids.

Résumé : Au sein de la communauté scientifique, l'étude des nombreuses applications des réseaux d'énergie suscite un vif intérêt puisqu'elles deviennent de plus en plus importantes dans notre monde moderne. Des outils mathématiques avancés et complexes sont nécessaires afin de bien concevoir et mettre en œuvre ces réseaux. La précision et l'optimalité sont deux caractéristiques essentielles pour la conception de tels réseaux. Bien que ces deux aspects sont dans le cœur des méthodes formelles, leur application effective reste largement inexplorée aux réseaux d'énergie. Cela motive fortement le travail développé dans cette thèse. Un accent particulier est placé sur le problème général de planification de la consommation d'énergie. Il s'agit d'un scénario dans lequel les consommateurs ont besoin d'une certaine quantité d'énergie et souhaitent que cette demande soit satisfaite dans un délai de temps spécifique (e.g., un Véhicule Électrique (VE) doit être rechargé dans une fenêtre de temps définie par son propriétaire). Par conséquent, chaque consommateur doit choisir une puissance de consommation à chaque instant, afin que l'énergie finale accumulée atteigne un niveau souhaité. La manière dont les puissances sont choisies est en fonction d'une « stratégie »qui prend en compte à chaque instant les informations pertinentes d'un consommateur afin de choisir un niveau de consommation approprié (e.g., l'énergie accumulée pour recharge le VE). Les stratégies peuvent être conçues selon une approche centralisée (dans laquelle il n'y a qu'un seul décideur qui contrôle toutes les stratégies des consommateurs) ou décentralisée (dans laquelle il y a plusieurs contrôleurs, chacun représentant un consommateur). Nous analysons ces deux scénarios dans cette thèse en utilisant des méthodes formelles, la théorie des jeux et l'optimisation. Plus précisément, nous modélisons le problème de planification de la consommation d'énergie à l'aide des processus de décision de Markov et des jeux stochastiques. Par exemple, l'environnement du système électrique, à savoir : la partie non contrôlable de la consommation totale (e.g., la consommation hors VEs), peut être représentée via un modèle stochastique. La partie contrôlable de la consommation totale peut s'adapter aux contraintes du réseau de distribution (e.g., pour ne pas dépasser la température maximale d'arrêt du transformateur électrique) et à leurs objectifs (e.g., tous les VEs soient rechargés). À première vue, cela peut être vu comme un système stochastique avec des multi-objectifs sous contraintes. Par conséquent, cette thèse concerne également une contribution aux modèles avec des objectives multicritères, ce qui permet de poursuivre plusieurs objectifs à la fois et une conception des stratégies qui sont fonctionnellement correctes et robustes aux changements de l'environnement.

Title: Stochastic games on graphs with applications to smart-grids optimization
Keywords: Formal methods, Markov decision processes, Quantitative analysis, Game theory, Stochastic optimization, Smart grids.

Abstract: Within the research community, there is a great interest in exploring many applications of energy networks since these become more and more important in our modern world. To properly design and implement these networks, advanced and complex mathematical tools are necessary. Two key features for their design are correctness and optimality. While these last two properties are in the core of formal methods, their effective application to energy networks remains largely unexploited. This constitutes one strong motivation for the work developed in this thesis. A special emphasis is made on the generic problem of power consumption scheduling. This is a scenario in which the consumers have a certain energy demand and want to have this demand to be fulfilled before a set deadline (e.g., an Electric Vehicle (EV) has to be recharged within a given time window set by the EV owner). Therefore, each consumer has to choose at each time the consumption power so that the final accumulated energy reaches a desired level. The way in which the power levels are chosen is according to a "strategy" mapping at any time the relevant information of a consumer (e.g., the current accumulated energy for EV-charging) to a suitable power consumption level. The design of such strategies may be either centralized (in which there is a single decision-maker controlling all strategies of consumers), or decentralized (in which there are several decision-makers, each of them representing a consumer). We analyze both scenarios by exploiting the theory from formal methods, game theory and optimization. More specifically, the power consumption scheduling problem can be modelled using Markov decision processes and stochastic games. For instance, probabilities provide a way to model the environment of the electrical system, namely: the non-controllable part of the total consumption (e.g., the non-EV consumption). The controllable consumption can be adapted to the constraints of the distribution network (e.g., to the maximum shutdown temperature of the electrical transformer), and to their objectives (e.g., all EVs are recharged). At first glance, this can be seen as a stochastic system with multi-constraints objectives. Therefore, the contributions of this thesis also concern the area of multi-criteria objective models, which allows one to pursue several objectives at the time such as having strategy designs functionally correct and robust against changes of the environment.