

Stochastic Optimal Control of a Battery with Degradation in Stochastic Electricity Markets

John Huynh
Matrikelnummer: 401283

Ausarbeitung zur Erlangung des akademischen Grades

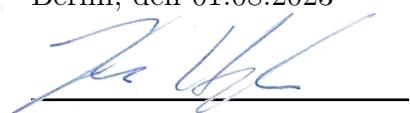
**Master of Science
Mathematik**

Erstgutachter: Dr. Christian Bayer
Zweitgutachter: Prof. Dr. Peter Friz

Eidesstattliche Erklärung zur Masterarbeit

Hiermit erkläre ich, dass ich die vorliegende Arbeit selbstständig und eigenhändig sowie ohne unerlaubte fremde Hilfe und ausschließlich unter Verwendung der aufgeführten Quellen und Hilfsmittel angefertigt habe.

Berlin, den 01.08.2023



John Huynh

Zusammenfassung

Im Zuge des weltweiten Strebens nach einer nachhaltigeren Zukunft stellt die Integration erneuerbarer Energiequellen in das Stromnetz eine komplexe Herausforderung dar. In diesem Zusammenhang erweisen sich Batteriespeichersysteme als eine Schlüsselkomponente für die erfolgreiche Energiewende. In dieser Arbeit wird ein stochastisch optimales Kontrollproblem für Batteriespeichersysteme aus der Perspektive eines Verbrauchers vorgestellt, dessen Ziel die Minimierung seiner Stromkosten über einen festgelegten Zeithorizont ist. Das Kontrollproblem umfasst ein realistisches Batteriedegradierungsmodell und ermöglicht die Einbeziehung von stochastischen Strompreisen, Wettervorhersagen und ungewissen Strombedarf. Wir formulieren dieses optimale Batteriekontrollproblem als ein Markov-Entscheidungsproblem und präsentieren einen kleinsten Quadrate Monte-Carlo-Algorithmus als Methode zur numerischen Lösung des Problems. Unsere numerischen Ergebnisse zeigen die Anwendbarkeit unseres Ansatzes und verdeutlichen, dass die Einbeziehung von Batteriedegradierung, stochastischen Strompreisen und Wettervorhersagen für die optimale Kontrollstrategie von Bedeutung sind.

Abstract

As the world strives for a more sustainable future, integrating renewable energy sources into power systems presents complex challenges. In this context, battery energy storage systems emerge as a key component in ensuring a successful transition. This thesis presents a stochastic optimal control framework for battery energy storage systems from the perspective of a consumer who seeks to minimize her cost of electricity over a fixed time horizon. The framework includes a realistic battery degradation model and allows for the incorporation of stochastic electricity prices, weather forecasts and uncertain electricity demand. We formulate this optimal battery control problem as a Markov decision problem and present a least squares Monte Carlo algorithm as a method to solve the problem numerically. Our numerical results demonstrate the feasibility of our approach and highlight the importance of incorporating battery degradation, stochastic electricity prices, and weather forecasts into the optimal control strategy.

Contents

1	Introduction	1
2	Modelling a battery energy storage system	3
2.1	Considerations in modeling the battery degradation	3
2.2	A battery degradation model	5
2.2.1	Battery variables	6
2.2.2	Dimensionless battery variables	7
2.2.3	Modelling the degradation state	8
2.2.4	Modelling the degradation state in discrete-time	14
3	Stochastic modelling of electricity markets	17
3.1	Modelling electricity prices	18
3.1.1	The Schwartz model	20
3.1.2	Modelling price spikes	21
3.1.3	Stochastic volatility models	25
3.2	Simulation	28
3.2.1	The seasonality component	29
3.2.2	The stochastic component	30
3.3	Modelling the impact of wind forecasts on electricity prices	32
4	The optimal battery control problem	35
4.1	Stochastic optimal control	35
4.1.1	Bellman equation	39
4.2	Numerical methods	44
4.2.1	Computational cost and convergence analysis	50
4.3	Optimal battery control	53
4.3.1	Case without degradation	55
4.3.2	Case with degradation	57
5	Numerical Results	60
5.1	Degradation experiments	60
5.2	Forecast experiments	67
6	Conclusion	70
References		72

1 Introduction

The global energy landscape has undergone a transformative shift towards renewable energy sources, such as wind and solar power, in response to the pressing need for sustainable and environmentally friendly alternatives. This ongoing global energy transition has brought about significant changes to the electricity sector, as these intermittent and variable renewable energy sources become increasingly prevalent in power systems. However, the integration of renewable energy into existing power systems presents a new set of challenges. The inherent intermittency and variability of renewable energy sources pose significant obstacles to the stability and reliability of electricity grids. The supply of renewable energy is subject to natural fluctuations, which can lead to mismatches between energy generation and demand. To address this issue and ensure a seamless integration of renewable energy, energy storage systems have emerged as a crucial solution. Among various energy storage technologies, battery energy storage systems, e.g. lithium-ion batteries, stand out as a key technology as they feature quick response time, high round-trip efficiency, pollution-free operation, and flexible power and energy ratings. These characteristics make it an ideal choice for a wide range of power system applications, including the integration of renewable energy, energy price arbitrage, and various ancillary services, such as load shifting, frequency regulation, and grid stabilization.

In this thesis, we take the view of a consumer operating a battery energy storage system whose objective is to minimize her cost of electricity over a fixed time horizon. Achieving this goal necessitates the implementation of advanced control strategies that are capable of accounting for various dynamic factors, such as the energy demand and the electricity prices on the market. In addition, a key factor is the precise estimation of the battery operating cost, which consists of the gradual loss of battery capacity and performance over time. Consequently, it becomes essential to develop accurate degradation models, since disregarding the battery degradation in the optimization process can lead to suboptimal control strategies that fail to accurately incorporate the true costs of operating the battery as well as the long-term implications on battery performance. Moreover, renewable energy sources are highly dependent on weather conditions. As such, weather forecasts provide valuable information as they predict the availability of cost-effective generated power by wind and solar. As a result, by considering forecasted weather data, one can optimize their battery operation in anticipation of the price fluctuations due to the renewable energy generation.

The aim of this thesis is to address these challenges associated with the optimal control of a battery storage system. We model this problem as a stochastic optimal control problem, which includes both a battery degradation model and an accurate stochastic model of electricity prices. Notably, we propose a realistic battery degradation model which also

takes into account the charging and discharging speeds. Additionally, our framework accommodates the incorporation of various stochastic factors, including (weather) forecasts and electricity demands. In order to solve the optimal battery control problem numerically, we also present a least squares Monte Carlo algorithm as a solution method.

The remainder of this thesis is organized as follows. In Chapter 2, we first briefly discuss the essential factors that must be considered while modelling battery degradation. In addition, we develop an approach to model the battery degradation based on the average charging and discharging speeds and total energy throughput. Chapter 3 introduces several stochastic electricity price models. In particular, we discuss their simulation and how wind forecasts can be incorporated into these models. The optimal battery control problem is formulated as a Markov decision problem in Chapter 4. We also recall fundamental theoretical results of Markov decision models to provide a solid foundation for our approach. Moreover, we describe a least squares Monte Carlo algorithm which serves as our key numerical tool to solve the optimal battery control problem. Furthermore, we offer a computational cost analysis and state a convergence result to validate the effectiveness of our proposed solution method. Finally, Chapter 5 presents our numerical results, where we demonstrate the importance of incorporating both degradation and forecasts into the models.

2 Modelling a battery energy storage system

In order to optimize the usage of a battery energy storage system, it is essential to dynamically take into account the state of the battery. This includes its current state of charge level and its capacity, and crucially, the effects of battery degradation. Battery degradation represents the gradual loss of battery capacity and performance over time. For instance, certain charge or discharge operations of the battery may appear attractive due to their immediate benefits, however, they could be outweighed by the adverse effects on the battery lifetime. In other words, different operational profiles may yield the same overall benefit during a specific period, yet their impact on battery degradation can significantly vary. In light of these complexities, the precise estimation of the battery operating cost, which depends on an accurate battery degradation model, becomes imperative. The degradation process is influenced by various factors, such as charge and discharge profiles, operating conditions, and temperature. As a result, disregarding battery degradation in the optimization process can lead to suboptimal control strategies that fail to accurately incorporate the cost of operating the battery but also the long-term implications on battery performance.

2.1 Considerations in modeling the battery degradation

In this section, we delve into the crucial factors to be considered when modelling battery degradation. Subsequently, we provide a concise overview of existing battery degradation models documented in the literature. For a more extensive examination of the topic, we direct readers to [Sha+22], which offers a comprehensive survey encompassing a broader scope of degradation modelling approaches.

Battery degradation is a gradual process characterized by a decline in capacity, decreasing operating voltage, increasing internal resistance, and energy loss, which occurs both during operation and during storage. Typically, the degradation can be divided into two factors: calendar ageing and cycling ageing. Calendar ageing is mainly time-dependent, in the sense that the battery capacity fades while the energy is stored rather than used. Here, factors such as cell temperature, the state of charge (SoC), i.e. the current level of stored energy in a battery relative to its total capacity, and the no-load voltage play crucial roles. More precisely, a high SoC leads to increased cell voltages, exerting greater mechanical stress on the battery materials, which reduces battery life. Temperature is probably the most important factor in calendar degradation, which is due to the fact that higher temperatures accelerate chemical degradation. This is why it usually makes sense to install a cooling device alongside a battery storage. In contrast to calendar ageing,

cycling ageing is mainly dependent on how the battery is operated. Important factors include the depth of discharge (DoD), and the charge and discharge rate (C-rate). Let us briefly explain the meaning behind these factors. The depth of discharge refers to the amount of capacity that has been utilized from the total available capacity of a battery during a discharge cycle. It is typically expressed as the ratio of the discharged capacity to the total capacity of the battery. A charge and discharge cycle represents the process of charging a battery with electrical energy and then discharging it to utilize that energy. Each complete charge and discharge cycle contributes to the overall usage and ageing of the battery. The C-rate is a measure of the charging or discharging current relative to the battery's capacity. It indicates the rate at which energy is being charged into or discharged the battery. Specifically, a C-rate of 1 means that the current flowing into or out of the battery is equal to its capacity, resulting in a full charge or discharge in one hour. For instance, a battery with a capacity of 100Ah charging or discharging at a C-rate of 0.5 would have a current flow of 50A.

Accurately predicting battery degradation in real-life applications requires a good understanding of the non-linear dependencies among the various ageing factors. And in particular, the impact of each of these stress factors on the battery can vary depending on the specific lithium-ion battery chemistry and design. Various approaches have been proposed in the literature to capture the complex process involved in battery degradation. The published research generally distinguishes between different basic models. These models aim to provide insights into the factors that contribute to capacity loss.

An important approach is the physics-based modelling, which leverages fundamental electrochemical laws to simulate the complex mechanisms representing the most important state variables at any point in the battery cell and at any time. In particular, these models consider the underlying chemical reactions, material properties, and structural changes that occur during battery operation. As a result, such models have high accuracy, but at the same time have high computational complexity, which makes it infeasible to apply them in practice.

An alternative approach, known as empirical modelling, provides a practical means of characterizing battery degradation without delving into the underlying electrochemical processes. Empirical models utilize mathematical equations to describe the relationship between battery capacity and cycles. These models heavily rely on experimental data obtained from battery testing conducted under various conditions. By analysing this data, mathematical models can be developed based on the empirical correlation of the observed capacity and the ageing factors, such as temperature, depth of discharge, cycling conditions, and calendar ageing effects. Empirical models commonly employ polynomial

and exponential power laws, logarithmic, and trigonometric functions.

In recent years, data-driven modelling approaches, that apply machine learning based techniques, have gained popularity in battery degradation modelling. These methods utilize large datasets to train models that can capture complex patterns and correlations between operational parameters and degradation indicators. Data-driven models have shown promise in predicting battery ageing based on historical data, but are often limited by the large amount of experimental data that is required to build these models.

Event-oriented models are an important addition to battery degradation modelling approaches due to their simplicity. These models only focus on the cycling ageing of the battery and are mainly based on the number of cycles and the depth of discharge. An important example is the so-called rainflow-counting algorithm, which can be used to quantify cycles in the battery's state of charge profile.

Nowadays, semi-empirical models, which combine empirical and physics-based approaches, are widespread. This is because they have relatively high accuracy and do not require a large learning sample compared to data-driven models. And thus they represent a trade-off between computational complexity and accuracy.

2.2 A battery degradation model

The selection of an appropriate battery degradation model plays a crucial role in optimizing battery control. As discussed in the previous section, there exist various battery degradation models, but many of them tend to be overly complex for optimal control purposes. Since our objective is to formulate the optimal battery control problem as a Markov decision problem, it is essential to choose a model that aligns with the Markovian dynamics assumption. More precisely, at each point in time where we need to decide how to operate the battery, we would like to compute the degradation cost given a certain decision. In particular, this computation should not rely on the entire battery state of charge profile trajectory, but rather the degradation cost should be computable based on the current battery state and the performed operation. To simplify the degradation model, we will mainly focus on the cycling ageing effect.

It is worth noting that the degradation model based on the rainflow-counting algorithm typically strongly depends on the past trajectory of battery status. Even so, [KZ21] show that by introducing additional battery state variables, the method can be modified in such a way that it fits into the Markovian framework. However, the rainflow-counting algorithm mainly considers the depth of discharge and the number of cycles, but overlooks the influence of the C-rate on the degradation. In this section, we introduce a specific battery degradation model that addresses this limitation by leveraging the average C-rate and total (Ah)-throughput as additional battery states to model the battery degradation.

By focusing on the these two ageing factors, this model ensures computational tractability while still providing a more realistic representation of the battery degradation dynamics.

2.2.1 Battery variables

In this section, we will introduce important variables that represent aspects of the battery system to model the battery degradation. It is important to note that these variables are expressed in their physical units.

We consider a time interval $[0, T]$ with some finite time horizon $T > 0$ and we assume that the time is measured in hours h. We denote by Q_t the capacity of the battery, that is, the amount of energy that is stored in the battery at time t . And by Q_t^{\max} , we denote the maximum potential amount of energy that can be stored in the battery, which we refer to as the maximum capacity of the battery at time t . In particular, Q_0^{\max} denotes the initial maximum capacity of the battery. Of course, at any given time $t \in [0, T]$ it holds $0 \leq Q_t \leq Q_t^{\max}$. The capacity and the maximum capacity of the battery are measured in Ampere-hours Ah. We further denote by I_t the current at time t , which is measured in Ampere A. To simplify the notation, we assume that $I_t > 0$ means that the battery is being charged, whereas we interpret $I_t < 0$ as the battery being discharged, both with a current of $|I_t|$. Of course, the capacity Q_t at time t is influenced by the charge or discharge current I_t that is applied to the battery. More precisely, the dynamics of the battery capacity Q_t can be formulated by

$$\frac{d}{dt}Q_t = I_t, \quad \text{or equivalently,} \quad Q_t = Q_0 + \int_0^t I_s dt, \quad (2.1)$$

where $Q_0 \in [0, Q_0^{\max}]$ is the initial capacity of the battery. At this point, let us note that in our setting, we make the important simplifying assumption that the voltage stays constant throughout and does not depend on the state of charge or the charge and discharge current.

Recall that we would like to describe the battery degradation in terms of the Ah-throughput and the average C-rate. For this purpose, let us introduce the following battery variable

$$Q_t^{\text{tot}} := \int_0^t |I_s| ds \quad t \in [0, T],$$

which represents the total Ah-throughput up until time t , i.e. the total amount of electricity flown through the battery measured in Ah. Moreover, we consider the average

current up until time t defined by

$$I_t^{\text{av}} := \frac{Q_t^{\text{tot}}}{T_t^{\text{ac}}} = \frac{\int_0^t |I_s| \, ds}{T_t^{\text{ac}}}, \quad \text{where } T_t^{\text{ac}} := \int_0^t \mathbb{1}_{|I_s| > 0} \, ds.$$

We adopt the convention $\frac{0}{0} := 0$. Note that, T_t^{ac} describes the time in which the battery was actively charged or discharged up until time t , i.e. for the computation of the average current we ignore the times when the battery is neither charged nor discharged.

2.2.2 Dimensionless battery variables

We stress that the battery variables introduced in the previous section are expressed in their physical units. While these physical units provide a tangible representation of the quantities involved in battery dynamics, we will now transition to a more mathematically rigorous framework that employs dimensionless variables. A dimensionless variable is a quantity without any specific physical units and is often expressed relative to a reference value. By adopting dimensionless variables, we can focus on the underlying mathematical relationships without having to think about specific physical units. In particular, this abstraction enables us to easily apply the battery variables in the optimal battery control setting.

We consider the dimensionless time interval $[0, T]$, where $T := T/1 \text{ h}$ is the dimensionless equivalent to the terminal time T . To clarify the notation, we use the variable $t \in [0, T]$ to denote the dimensionless time, and we write $t \in [0, T]$ for the time measured in hours. Note that for every time $t \in [0, T]$ there exists a corresponding dimensionless time \mathbf{t} by the following relationship $\mathbf{t} = t/1 \text{ h}$ and vice versa. And hence, in order to draw a connection between the dimensionless variables and the battery variables introduced in the previous section, we will use the convention that whenever \mathbf{t} and t appear in the same equation, then t is given by $t = \mathbf{t} \cdot 1 \text{ h}$, unless otherwise stated.

Recall that the normalized capacity, i.e. the battery capacity relative to its maximum capacity, is referred to as the state of charge. More precisely, we define the state of charge w.r.t. Q_0^{\max} at time \mathbf{t} by $y_{\mathbf{t}}^{\text{SoC}} := Q_t/Q_0^{\max} \in [0, Q_t^{\max}/Q_0^{\max}]$ or alternatively the state of charge w.r.t Q_t^{\max} by $\tilde{y}_{\mathbf{t}}^{\text{SoC}} := Q_t/Q_t^{\max} \in [0, 1]$. In what follows, we will mainly focus on $y_{\mathbf{t}}^{\text{SoC}}$, i.e. the state of charge w.r.t. Q_0^{\max} . Remember that the C-rate is a way to quantify the rate at which a battery is charged or discharged relative to its maximum capacity. Consequently, we define the C-rate at time \mathbf{t} by

$$C_{\mathbf{t}} := I_t/I_C \quad \text{where} \quad I_C := \frac{Q_0^{\max}}{1 \text{ h}}.$$

Here, I_C is a normalization constant, which can be interpreted as the current required to

fully charge or discharge the battery in exactly one hour. For instance, if we are given a battery with 10 Ah maximum capacity and a 5 A discharge current, then the normalization constant I_C is 10 A and the C-rate is 0.5. Similarly to the charge/discharge current I_t , we interpret a positive C-rate as the battery being charged, while a negative C-rate implies a discharge of the battery. In particular, we can describe the dynamics of the state of charge with the C-rate, namely we have

$$\frac{d}{dt}y_t^{\text{SoC}} = C_t, \quad \text{or equivalently,} \quad y_t^{\text{SoC}} = y_0^{\text{SoC}} + \int_0^t C_s ds. \quad (2.2)$$

Moreover, we consider the normalized version of the Ah-throughput Q^{tot} , more precisely

$$y_t^{\text{tot}} := Q_t^{\text{tot}}/Q_0^{\text{max}} = \frac{1}{Q_0^{\text{max}}} \int_0^t |I_s| ds = \frac{1}{I_C \cdot 1 \text{ h}} \int_0^t |I_s| ds = \int_0^t |C_s| ds, \quad t \in [0, T]$$

and further, we can define the average C-rate C_t^{av} at time t by

$$C_t^{\text{av}} := \frac{\int_0^t |C_s| ds}{T_t^{\text{ac}}} = \frac{y_t^{\text{tot}}}{T_t^{\text{ac}}} \quad \text{where} \quad T_t^{\text{ac}} := \int_0^t \mathbb{1}_{|I_s| > 0} ds, \quad (2.3)$$

where T_t^{ac} describes the dimensionless time in which the battery was actively charged or discharged up until time t . Additionally, we refer to the normalized maximum capacity of the battery as the degradation state, defined by $\eta_t := Q_t^{\text{max}}/Q_0^{\text{max}}$. Of course, it is reasonable to assume that the degradation state η_t is monotonically decreasing, and hence it holds $\eta_t \in [0, 1]$. As a sidenote, we have $y_t^{\text{SoC}} = \eta_t \cdot \tilde{y}_t^{\text{SoC}}$.

In what follows, we assume that the initial maximum battery capacity Q_0^{max} is fixed as a battery parameter. Consequently, the task of modelling the battery degradation boils down to modelling the degradation state η_t .

2.2.3 Modelling the degradation state

From now on, we will only consider the dimensionless variables, which were introduced in the previous section, to model the degradation state. Thus, to improve the readability, we simply write t instead of \mathbf{t} for the dimensionless time.

To derive a model for the degradation state, we first look at the case of periodic cycling. Periodic cycling refers to a battery operation pattern where the battery is cycled between its maximum and minimum state of charge with a constant C-rate. This cycling pattern is used in many battery degradation experiments, and one can typically observe that the degradation can be modelled well by the number of charge cycles and the constant C-rate. In fact, in this controlled setting of periodic cycling, the degradation state can be approximated well by a function in the product of the number of charge cycles and the

constant C-rate. Accordingly, we define the following cycle counting process

$$N_t := \max \{n \in \mathbb{N}: y_t^{\text{tot}} \geq 2n\}, \quad (2.4)$$

that is, N_t counts the number of charge and discharge cycles completed up until time t . Notice that N_t is actually a stepwise constant function in t . Intuitively, however, it makes more sense to consider a continuous version of the cycle counting process. It is easy to see that this continuous version is simply given by $y_t^{\text{tot}}/2$. In other words, in the case of periodic cycling, we may consider the normalized Ah-throughput instead of the number of cycles to model the battery degradation. If we now assume that the battery is operated according to the periodic cycling pattern with constant C-rate C , then we can model the degradation state by

$$\eta_t = \hat{\eta} \left(\frac{y_t^{\text{tot}}}{2} \cdot C \right), \quad t \in [0, T], \quad (2.5)$$

where $\hat{\eta}: \mathbb{R}_{\geq 0} \rightarrow [0, 1]$ is some suitable degradation function.

Example 2.1. Clearly, the degradation function $\hat{\eta}: \mathbb{R}_{\geq 0} \rightarrow [0, 1]$ should be strictly decreasing. Let us consider two simple examples. Firstly, we can adopt a linear degradation function of the form $\hat{\eta}(x) = \max \{1 - a_{\text{deg}}x, 0\}$, where $a_{\text{deg}} > 0$ is some degradation parameter to be determined. A simple calculation shows that by choosing $a_{\text{deg}} = 2 \cdot 10^{-4}/C$, we can ensure that the battery's capacity is at 80% of its original after 10^3 charge cycles when charging with a constant C-rate of C . Another option is to model $\hat{\eta}$ as an exponential function, namely $\hat{\eta}(x) = e^{-a_{\text{deg}}x}$, where $a_{\text{deg}} > 0$. Since the linear degradation function provides the advantage of being additive, we will primarily focus on it in subsequent analysis.

Of course, it is unlikely that for the battery control problem the optimal state of charge profile will match the periodic cycling pattern. Hence, our goal is now to construct a simple equivalent model for the case of non-periodic cycling, i.e. an arbitrary battery control pattern. Ideally, this equivalent model should yield the same result upon insertion of periodic conditions. In the setting of non-periodic cycling, the C-rate is not guaranteed to be constant. Our approach is now to replace it with the average C-rate C_t^{av} as defined in (2.3). Moreover, we simply keep $y_t^{\text{tot}}/2$ as a factor to model the degradation state, even though it cannot be interpreted as the number of charge cycles any more. Thus, we obtain the following model for the degradation state

$$\eta_t = \hat{\eta} \left(\frac{y_t^{\text{tot}}}{2} \cdot C_t^{\text{av}} \right) = \hat{\eta} \left(\frac{(y_t^{\text{tot}})^2}{2T_t^{\text{ac}}} \right), \quad t \in [0, T], \quad (2.6)$$

where $\hat{\eta}$ is again some degradation function to be specified, see Example 2.1. Note that

the degradation state η_t modelled above essentially only depends on the past trajectory of the C-rate, i.e. $(C_s)_{s \in [0,t]}$. At first sight, the degradation state model (2.6) seems to be a reasonable generalization to the non-periodic setting. However, it turns out that this model is not quite realistic.

Remark 2.2. It can be shown that for the degradation state model (2.6) the ageing effect reverses under certain conditions, i.e. η_t is not monotonically decreasing in t . Intuitively, this can be explained as follows: The degradation state is a monotone decreasing function of the product of the normalized Ah-throughput and the average C-rate. Hence, for a sufficiently small C-rate C_t , the normalized total Ah-throughput y_t^{tot} increases but at the same time the average C-rate decreases C_t^{av} . If the effect of the decreasing average C-rate surpasses the effect of the increasing normalized Ah-throughput, then the degradation state increases. Indeed, this can be observed formally as follows. We assume that the degradation state is modelled by a linear degradation function, i.e. $\eta_t = 1 - a_{\text{deg}} \cdot \frac{y_t^{\text{tot}}}{2} C_t^{\text{av}}$, see Example 2.1. Then, the degradation dynamics are described by

$$\begin{aligned}\frac{\partial}{\partial t} \eta_t &= \frac{\partial}{\partial t} \left(1 - a_{\text{deg}} \cdot \frac{y_t^{\text{tot}}}{2} C_t^{\text{av}} \right) = -\frac{a_{\text{deg}}}{2} \frac{\partial}{\partial t} (y_t^{\text{tot}} \cdot C_t^{\text{av}}) \\ &= -\frac{a_{\text{deg}}}{2} \left(\frac{\partial}{\partial t} y_t^{\text{tot}} \cdot C_t^{\text{av}} + y_t^{\text{tot}} \cdot \frac{\partial}{\partial t} C_t^{\text{av}} \right), \quad t \in [0, T].\end{aligned}$$

In other words, the degradation dynamic can be separated into two penalization components. The first describes the degradation due to the change in the normalized Ah-throughput of the battery. Meanwhile, the second term models the effect of the change in the average C-rate on the degradation of the battery. Note that the normalized Ah-throughput is monotonically increasing, namely we have

$$\frac{\partial}{\partial t} y_t^{\text{tot}} = \frac{\partial}{\partial t} \int_0^t |C_s| \, ds = |C_t| \geq 0, \quad t \in [0, T].$$

This is true for T^{ac} as well, i.e.

$$\frac{\partial}{\partial t} T_t^{\text{ac}} = \frac{\partial}{\partial t} \int_0^t \mathbb{1}_{|C_s|>0} \, ds = \mathbb{1}_{|C_t|>0} \geq 0, \quad t \in [0, T].$$

In contrast, the change in the average C-rate can become negative. We have

$$\begin{aligned}\frac{\partial}{\partial t} C_t^{\text{av}} &= \frac{\partial}{\partial t} \left(\frac{y_t^{\text{tot}}}{T_t^{\text{ac}}} \right) = \frac{\frac{\partial}{\partial t} y_t^{\text{tot}} T_t^{\text{ac}} - y_t^{\text{tot}} \frac{\partial}{\partial t} T_t^{\text{ac}}}{(T_t^{\text{ac}})^2} \\ &= \frac{|C_t| T_t^{\text{ac}} - y_t^{\text{tot}} \mathbb{1}_{|C_t|>0}}{(T_t^{\text{ac}})^2} = \frac{1}{T_t^{\text{ac}}} (|C_t| - C_t^{\text{av}} \mathbb{1}_{|C_t|>0}), \quad t \in [0, T].\end{aligned}$$

Consequently, we obtain for degradation state

$$\begin{aligned}
\frac{\partial}{\partial t} \eta_t &= -\frac{a_{\text{deg}}}{2} \left(\frac{\partial}{\partial t} y_t^{\text{tot}} \cdot C_t^{\text{av}} + y_t^{\text{tot}} \cdot \frac{\partial}{\partial t} C_t^{\text{av}} \right) \\
&= -\frac{a_{\text{deg}}}{2} \left(|C_t| \cdot C_t^{\text{av}} + y_t^{\text{tot}} \cdot \frac{1}{T_t^{\text{ac}}} (|C_t| - C_t^{\text{av}} \mathbb{1}_{|C_t|>0}) \right) \\
&= -\frac{a_{\text{deg}}}{2} (|C_t| \cdot C_t^{\text{av}} + C_t^{\text{av}} (|C_t| - C_t^{\text{av}} \mathbb{1}_{|C_t|>0})) \\
&= -\frac{a_{\text{deg}}}{2} C_t^{\text{av}} (2|C_t| - C_t^{\text{av}} \mathbb{1}_{|C_t|>0}), \quad t \in [0, T].
\end{aligned}$$

We observe that $\frac{\partial}{\partial t} \eta_t$ becomes positive if $0 < |C_t| < C_t^{\text{av}}/2$ and hence, there exist C-rate trajectories such that η_t is not monotonically decreasing. In other words, the degradation reverses if the absolute C-rate $|C_t|$ is sufficiently small. In this case, the effect of the change in the average C-rate exceeds the effect of the added normalized Ah-throughput.

Of course, in reality, it is not possible that a battery recovers its maximum energy capacity by utilizing it in a certain way. As such, the proposed model is insufficient to capture the battery degradation effects realistically. Alternatively, a degradation state model which does not have the shortcomings of (2.6) is illustrated in the following example.

Example 2.3. We model degradation state as follows

$$\eta_t = \hat{\eta} \left(\frac{1}{2} \int_0^t |C_s|^2 ds \right) \quad (2.7)$$

where $\hat{\eta}$ is again some degradation function, see Example 2.1. Indeed, if we consider the setting of periodic cycling, that is we assume $|C_t| \equiv C$ is constant and periodic charge and discharge cycles. Then, we have

$$\frac{1}{2} \int_0^t |C_s|^2 ds = \frac{1}{2} \int_0^t |C_s| ds \cdot C = \frac{1}{2} y_t^{\text{tot}} \cdot C, \quad t \in [0, T],$$

i.e. this degradation state model is equivalent to (2.5) under periodic conditions. Taking a linear degradation function from Example 2.1 to model the degradation state in (2.7), we obtain

$$\frac{\partial}{\partial t} \eta_t = \frac{\partial}{\partial t} \left(1 - \frac{a_{\text{deg}}}{2} \int_0^t |C_s|^2 ds \right) = -\frac{a_{\text{deg}}}{2} |C_t|^2 \leq 0, \quad t \in [0, T],$$

i.e. η_t is monotonically decreasing. As a result, this model does not allow for a reversal of the degradation effect compared to (2.6).

While the above degradation state model does not have the same insufficiency as (2.6),

it seems to be too simplistic. To be more precise, the instantaneous degradation cost at any time t only depends on the C-rate $|C_t|$ at that time. In contrast, (2.6) also takes into account, the average C-rate C_t^{av} at time t which is a function of the entire history of the C-rate $(C_s)_{s \in [0, t]}$, see Remark 2.2. Nevertheless, the crucial observation here is that a degradation state should be a monotonically decreasing function, i.e. $\frac{\partial}{\partial t} \eta_t \leq 0$.

With this in mind, let us consider a different approach to model the degradation state. Namely, we choose to model the instantaneous degradation cost $\frac{\partial}{\partial t} \eta_t$ directly instead. More precisely, we model it as a function of the normalized Ah-throughput y_t^{tot} , the average C-rate C_t^{av} and the C-rate $|C_t|$ at time t , i.e.

$$\frac{\partial}{\partial t} \eta_t = \eta'(y_t^{\text{tot}}, C_t^{\text{av}}, |C_t|),$$

where $\eta' : \mathbb{R}_{\geq 0}^3 \rightarrow \mathbb{R}$ is some suitable function. Then, the degradation state at time t is given by

$$\eta_t = \eta_0 + \int_0^t \eta'(y_s^{\text{tot}}, C_s^{\text{av}}, |C_s|) ds = 1 + \int_0^t \eta'(y_s^{\text{tot}}, C_s^{\text{av}}, |C_s|) ds.$$

Indeed, this is a much more flexible approach since it encompasses various degradation state models by using different functions η' . For instance, by choosing

$$\eta'_{\mathbf{A1}}(y_t^{\text{tot}}, C_t^{\text{av}}, |C_t|) := -\frac{a_{\text{deg}}}{2} C_t^{\text{av}} (2|C_t| - C_t^{\text{av}} \mathbb{1}_{|C_t|>0}), \quad (2.8)$$

we obtain the initially proposed model (2.6), as shown in Remark 2.2, and in what follows we will refer to it as model **(A1)**. In comparison, by setting

$$\eta'_{\mathbf{B}}(y_t^{\text{tot}}, C_t^{\text{av}}, |C_t|) := -\frac{a_{\text{deg}}}{2} |C_t|^2,$$

we obtain the degradation state model from Example 2.3, which we refer to as model **(B)**.

Remark 2.4. Note that for any arbitrary C-rate $C_t \in \mathbb{R}$, and average C-rate $C_t^{\text{av}} \geq 0$, a simple calculation yields

$$\begin{aligned} |C_t|^2 - C_t^{\text{av}} (2|C_t| - C_t^{\text{av}} \mathbb{1}_{|C_t|>0}) &= |C_t|^2 - 2|C_t|C_t^{\text{av}} + (C_t^{\text{av}})^2 \mathbb{1}_{|C_t|>0} \\ &= |C_t|^2 - 2|C_t|C_t^{\text{av}} \mathbb{1}_{|C_t|>0} + (C_t^{\text{av}} \mathbb{1}_{|C_t|>0})^2 \\ &= (|C_t| - C_t^{\text{av}} \mathbb{1}_{|C_t|>0})^2 \geq 0, \end{aligned}$$

and in particular, this implies that $\eta'_{\mathbf{B}} \leq \eta'_{\mathbf{A1}}$. In other words, the degradation cost in model **(B)** is always higher compared to model **(A1)**.

Recall that, as seen in Remark 2.2, the major problem of the degradation model **(A1)** is the fact that $\frac{\partial}{\partial t}\eta_t = \eta'_{\mathbf{A}1}(y_t^{\text{tot}}, C_t^{\text{av}}, |C_t|) > 0$ if $0 < |C_t| < C_t^{\text{av}}/2$, i.e. the degradation reverses when the absolute C-rate $|C_t|$ is sufficiently small. A simple measure to solve this problem is to model the instantaneous degradation cost by

$$\frac{\partial}{\partial t}\eta_t = \eta'_{\mathbf{A}1}(y_t^{\text{tot}}, C_t^{\text{av}}, |C_t|) \wedge 0,$$

which would ensure $\frac{\partial}{\partial t}\eta_t \leq 0$. However, this model does not capture the impact of the current C-rate on the degradation realistically. In general, a larger C-rate $|C_t|$ suggests a stronger degradation. In other words, we would like $|C_t| \mapsto \eta'(y_t^{\text{tot}}, C_t^{\text{av}}, |C_t|)$ to be a strictly decreasing function.

With this in mind, we propose the following degradation state model

$$\eta'_{\mathbf{A}2}(y_t^{\text{tot}}, C_t^{\text{av}}, |C_t|) := \begin{cases} \eta'_{\mathbf{A}1}(y_t^{\text{tot}}, C_t^{\text{av}}, |C_t|) & \text{if } |C_t| > C_t^{\text{av}}, \\ \eta'_{\mathbf{B}}(y_t^{\text{tot}}, C_t^{\text{av}}, |C_t|) & \text{if } |C_t| \leq C_t^{\text{av}}, \end{cases}$$

which we refer to as model **(A2)**. Indeed, this model satisfies our desired properties.

Remark 2.5. (i) The degradation state in model **(A2)** is monotonically decreasing, i.e. $\frac{\partial}{\partial t}\eta_t \leq 0$. This is easily seen since, by Remark 2.2, $\eta'_{\mathbf{A}1}(y_t^{\text{tot}}, C_t^{\text{av}}, |C_t|) \leq 0$ if $|C_t| > C_t^{\text{av}} \geq C_t^{\text{av}}/2$. As a result, model **(A2)** does not allow for a reversal in the degradation.

(ii) Moreover, for fixed $y_t^{\text{tot}} > 0$ and $C_t^{\text{av}} > 0$, we have that $|C_t| \mapsto \eta'_{\mathbf{A}2}(y_t^{\text{tot}}, C_t^{\text{av}}, |C_t|)$ is a strictly decreasing function. Namely, we observe

$$\frac{\partial}{\partial |C_t|}\eta'_{\mathbf{B}}(y_t^{\text{tot}}, C_t^{\text{av}}, |C_t|) = \frac{\partial}{\partial |C_t|}\left(-\frac{a_{\text{deg}}}{2}|C_t|^2\right) = -a|C_t| < 0$$

for all $|C_t| > 0$ and in addition we have

$$\begin{aligned} \frac{\partial}{\partial |C_t|}\eta'_{\mathbf{A}1}(y_t^{\text{tot}}, C_t^{\text{av}}, |C_t|) &= \frac{\partial}{\partial |C_t|}\left(-\frac{a_{\text{deg}}}{2}(2|C_t| - C_t^{\text{av}})C_t^{\text{av}}\right) \\ &= -aC_t^{\text{av}} < 0, \end{aligned}$$

which implies that $|C_t| \mapsto \eta'_{\mathbf{A}2}(y_t^{\text{tot}}, C_t^{\text{av}}, |C_t|)$ must be a strictly decreasing function.

(iii) In fact, we can even show that $|C_t| \mapsto \eta'_{\mathbf{A}2}(y_t^{\text{tot}}, C_t^{\text{av}}, |C_t|)$ is a continuously differentiable function for fixed y_t^{tot} and C_t^{av} . It is easy to check that $|C_t| \mapsto \eta'_{\mathbf{A}1}(y_t^{\text{tot}}, C_t^{\text{av}}, |C_t|)$, and $|C_t| \mapsto \eta'_{\mathbf{B}}(y_t^{\text{tot}}, C_t^{\text{av}}, |C_t|)$ are continuously differentiable. Hence, we are left with showing that $|C_t| \mapsto \eta'_{\mathbf{A}2}(y_t^{\text{tot}}, C_t^{\text{av}}, |C_t|)$ is continuously differentiable in $|C_t| = C_t^{\text{av}}$,

i.e. when the absolute C-rate $|C_t|$ is equal to the average C-rate C_t^{av} . Indeed, we have

$$\begin{aligned}\eta'_{\mathbf{A}2}(y_t^{\text{tot}}, C_t^{\text{av}}, |C_t|) &= -\frac{a_{\text{deg}}}{2}(2|C_t| - C_t^{\text{av}} \mathbb{1}_{|C_t|>0})C_t^{\text{av}} \\ &= -\frac{a_{\text{deg}}}{2}(2|C_t| - |C_t|\mathbb{1}_{|C_t|>0})|C_t| \\ &= -\frac{a_{\text{deg}}}{2}|C_t|^2 = \eta'_{\mathbf{B}}(y_t^{\text{tot}}, C_t^{\text{av}}, |C_t|).\end{aligned}$$

Moreover, as shown in (ii), we have

$$\frac{\partial}{\partial|C_t|}\eta'_{\mathbf{A}2}(y_t^{\text{tot}}, C_t^{\text{av}}, |C_t|) = -aC_t^{\text{av}} = -a|C_t| = \frac{\partial}{\partial|C_t|}\eta'_{\mathbf{B}}(y_t^{\text{tot}}, C_t^{\text{av}}, |C_t|).$$

Hence, $|C_t| \mapsto \eta'_{\mathbf{A}2}(y_t^{\text{tot}}, C_t^{\text{av}}, |C_t|)$ is continuously differentiable.

Note that the models **(A1)**, **(A2)** and **(B)** all implicitly model the linear degradation function, see Example 2.1. That is, our models aim to replicate the degradation in the periodic setting (2.5) using a linear degradation function. Of course, we also could have chosen an exponential degradation function as an alternative. However, the linearity allowed a simpler analysis, which is why we chose the linear degradation function as modelling basis instead.

2.2.4 Modelling the degradation state in discrete-time

Recall that our objective is the optimal control of a battery storage system, mainly by varying the C-rate. However, finding an optimal strategy $(C_t)_{t \in [0,T]}$ in continuous time proves challenging due to its infinite-dimensional nature. To address this, Chapter 4 takes a discretized approach and formulates the problem as a Markov decision problem. Consequently, the goal of this section is to introduce the discrete time equivalent of the battery degradation models discussed in the previous section.

For this purpose, consider the following time discretization $\{t_k = \frac{k}{N}T \mid k = 0, \dots, N\}$ for some $N \in \mathbb{N}$ and $\Delta t := T/N$. For convenience, we denote by y_k^{SoC} the state of charge at time t_k , i.e. $y_k^{\text{SoC}} := y_{t_k}^{\text{SoC}}$ and analogously for any other time-dependent battery variable. As we would like the discretized setting to coincide with the continuous setting, we assume that the C-rate is step-wise constant. More precisely, C_k denotes the C-rate at which the battery is charged or discharged during the time interval $[t_k, t_{k+1})$. Let y_0^{SoC} be the initial state of charge level of the battery at time 0. Then, compared to (2.2), the state of charge y_{k+1}^{SoC} at time t_{k+1} is recursively given by

$$y_{k+1}^{\text{SoC}} = y_k^{\text{SoC}} + C_k \Delta t \tag{2.9}$$

for all $k = 0, \dots, N - 1$. Moreover, we have $y_0^{\text{tot}} = T_0^{\text{ac}} = 0$ and

$$y_{k+1}^{\text{tot}} = y_k^{\text{tot}} + |C_k| \Delta t, \quad \text{and} \quad T_{k+1}^{\text{ac}} = T_k^{\text{ac}} + \mathbb{1}_{|C_k| > 0} \Delta t$$

for all $k = 0, \dots, N - 1$, and the average C-rate is given by $C_k^{\text{av}} = y_k^{\text{tot}} / T_k^{\text{ac}}$ for $k = 0, \dots, N$. Note that a small computation yields

$$\begin{aligned} C_{k+1}^{\text{av}} - C_k^{\text{av}} &= \frac{y_k^{\text{tot}} + |C_k| \Delta t}{T_k^{\text{ac}} + \mathbb{1}_{|C_k| > 0} \Delta t} - C_k^{\text{av}} \\ &= \frac{y_k^{\text{tot}} + |C_k| \Delta t - C_k^{\text{av}} (T_k^{\text{ac}} + \mathbb{1}_{|C_k| > 0} \Delta t)}{T_k^{\text{ac}} + \mathbb{1}_{|C_k| > 0} \Delta t} \\ &= \frac{|C_k| \Delta t - C_k^{\text{av}} \mathbb{1}_{|C_k| > 0} \Delta t}{T_k^{\text{ac}} + \mathbb{1}_{|C_k| > 0} \Delta t} \end{aligned} \tag{2.10}$$

and hence we can also define the average C-rate recursively by

$$C_{k+1}^{\text{av}} = C_k^{\text{av}} + \frac{|C_k| \Delta t - C_k^{\text{av}} \mathbb{1}_{|C_k| > 0} \Delta t}{T_k^{\text{ac}} + \mathbb{1}_{|C_k| > 0} \Delta t}, \quad k = 0, \dots, N.$$

In the preceding section, we described the degradation state η_t by directly modelling the instantaneous degradations cost $\frac{\partial}{\partial t} \eta_t$. Here, in the discretized setting, it makes sense to model the degradation cost for each time step instead, namely

$$\Delta_k \eta := \eta_{k+1} - \eta_k = \int_{t_k}^{t_{k+1}} \eta'(y_s^{\text{tot}}, C_s^{\text{av}}, |C_s|) \, ds$$

for every $k = 0, \dots, N - 1$. Similarly to the continuous case, we model the degradation cost by some suitable function $\tilde{\eta}: \mathbb{R}_{\geq 0}^3 \rightarrow \mathbb{R}$ taking y_k^{tot} , C_k^{av} and $|C_k|$ as an input, i.e.

$$\Delta_k \eta = \tilde{\eta}(y_k^{\text{tot}}, C_k^{\text{av}}, |C_k|), \quad k = 0, \dots, N - 1.$$

Our main goal is to replicate model **(A2)** in this discretized setting. We first observe for model **(B)** that

$$\begin{aligned} \Delta_k \eta &= \int_{t_k}^{t_{k+1}} \eta'_{\mathbf{B}}(y_s^{\text{tot}}, C_s^{\text{av}}, |C_s|) \, ds = \int_{t_k}^{t_{k+1}} -\frac{a_{\text{deg}}}{2} |C_s|^2 \, ds \\ &= -\frac{a_{\text{deg}}}{2} |C_k|^2 \Delta t =: \tilde{\eta}_{\mathbf{B}}(y_k^{\text{tot}}, C_k^{\text{av}}, |C_k|) \end{aligned}$$

for every $k = 0, \dots, N - 1$. In addition, we recall that for model **(A1)**, we have $\eta_t =$

$1 - a_{\text{deg}} \cdot \frac{y_t^{\text{tot}}}{2} C_t^{\text{av}}$ and thus, we obtain

$$\begin{aligned}
\Delta_k \eta &= \eta_{k+1} - \eta_k = \left(1 - a_{\text{deg}} \cdot \frac{y_{k+1}^{\text{tot}}}{2} C_{k+1}^{\text{av}} \right) - \left(1 - a_{\text{deg}} \cdot \frac{y_k^{\text{tot}}}{2} C_k^{\text{av}} \right) \\
&= -\frac{a_{\text{deg}}}{2} (y_{k+1}^{\text{tot}} \cdot C_{k+1}^{\text{av}} - y_k^{\text{tot}} \cdot C_k^{\text{av}}) \\
&= -\frac{a_{\text{deg}}}{2} ((y_{k+1}^{\text{tot}} - y_k^{\text{tot}}) \cdot C_{k+1}^{\text{av}} + y_k^{\text{tot}} \cdot (C_{k+1}^{\text{av}} - C_k^{\text{av}})) \\
&= -\frac{a_{\text{deg}}}{2} \left(|C_k| \Delta t \cdot \frac{y_k^{\text{tot}} + |C_k| \Delta t}{T_k^{\text{ac}} + \mathbb{1}_{|C_k| > 0} \Delta t} + y_k^{\text{tot}} \cdot \frac{|C_k| \Delta t - C_k^{\text{av}} \mathbb{1}_{|C_k| > 0} \Delta t}{T_k^{\text{ac}} + \mathbb{1}_{|C_k| > 0} \Delta t} \right) \\
&= -\frac{a_{\text{deg}}}{2} \left(\frac{(2|C_k| - C_k^{\text{av}} \mathbb{1}_{|C_k| > 0}) y_k^{\text{tot}} \Delta t + |C_k|^2 \Delta t^2}{y_k^{\text{tot}} / C_k^{\text{av}} + \Delta t} \right) \\
&=: \tilde{\eta}_{\mathbf{A1}}(y_k^{\text{tot}}, C_k^{\text{av}}, C_k), \quad k = 0, \dots, N-1,
\end{aligned}$$

where we used (2.10) in the fourth equation. It is now easy to see that

$$\tilde{\eta}_{\mathbf{A2}}(y_k^{\text{tot}}, C_k^{\text{av}}, |C_k|) := \begin{cases} \tilde{\eta}_{\mathbf{A1}}(y_k^{\text{tot}}, C_k^{\text{av}}, |C_k|) & \text{if } |C_k| > C_k^{\text{av}}, \\ \tilde{\eta}_{\mathbf{B}}(y_k^{\text{tot}}, C_k^{\text{av}}, |C_k|) & \text{if } |C_k| \leq C_k^{\text{av}}, \end{cases}$$

reproduces model **(A2)** in the discretized setting. And similarly to continuous case, a straightforward calculation shows that $|C_k| \mapsto \tilde{\eta}_{\mathbf{A2}}(y_k^{\text{tot}}, C_k^{\text{av}}, |C_k|)$ is strictly decreasing and continuously differentiable.

3 Stochastic modelling of electricity markets

Starting in the early 1990s, the deregulation of electricity markets in many countries was a significant development that revolutionized the way electricity was bought and sold. This market liberalization aimed to introduce competition, opening up the supply of electricity to a broader range of market participants. Notably, the United Kingdom underwent a radical reform of its electricity supply industry, establishing a competitive market structure and empowering consumers with the ability to switch suppliers. Since then, this approach of opening up the markets has affected most developed countries in the world. This shift marked a fundamental change in the dynamics of electricity markets, as market participants were now faced with the challenges of forecasting and managing the uncertainties inherent in electricity supply and demand, as well as navigating the risk and complexities of competing in an open market environment. As a result, the demand for mathematical models capable of capturing the dynamic nature of electricity markets has become increasingly important.

In order to model electricity markets, it is important to understand how electricity is traded. The typical products traded on energy markets are spot prices and long-term contracts such as forward and futures contracts, which usually have a delivery period varying from a week up to a year. In what follows, we mainly focus on electricity spot prices. However, since electricity is not efficiently storable there is no actual spot price, but only short-term contracts. These contracts are traded on day-ahead markets, which involve delivery of electricity on the next day, and on intraday markets that involve delivery in 15 or 30 minutes after the transaction. Let us briefly introduce how these kinds of spot markets are structured. As an example, we look at a major central European energy exchange, the European Energy Exchange (EEX) located in Leipzig. The spot market of EEX, also called EPEX spot market, trades contracts for the physical delivery of electricity in Austria, Germany, France, or Switzerland. There are two types of markets the EPEX operates, namely the day-ahead market and the intraday market. The day-ahead market is organized as a daily auction which takes place at 12:00 pm, 7 days a week all year. The underlying quantity to be traded is the electricity for delivery the following day in 24 one hour intervals. After the close of the day-ahead auction, market participants can trade electricity on the intraday market at very short notice, up to 5 minutes before delivery, for time periods between a quarter of an hour and blocks of one hour. This allows for markets participants to adjust from hourly positions to finer granularities. In particular, the intraday market empowers market participants to promptly correct their positions in response to more accurate estimates of electricity generation and consumption. Allowing these real-time adjustments significantly enhances the overall accuracy of supply and demand balancing in the electricity market, improving the efficiency and reliability

of the entire system.

The rest of this chapter is organized as follows. First, we introduce several stochastic models for electricity prices. We then deal with the simulation of the presented electricity price models. Finally, we investigate how weather forecasts such as wind can be incorporated into these models.

3.1 Modelling electricity prices

Before we can even think of modelling electricity prices, it is essential that we are aware of the specific characteristics that electricity prices exhibit. After all, electricity is a quite unique commodity, in the sense that it is grid-based and very difficult to store because most technologies for the large-scale storage of electricity are still quite expensive and limited. Therefore, the electricity grid has to stay in balance at all times, that is, as soon as electricity is produced, it has to be delivered and consumed by the end consumers. This leads to certain stylized facts which can be consistently observed in the time series of electricity prices.

- ▷ *Seasonality*: One prominent stylized fact of electricity prices is seasonality. This is attributable to traditional seasonal influences, such as weather conditions, economic activity, and specific events like holidays or industrial shutdowns, that impact electricity demand and prices accordingly. Moreover, renewable energy sources, such as wind and solar power, exhibit seasonal generation patterns, leading to fluctuations in electricity supply and subsequent price variations. In addition, day-to-day changes in electricity demand, driven by factors like industrial activity and residential consumption patterns, also contribute to seasonal price dynamics.
- ▷ *Extreme variability in prices*: Unlike other commodities, electricity cannot be easily stored in large quantities. Therefore, the supply and demand balance must be continuously matched in real-time. Small changes in supply or demand can lead to significant price fluctuations, as the system operator must maintain a delicate equilibrium between generation and consumption. As a result, this leads to extreme variability in prices.
- ▷ *Price spikes*: These are substantial up- or downwards movements in electricity prices followed by rapid reversion to previous levels that occur sporadically in the market. These spikes can be caused by a variety of factors, including unexpected changes in supply or demand, e.g. due to extreme weather conditions, generation or transmission failures, or market manipulation.

- ▷ *Mean-reversion*: This refers to the tendency of prices to revert to a long-term average or equilibrium level over time. This mean reversion can be attributed to various factors, such as the presence of arbitrage opportunities that incentivize market participants to exploit deviations from the mean, economic forces that drive prices towards cost-based levels, and regulatory mechanisms designed to promote price stability and prevent market manipulation.
- ▷ *Roughness/antipersistence*: Another important stylized fact observed in electricity price time series is roughness or antipersistence. Specifically, the so-called roughness index of electricity spot price time series is negative, i.e. the paths of the spot prices are very rough as compared to a model based on the Brownian motion. Intuitively, this means that consecutive price changes tend to alternate between increases and decreases, rather than following a persistent trend. For more details, we refer to [Ben17].

In view of these stylized facts, it is clear that a good electricity price model should be able to reproduce these main characteristics. For a review of electricity spot (and future) price models, we refer to [DFG21].

The electricity price models that we consider here will be defined in continuous time. At first sight, a continuous-time model seems inappropriate, since electricity prices, as described previously, are specified for time intervals where the electricity is delivered. Indeed, the prices are determined as a set of hourly prices in the day-ahead market, and as half-hourly, or even quarter-hourly prices in the intraday markets. Hence, electricity prices are inherently given on a time-discretized grid. Continuous-time models, however, are prevalent in the literature due to their practicality, as a significant amount of research is devoted to deriving forward prices from the dynamics of spot prices. After all, forward prices evolve in a continuous-time market, and not at discrete hourly times. In particular, one may view the continuous price dynamics as an unobserved price for immediate delivery of electricity, and the actual hourly or quarterly prices are simply observations of this. For a deeper discussion on this, we refer to Remark 3.15. As such, in what follows, we proceed by modelling electricity spot prices by continuous time models.

For this purpose, we denote the electricity spot price at time $t \geq 0$ by S_t and throughout this chapter, we assume a given underlying complete probability space $(\Omega, \mathcal{F}, \mathbb{P})$. In addition, we fix a filtration $\mathbb{F} = (\mathcal{F}_t)_{t \geq 0}$ which satisfies the usual conditions, i.e. complete and right-continuous.

3.1.1 The Schwartz model

It is well known that the basic model for stock prices is the famous Black-Scholes model whose dynamics are described by a geometric Brownian motion, i.e.

$$dS_t = \mu S_t dt + \sigma S_t dB_t, \quad t \geq 0,$$

where $\mu \in \mathbb{R}$, $\sigma > 0$ and $(B_t)_{t \geq 0}$ is a one-dimensional Brownian motion. The constant μ is referred to as the *drift* and σ as the *volatility*.

Similarly, the canonical model in energy markets follows the so-called Schwartz model, which was first proposed by [Sch97]. The spot price is given by the stochastic process

$$S_t = \Lambda_t + Z_t, \quad t \geq 0, \tag{3.1}$$

where Λ_t is a positive deterministic function modelling the seasonality, and Z_t follows the Ornstein-Uhlenbeck process, i.e. the dynamics are given by

$$dZ_t = -\theta Z_t dt + \sigma dB_t, \quad t \geq 0. \tag{3.2}$$

Here, θ and σ are positive constants where θ specifies the *rate of reversion* and σ the *volatility*. The Ornstein-Uhlenbeck process is mean reverting in the sense that the process tends to drift towards 0. In particular, the electricity price S_t reverts towards its mean Λ_t with a rate θ . Note that (3.2) can be solved explicitly by applying Itô's formula which yields

$$Z_t = Z_0 e^{-\theta t} + \int_0^t e^{-\theta(t-s)} \sigma dB_s.$$

Interestingly, the Ornstein-Uhlenbeck process is both a Gaussian process and a Markov process.

Remark 3.1. When modelling the spot price, one typically considers two different modelling choices, the arithmetic approach and the geometric approach. In the arithmetic case, the goal is to model the dynamics of the spot price directly. Meanwhile, in the geometric case one aims to model the logarithm of the spot price, i.e. $\ln S_t$. Hence, if we use the Schwartz dynamics (3.1) to model the logarithm of the spot price for the geometric case, we obtain a model of the following form

$$\ln S_t = \tilde{\Lambda}_t + Z_t, \quad t \geq 0,$$

or equivalently,

$$S_t = \Lambda_t \exp(Z_t), \quad t \geq 0, \quad (3.3)$$

where $\Lambda_t = \exp \tilde{\Lambda}_t$ is a positive deterministic function modelling the seasonality factor, and Z_t an Ornstein-Uhlenbeck process as in (3.2). Note that the Schwartz model in the geometric case (3.3) is multiplicative, while the Schwartz model in the arithmetic case (3.1) is additive. While both additive and multiplicative models are widely used in the modelling of energy markets, we will focus on the arithmetic case in the following sections. This choice is motivated by the ability of arithmetic models to capture the occurrence of negative prices, which are increasingly observed in electricity markets.

The Schwartz model can be generalized to include multiple factors, instead of a single one, as studied in [SS00]. In this case, the extended model takes the following form

$$S_t = \Lambda_t + \sum_{i=1}^d Z_t^i, \quad \text{with} \quad Z_t^i = -\theta_i Z_t^i dt + dB_t^i, \quad \text{for } i = 1, \dots, d, \quad (3.4)$$

where θ_i is the corresponding rate of reversion for the Ornstein-Uhlenbeck process Z^i , $i = 1, \dots, d$, and $(B^i)_{i=1,\dots,d}$ are correlated Brownian motions.

Although the Schwartz model has quite simple dynamics, it possesses two major advantages, namely the ability to price derivatives explicitly and the existence of efficient calibration methods on market data. However, due to its simplicity, the model is limited in its ability to represent the specific characteristics of electricity prices. Namely, the Schwartz model (3.1) is driven by a Brownian motion, and hence unlikely to create the large price spikes observed in electricity markets. This problem will be the focus of the next section.

3.1.2 Modelling price spikes

Price spikes are characterized by large up and downward movements followed by a very rapid return to previous levels, and are a consequence of the difficulty of storing electricity. There are both upward and downward price spikes, especially for regions with a high proportion of renewable generation.

A natural way to model these kinds of price spikes, is to replace the Brownian motion $(B_t)_{t \geq 0}$ in (3.2) by a more general Lévy process $(L_t)_{t \geq 0}$. For the reader's convenience, let us quickly recall some basic properties of Lévy processes. For more details, we refer to [TC15] and [Pro13].

Definition 3.2. An \mathbb{F} -adapted stochastic process $(L_t)_{t \geq 0}$ taking values in \mathbb{R} with $L_0 = 0$ almost surely is called a Lévy process if it satisfies the following properties

- (i) L has independent increments, i.e. $L_t - L_s$ is independent of \mathcal{F}_s , $0 \leq s \leq t$,
- (ii) L has stationary increments, i.e. $L_t - L_s$ has the same law as L_{t-s} , $0 \leq s \leq t$,
- (iii) L is continuous in probability, i.e. $\lim_{h \rightarrow 0} \mathbb{P}(|L_{t+h} - L_t| \geq \varepsilon) = 0$ for all $\varepsilon > 0$.

Of course, it is immediately clear that a Brownian motion is a continuous Lévy process. In fact, it turns out that if a Lévy process L is continuous almost surely, then it is a Brownian motion with drift. In general, it can be shown that any Lévy process has a unique càdlàg modification which is also a Lévy process, see [Pro13, Theorem 30].

Example 3.3 (Compound Poisson process). Let $(N_t)_{t \geq 0}$ denote a Poisson process with intensity $\lambda > 0$, and let $(J_i)_{i=1}^\infty$ be a sequence of i.i.d. random variables independent of the Poisson process $(N_t)_{t \geq 0}$. Then, the stochastic process defined by $M_t = \sum_{i=1}^{N_t} J_i$, $t \geq 0$, is called a compound Poisson process. It is easy to see that the sample paths of $(M_t)_{t \geq 0}$ are càdlàg piecewise constant functions. In fact, it can be proven that a stochastic process is a compound Poisson process if and only if it is a Lévy process with sample paths that are piecewise constant functions. For a proof, we refer to [TC15, Proposition 3.3].

An important property of Lévy processes is that for any $t \geq 0$ the law L_t is infinitely divisible.

Definition 3.4. A probability distribution F on \mathbb{R} is said to be infinitely divisible, if for any $n \in \mathbb{N}$, there exist independent and identically distributed random variables Y_1, \dots, Y_n such that $Y_1 + \dots + Y_n$ has distribution F .

Indeed, for any $t \geq 0$ and $n \in \mathbb{N}$, we have that $L_t = \sum_{k=0}^{n-1} L_{(k+1)t/n} - L_{kt/n}$ and, by Definition 3.2, $(L_{(k+1)t/n} - L_{kt/n})_{k=0, \dots, n-1}$ are n independent and identically distributed random variables. Thus, the law of L_t is infinitely divisible. In particular, the fact that Brownian motion and the Poisson process are Lévy processes implies that the normal distribution and the Poisson distribution are infinitely divisible. Interestingly, it can be shown that any infinitely divisible distribution can be used to construct a Lévy process.

Proposition 3.5 ([TC15, Proposition 3.1]). *Let $(L_t)_{t \geq 0}$ be a Lévy process. Then, for every $t \geq 0$, L_t has an infinitely divisible distribution. Conversely, if F is an infinitely divisible distribution, then there exists a Lévy process $(L_t)_{t \geq 0}$ such that the distribution of L_1 is given by F .*

As a direct consequence of Proposition 3.5, we have that the distribution of $(L_t)_{t \geq 0}$ is uniquely determined by the distribution of L_1 .

The Lévy-Itô decomposition is another crucial result in the theory of Lévy processes

which gives an intuition on the structure of a Lévy process. Namely, it turns out that any Lévy process can be decomposed into three components. Recall that we have seen in Example 3.3 that Lévy processes actually can have jumps. It is well known that for every Poisson process, only finitely many jumps occur in every bounded interval with probability one, and hence this is true for every compound Poisson process as well. We say that a Lévy process which satisfies this condition has finite activity. And in the contrary case, we say that it has infinite activity. The Lévy-Itô decomposition states that a Lévy process can be decomposed into a diffusion, a process of finite activity, and a process of infinite activity.

Theorem 3.6 (Lévy-Itô decomposition). *Let $(L_t)_{t \geq 0}$ be a Lévy process. Then, there exist three independent Lévy processes $(B_t)_{t \geq 0}$, $(M_t)_{t \geq 0}$ and $(Y_t)_{t \geq 0}$ with*

$$L_t = B_t + M_t + Y_t, \quad t \geq 0,$$

such that $(B_t)_{t \geq 0}$ is a Brownian motion with drift, $(M_t)_{t \geq 0}$ is a compound Poisson process (the finite activity part), and $(Y_t)_{t \geq 0}$ is a pure jump martingale, with jumps bounded by some fixed number $\varepsilon > 0$ (the infinite activity part).

Proof. For a proof, we refer to [App09]. □

An immediate consequence of the Lévy-Itô decomposition is that every Lévy process is a combination of a Brownian motion with drift and a possibly infinite sum of independent compound Poisson processes. And in particular, this means that every Lévy process can be approximated by a superposition of a Brownian motion with drift and an independent compound Poisson processes.

We now turn our attention to an important class of Lévy processes which we will use to model electricity prices. Recall that by Proposition 3.5 any infinitely divisible distribution induces a Lévy process. One very general class of infinitely divisible distributions is the so-called generalized hyperbolic distribution. Due to its many desirable properties such as fat tails and skewness, it is especially suited for our application of modelling electricity prices. In fact, many well-known probability distributions are generalized by the generalized hyperbolic distribution, such as the hyperbolic distribution, the Student's t -distribution, the variance-gamma distribution and the normal inverse Gaussian distribution.

The generalized hyperbolic distribution is obtained as a normal variance-mean mixture, where the mixing distribution is the generalized inverse Gaussian distribution.

Definition 3.7 (Generalized inverse Gaussian distribution). Let $(\lambda, \chi, \psi) \in \mathbb{R}^3$ be some

parameters that satisfy one of the following three restrictions

$$\chi > 0, \psi \geq 0, \lambda < 0, \quad \text{or} \quad \chi > 0, \psi > 0, \lambda = 0, \quad \text{or} \quad \chi \geq 0, \psi > 0, \lambda > 0.$$

Then, the density of the generalized inverse Gaussian distribution with parameters (λ, χ, ψ) is defined by

$$f_{GIG}(x) = \left(\frac{\psi}{\chi} \right)^{\frac{\lambda}{2}} \frac{x^{\lambda-1}}{2K_\lambda(\sqrt{\chi\psi})} \exp \left(-\frac{1}{2} \left(\frac{\chi}{x} + \psi x \right) \right),$$

where K_λ denotes the modified Bessel function of the third kind.

Definition 3.8 (Generalized hyperbolic distribution). Let X be a random variable. We say that the law of X is given by a (univariate) generalized hyperbolic distribution if

$$X \stackrel{\text{law}}{=} \mu + W\gamma + \sqrt{W}\sigma Z,$$

where $Z \sim N(0, 1)$, $\sigma, \mu, \gamma \in \mathbb{R}$ and W is a random variable, independent of Z , with generalized inverse Gaussian distribution and parameters (λ, χ, ψ) , i.e. $W \sim GIG(\lambda, \chi, \psi)$.

Remark 3.9. In Definition 3.8 above, we refer to μ as the location parameter, $\Sigma = \sigma^2$ as the dispersion matrix and γ as the symmetry or skewness parameter. If $\gamma = 0$, then this means that the distribution is symmetric. The three parameters (λ, χ, ψ) of the generalized inverse Gaussian distribution determine the shape of the generalized hyperbolic distribution. The parametrization described here is referred to as the so-called $(\lambda, \chi, \psi, \mu, \Sigma, \gamma)$ -parametrization of the generalized hyperbolic distribution.

As noted above, the generalized hyperbolic distribution generalized many other well-known distributions, namely

- for $\lambda = 1$, we obtain the hyperbolic distribution,
- for $\lambda = -1/2$, we obtain the normal inverse Gaussian distribution,
- for $\lambda > 0$ and $\chi = 0$, we obtain the variance gamma distribution,
- for $\lambda < 0$ and $\psi = 0$, we obtain the (generalized hyperbolic) Students's t-distribution.

For more details, we refer to [WBL22].

To summarize, Lévy processes are an important class of stochastic processes since they naturally include big jumps as well as many small jumps, and can even include a Brownian motion. As mentioned in the beginning of the section, we can generalize the

(extended) Schwartz model by replacing the Brownian motion in the Ornstein-Uhlenbeck process (3.2) with a Lévy process, i.e.

$$dZ_t = -\theta Z_t dt + \sigma dL_t, \quad t \geq 0. \quad (3.5)$$

This process is often referred to as a Lévy-driven Ornstein-Uhlenbeck process and the Levy process $(L_t)_{t \geq 0}$ is often called a background-driving Lévy process, or BDLP for short. Similarly to before, by applying Itô's formula it can be proven that (3.5) is solved by

$$Z_t = Z_0 e^{-\theta t} + \int_0^t e^{-\theta(t-s)} dL_s.$$

For a deeper discussion on Lévy-driven Ornstein-Uhlenbeck processes, we refer the reader to [TC15, Chapter 15.3]. Finally, by generalizing (3.4) we obtain a model with dynamics given by

$$S_t = \Lambda_t + \sum_{i=1}^d Z_t^i, \quad \text{where} \quad Z_t^i = -\theta_i Z_t^i dt + dL_t^i, \quad \text{for } i = 1, \dots, d, \quad (3.6)$$

where $(L_t^i)_{t \geq 0}$, $i = 1, \dots, d$ are Lévy processes. Note that the mean reversion property of the Ornstein-Uhlenbeck process ensures that a large jump in the price is followed by an exponentially fast decrease in prices.

Example 3.10. To give an example, [MT08] studies a two-factor model of the form (3.6) with $d = 2$, and

$$dZ_t^1 = -\theta_1 Z_t^1 dt + \sigma dB_t \quad \text{and} \quad dZ_t^2 = -\theta_2 Z_t^2 dt + dM_t,$$

for $t \geq 0$, where $(M_t)_{t \geq 0}$ is a compound Poisson process, see Example 3.3. In particular, [MT08] proposes a method to filter out the different factors of the model, as well as statistical procedure for estimating the parameters of the model.

3.1.3 Stochastic volatility models

Another natural way to extend the Schwartz model (3.4) is to consider stochastic dynamics for the volatility parameter. In classical mathematical finance models, stochastic volatility is introduced because it allows to better characterize option prices observed in the market, such as the volatility smile. An additional advantage is that stochastic volatility leads to asymmetry in the price distribution as well as fat tails, and thus we can apply it to improve our electricity price models. Adding stochastic volatility to the one-factor Schwartz-model

yields

$$\begin{aligned} S_t &= \Lambda(t) + Z_t, \\ dZ_t &= (\mu - \theta X_t) dt + \sigma_t dB_t, \end{aligned} \tag{3.7}$$

for $t \geq 0$. There are a lot of options in modelling the stochastic volatility. For example, in the classical Heston model the volatility is modelled as a CIR process. For the purpose of modelling energy prices, [Ben13] proposes the so-called Barndorff-Nielsen and Shephard stochastic volatility model, which was first introduced in [BS01].

Example 3.11. In the Barndorff-Nielsen and Shephard stochastic volatility model the square of the volatility is represented by a multifactor model

$$\sigma_t^2 = \sum_{i=1}^I w_i v_t^{(i)} \quad \text{with} \quad dv_t^{(i)} = -\kappa_i v_t^{(i)} dt + dU_t^i, \quad t \geq 0, \tag{3.8}$$

where $\kappa_i > 0$ are constants and $w_i > 0$ weights such that $\sum_{i=1}^I w_i = 1$, and U^i are independent subordinators, i.e. Lévy process that are non-negative. Notice that $v^{(i)}$ is in fact an Ornstein-Uhlenbeck process driven by the subordinator $(U_t^{(i)})_{t \geq 0}$. In particular, since U_t^i is non-negative, it follows that $v_t^{(i)}$ is non-negative for all $i = 1, \dots, I$, and thus σ_t^2 is non-negative as well. Therefore, σ_t is well-defined as the square-root of σ_t^2 .

We now turn to a very general class of processes which also, by their nature, allow us to represent stochastic volatility, namely the so-called Lévy semi-stationary (\mathcal{LSS}) processes. In fact, this class provides a very complete theoretical framework that encompasses all the models presented so far. The idea to apply these kinds of processes as a framework to model energy spot prices was first described in detail by [BBV13].

Let $g, q: \mathbb{R}_+ \mapsto \mathbb{R}_+$ be two measurable deterministic functions such that $\lim_{t \rightarrow \infty} g(t) = \lim_{t \rightarrow \infty} q(t) = 0$, then an \mathcal{LSS} process $Z = \{Z_t\}_{t \in \mathbb{R}}$ is given by

$$Z_t = \mu + \int_{-\infty}^t g(t-s) \sigma_{s-} dL_s + \int_{-\infty}^t q(t-s) a_s ds, \quad t \in \mathbb{R}, \tag{3.9}$$

where $\mu \in \mathbb{R}$ is a constant and $L = (L_t)_{t \in \mathbb{R}}$ a two-sided Lévy process, i.e. a Lévy process which is defined on the entire real line. This can be constructed by taking an independent copy of $(L_t)_{t \geq 0}$, which we denote by $(\tilde{L}_t)_{t \geq 0}$ and defining $L_t := -\tilde{L}_{-(t-)}$ for $t < 0$. Further, $\sigma = (\sigma_t)_{t \in \mathbb{R}}$ and $a = (a_t)_{t \in \mathbb{R}}$ are càdlàg stochastic processes which are assumed to be independent of L . Typically, when using an \mathcal{LSS} process to model energy spot prices, only the stochastic integral part is used as a factor. That is, the drift is disregarded, i.e.

$$Z_t = \int_{-\infty}^t g(t-s) \sigma_{s-} dL_s. \tag{3.10}$$

In order to ensure the existence of this stochastic integral, we require suitable integrability conditions. For this, we assume that L and σ have finite second moments and that the kernel function g satisfies the following conditions

$$(i) \int_{-\infty}^t g^2(t-s)\mathbb{E}[\sigma_s^2] ds < \infty, \text{ for all } t \in \mathbb{R}, \text{ and}$$

(ii) there exists some $a \in (0, 1)$ such that

$$\int_0^\infty g^{2a}(x) dx < \infty, \quad \text{and} \quad \int_{-\infty}^t g^{2(1-a)}(t-s)\mathbb{E}[\sigma_s^2] ds < \infty, \quad t \in \mathbb{R}.$$

For a proof, we refer to [BBV13]. In particular, they show that under these conditions, the \mathcal{LSS} process (3.10) has finite second moments. Let us now consider a simple example.

Example 3.12. By choosing $g(t-s) = \exp(-\theta(t-s))$ for $\theta > 0$, constant volatility $\sigma \equiv 1$ and $L = B$ a two-sided Brownian motion for the Lévy process, we obtain the (stationary) Ornstein–Uhlenbeck process. Of course, it is easy to verify that the above integrability conditions are satisfied. Note that in the special case, where the driving Lévy process of an \mathcal{LSS} process is a two-sided Brownian motion, the \mathcal{LSS} process is also referred to as a Brownian semi-stationary (\mathcal{BSS}) process.

Notice that the lower integration bound in the definition of the \mathcal{LSS} process (3.10), is chosen to be $-\infty$ rather than 0. This is because it allows to have a stationary model. Indeed, the \mathcal{LSS} process (3.10) is stationary as soon as σ and a are stationary, which is also the reason they are named Lévy semi-stationary processes. Recall that all the energy spot price models we have examined until now rely on mean-reverting stochastic processes, specifically the Ornstein-Uhlenbeck process. This choice is driven by the understanding that energy spot prices represent equilibrium prices, which are determined by supply and demand. In that context, stationarity can be regarded as a weak form of mean-reversion since it ensures that the process can not move away from its long term mean indefinitely, but has to return to it since otherwise the stationarity would not be upheld. A further advantage of \mathcal{LSS} processes is that stochastic volatility and the possibility of jumps and extreme spikes are naturally incorporated into the model. Moreover, the kernel function g allows great flexibility in terms of modelling the autocorrelation structure, which many other models lack. For more details, we refer to [BBV13].

We consider another example.

Example 3.13 (CARMA(2, 1)). A general class of processes which are also included in the \mathcal{LSS} framework are the so-called continuous-time autoregressive moving average (CARMA) processes. Let the kernel g be given $g(t) = \alpha_1 e^{-\theta_1 t} + \alpha_2 e^{-\theta_2 t}$ for some

$\alpha_1, \alpha_2 \in \mathbb{R}$ and $\theta_1, \theta_2 > 0$. Then, the \mathcal{LSS} process takes the following form

$$Z_t = \alpha_1 \int_{-\infty}^t e^{\theta_1(t-s)} \sigma_{s-} dL_s + \alpha_2 \int_{-\infty}^t e^{\theta_2(t-s)} \sigma_{s-} dL_s$$

where L is a Lévy process with finite second moments. If we further assume that there is no stochastic volatility, i.e. $\sigma \equiv 1$, then Z is simply the sum of two (stationary) Ornstein-Uhlenbeck process driven by the same Lévy process L . In fact, this process is also referred to as a Lévy-driven CARMA(2, 1) process, which is a special case of the more general class of CARMA(p, q) processes.

Let us now consider a specific electricity spot price model which was investigated by [RVG21].

Example 3.14. Consider the following multifactor model

$$S_t = \Lambda_t + Z_t^{\text{long}} + Z_t^{\text{short}}, \quad t \geq 0$$

where Λ denotes a deterministic seasonality and trend function and Z^{long} describes the (non-stationary) long-term and low-frequency stochastic dynamics of the spot price modelled as a Lévy process. Namely, [RVG21] find that a symmetric normal inverse Gaussian distribution fits the data well. Further, Z^{short} is responsible for modelling the (stationary) short-term dynamics of the price. For this factor, two different modelling approaches were proposed, namely one scenario with exogenous variables and one scenario without. The approach with exogenous variables allows us to include other factors, such as wind forecasts, to model the electricity price. However, in this example, we only consider the setting without exogenous variables. The short-term factor is modelled as a Lévy-driven CARMA(2, 1) process, i.e.

$$Z_t^{\text{short}} = \alpha_1 \int_{-\infty}^t e^{\theta_1(t-s)} dL_s + \alpha_2 \int_{-\infty}^t e^{\theta_2(t-s)} dL_s, \quad t \geq 0,$$

see Example 3.13. For the driving Lévy process L , the authors test different generalized hyperbolic distributions and find that the asymmetric Student's t-distribution provides the best fit.

3.2 Simulation

To apply any of the introduced electricity price models, we need to be able to simulate them. Let us first take a look at the seasonality.

3.2.1 The seasonality component

Recall that seasonality is a significant characteristic found in electricity prices, and accordingly each model that we discussed was of the following form

$$S_t = \Lambda_t + \sum_{i=1}^d Z_t^{(i)}, \quad t \geq 0,$$

where Λ is a positive deterministic function modelling the seasonality, and $Z^{(i)}$, $i = 1, \dots, d$ stochastic processes modelling the stochastic component. In particular, the seasonality is modelled as an additive factor to the electricity price.

The two most commonly used seasonality functions are dummy variables and sinusoidal functions. A dummy variable is essentially an indicator function. For example, we may define a dummy variable for each day in the week $d_{\text{Monday}}, d_{\text{Tuesday}}, \dots, d_{\text{Sunday}}$. For this purpose, we assume that $t = 0$ corresponds to a certain date and time in the calendar. Then, each $t \geq 0$ corresponds to a certain date and time in the calendar which allows us to define the weekday dummy variables, e.g.

$$d_{\text{Monday}}(t) = \begin{cases} 1 & \text{if the date corresponding to } t \text{ is a Monday,} \\ 0 & \text{else,} \end{cases} \quad t \geq 0,$$

and respectively for every other weekday.

To deal with the yearly seasonality, we follow [RVG21] by modelling it with a sinusoidal, namely a cosine function with yearly frequency, combined with an intercept and a linear trend. The weekly seasonality can be modelled using the dummy variables defined above. However, the modelling of the daily seasonality is more involved.

Remark 3.15. It is important to mention that all the electricity price models presented in the previous sections are mainly focused on modelling daily average prices. This focus in the literature is because the dynamics of daily average prices are used as a reference price for various purposes, such as option contracts. As a result, these continuous-time price models are usually fitted using daily average prices, which are typically the averages over the 24 hourly prices determined in the day-ahead auction. As mentioned previously, the continuous time models are interpreted as continuous-time view of the continuous price dynamics as an unobserved price for immediate delivery of electricity. However, due to the fact that these models are fitted to daily prices, they are not capable of capturing the dynamics for hourly prices very well. For instance, the hourly prices typically revert to an hourly specific mean price that does not coincide with the daily mean price. Moreover, the level of mean-reversion may not be constant over the day and can be different depending

on the hour, and also the volatility structure might differ throughout the day. In addition, there may be correlation patterns that exist between specific hours. In the literature, there are different approaches to model hourly electricity prices. However, this would be out of scope of this thesis and hence, while not quite accurate, we will proceed by applying the presented continuous-time models to describe the hourly dynamics. By incorporating the hourly mean prices into the seasonality function, we can capture the most important feature of the hourly dynamics, but we neglect other characteristics such as differing levels of mean-reversion and cross-correlation.

Nevertheless, let us briefly discuss some of the existing approaches to deal with the modelling of hourly electricity prices found in the literature. For example, [VV14] use multivariate Lévy semi-stationary processes to describe the different day-ahead prices. [Hir09] considers a discrete stochastic two-factor 24-dimensional vector process to model the characteristics of the day-ahead prices and [HHM07] introduces a panel model for hourly electricity prices in the day-ahead markets.

In light of the previous remark, we define for each hour $h = 1, \dots, 24$ in the day the following hourly dummy variables

$$d^h(t) = \begin{cases} 1 & \text{if the time corresponding to } t \text{ is during the } h\text{-th hour of the day,} \\ 0 & \text{else,} \end{cases}$$

for $t \geq 0$. In particular, for $h = 1, \dots, 24$, we denote by $d_{\text{Monday}}^h(t) := d_{\text{Monday}}(t) \cdot d^h(t)$ the hourly Monday dummy variable, and respectively for every other weekday. To summarize, we model the seasonality function as follows

$$\Lambda(t) = c_0 + c_1 t + c_2 \cos\left(\frac{c_3 + 2\pi t}{365}\right) + \sum_{h=1}^{24} c_{\text{Monday}}^h d_{\text{Monday}}^h(t) + \dots + \sum_{h=1}^{24} c_{\text{Sunday}}^h d_{\text{Sunday}}^h(t). \quad (3.11)$$

Here c_0, c_1, c_2, c_3 and $c_{\text{Monday}}^h, \dots, c_{\text{Sunday}}^h$, $h = 1, \dots, 24$ are parameters of the seasonality function which need to be estimated.

3.2.2 The stochastic component

We now turn to the simulation of the stochastic component of electricity price models. More precisely, our goal is to generate sample paths of the stochastic factors presented in the previous sections. Recall that the dynamics of most of the stochastic factors seen so far can be described by a stochastic differential equation (SDE), e.g. the Ornstein-Uhlenbeck process. These SDEs are formulated in terms of a Brownian motion or, more generally, a Lévy process when modelling price spikes. In order to solve these SDEs

numerically, we thus first need to efficiently simulate sample paths of Brownian motion and Lévy processes. However, it is important to note that we cannot sample the entire path, e.g. $(B_t)_{t \geq 0}$, as it is an infinite-dimensional object. Instead, we focus on simulating the sample path on a discrete time grid. Fortunately, this simplification is not significant since electricity prices in spot markets, as we have discussed previously, are typically provided at hourly (or even quarterly) intervals. Hence, it is more than adequate to be able to generate discrete-time observations of the continuous-time process.

For this purpose, let $N \in \mathbb{N}$ and for some terminal time $T > 0$, consider the time discretization $\mathcal{T} := \{t_k = k\Delta t \mid k = 0, \dots, N\}$ where $\Delta t := T/N$. We first start with the simulation of Brownian motion. Our objective is thus to generate the exact joint distribution of $(B_{t_0}, B_{t_1}, \dots, B_{t_N})$. A straightforward way to achieve this is to use the fact that Brownian motion has independent, normally distributed increments. We start by setting $B_{t_0} = 0$. Then, by sampling independent increments $\Delta B_k := B_{t_k} - B_{t_{k-1}}$, $k = 1, \dots, N$, we can iteratively construct the sample Brownian motion by $B_{t_k} = B_{t_{k-1}} + \Delta B_k$. Note that the increments ΔB_k are normal distributed with mean 0 and variance Δt , and in particular, they can be generated by setting $\Delta B_k = \sqrt{\Delta t} \cdot \xi_k$, where ξ_k , $k = 1, \dots, N$, are independent, standard normal random variables. This way of sampling a trajectory of a Brownian motion is also called the random walk approach. In particular, since it is based on the independent increments property of Brownian motion, it is also applicable for Lévy processes. The only difference being that we need to replace the normal distribution with the corresponding infinitely divisible distribution of the Lévy process. However, it turns out that this is not as straightforward as it seems at first sight. This is due to the fact that not every infinitely divisible distributions is closed under convolution, i.e. if the distribution of L_1 follows a specific law, then L_t follows the same law (up to shifting and scaling) for every other point in time $t \geq 0$. It is easy to see that this property is indeed satisfied for Brownian motion. Another example is the Poisson process, which is Poisson distributed at all points in time. In particular, this makes compound Poisson processes simple to simulate. However, it turns out that the generalized hyperbolic distribution fails to be convolution closed. Notably, the only subclasses of the generalized hyperbolic distributions that do exhibit closure under convolution are the generalized Laplace distributions and the normal inverse Gaussian (NIG) distributions. Indeed, the NIG Lévy process is quite popular since it allows simulation via the random walk method.

Let us now turn to the numerical approximation of solutions of SDEs. For a deeper discussion of this topic, we refer the reader to [Gla13]. The simplest approximation method is the so-called Euler-Maruyama method. We first consider a stochastic process Z which

satisfies the following stochastic differential equation

$$dZ_t = a(t, Z_t)dt + b(t, Z_t)dB_t$$

with some fixed start value $Z_0 = z \in \mathbb{R}$, where B is a Brownian motion, and $a, b: \mathbb{R}_+ \times \mathbb{R} \rightarrow \mathbb{R}$ satisfy some suitable conditions that ensure the existence (and uniqueness) of the solution of the SDE, e.g. Lipschitz conditions. Our goal is now to derive an approximation \widehat{Z} of the solution Z on the time grid \mathcal{T} . Analogously to the Euler scheme in the ordinary differential equation (ODE) setting, we first set $\widehat{Z}_{t_0} = z$ and iteratively define

$$\widehat{Z}_{t_k} = \widehat{Z}_{t_{k-1}} + a(t_{k-1}, \widehat{Z}_{t_{k-1}})\Delta t + b(t_{k-1}, \widehat{Z}_{t_{k-1}})\Delta B_k, \quad k = 1, \dots, N. \quad (3.12)$$

Recall that ΔB_k are the Brownian increments, as defined above, and can be simulated by $\Delta B_k = \sqrt{\Delta t} \cdot \xi_k$, where $\xi_k, k = 1, \dots, N$, are independent, standard normal random variables. One can prove that the scheme \widehat{Z} converges to the solution Z . To be more precise, the Euler-Maruyama approximation \widehat{Z} converges with strong order 1/2, that is, it holds,

$$\mathbb{E} \left[\sup_{k=0, \dots, N} \left| \widehat{Z}_{t_k} - Z_{t_k} \right| \right] \leq C\sqrt{\Delta t},$$

where $C > 0$ is some constant only depending on the coefficients, a and b , the initial value z , and the time horizon T .

Of course, the Euler-Maruyama scheme can be extended for SDEs driven by more general processes such as Lévy processes. However, obtaining general convergence results for this case are difficult. Note that in our setting, we typically aim to simulate hourly electricity prices, which provides a natural time discretization. Nonetheless, despite this predefined frequency, it remains crucial to choose a sufficiently refined grid for the Euler-Maruyama approximation to enhance accuracy. Once the simulation with the finer time discretization is obtained, we can easily downsample the data to achieve the desired frequency.

Remark 3.16 (Simulation of \mathcal{LSS} processes). The simulation of general \mathcal{LSS} processes is much more involved and out of scope of this thesis. As such, we refer the reader to [BLP17] which proposes the simulation scheme for \mathcal{BSS} processes.

3.3 Modelling the impact of wind forecasts on electricity prices

Renewable energy sources are becoming increasingly important in modern electricity markets. Since such renewable sources are highly dependent on weather conditions, they tend to increase the volatility of prices. In Germany, the most important source of renewable

energy is wind. Hence, incorporating wind forecasts is crucial as they provide information about the expected wind production, which directly influences the availability of cheap wind power in the electricity grid. As a result, by considering the forecasted wind data, one can optimize their battery operation in anticipation of the fluctuating wind generation.

To effectively integrate wind forecasts into the price models, several relevant wind variables can be considered. For instance, the wind production WD_t represents the amount of electricity generated by wind turbines. The daily load demand LD_t refers to the electricity consumption profile of households. Additionally, the wind penetration index WPI_t measures the proportion of wind power in the total power production. In other words, it corresponds to the ratio between the wind production and the load demand, and hence it takes values between zero and one. In particular, the predicted wind penetration index can be viewed as forward-looking information, since it is available before the prices are determined in the day-ahead auction or intraday markets. Hence, it is reasonable to try to incorporate this information in the model.

As mentioned in Example 3.14, [RVG21] propose to incorporate wind data into the model by adding an additional factor to the short-term component. For this purpose, we define the wind variable $W_t = (WD_t, LD_t, WPI_t)^\top$ which combines all the above-mentioned wind factors. Since these wind variables have clear seasonal patterns, it is useful to first remove it by fitting suitable seasonality and trend functions, similarly to the method described in section 3.2.1, see also [RVG21]. Then, by subtracting the fitted seasonality function, we obtain a deseasonalized version of W_t , which we denote by $\bar{W}_t = (\bar{WD}_t, \bar{LD}_t, \bar{WPI}_t)^\top$. [RVG21] now propose to model the short-term component by

$$Z_t^{\text{short}} = \alpha_1 \int_{-\infty}^t e^{\lambda_1(t-s)} dL_s + \alpha_2 \int_{-\infty}^t e^{\lambda_2(t-s)} dL_s + \beta^\top \bar{W}_t, \quad t \geq 0, \quad (3.13)$$

where β are some coefficients that need to be estimated, e.g. using least squares regression. In other words, the short-term factor is modelled as a CARMA(2, 1) process and a linear combination of the wind variables. Note that we could also consider an alternative (deseasonalized) wind variable \bar{W}_t . For instance, we may only consider a subset, e.g. $\bar{W}_t = \bar{WPI}_t$, or polynomials such as $\bar{W}_t = (1, \bar{WD}_t, \bar{LD}_t, \bar{WPI}_t, \bar{WD}_t^2, \bar{LD}_t^2, \bar{WPI}_t^2)^\top$. In order to simulate these wind variables, [RVG21, Chapter 6] suggests modelling them using Brownian semi-stationary processes with gamma kernels as defined in (3.9). However, a drawback of this model is its general non-Markovian nature. In the subsequent chapter, we will introduce the optimal battery control problem as a Markov decision problem, assuming Markovian dynamics for the price process. Consequently, the approach proposed in [RVG21, Chapter 6] is unsuitable for our specific case.

To simplify the modelling, we focus on the case $\bar{W}_t = \bar{WPI}_t$ in what follows. In-

deed, [RVG21] find that a model incorporating solely the wind penetration index already improves the stochastic electricity price model and provides favourable results when compared to considering other additional wind-related variables. In order to model the wind penetration index, we follow [BDL18] in modelling the wind variables as an Ornstein-Uhlenbeck process. More precisely, they model the wind speed and wind power production WD with a Gaussian Ornstein-Uhlenbeck process. We will take a similar approach and model the wind penetration index using an Ornstein-Uhlenbeck process with a normal inverse Gaussian distributed Lévy process $(L_t^{(W)})_{t \geq 0}$ as BDLP, as this allows for a more flexible model. Thus, we assume the dynamics

$$d\bar{W}_t = -\theta\bar{W}_t dt + dL_t^{(W)}, \quad t \geq 0, \quad (3.14)$$

for the (deseasonalized) wind penetration index. It is important to note that we consider the Lévy process $L^{(W)}$ to be independent of the Lévy process L used for the CARMA(2, 1) process in (3.13). In particular, this chosen model allows for straightforward simulation, which can be achieved by applying the Euler-Maruyama scheme as described in the previous section.

Note that in the proposed model, the electricity price S_t at time t directly depends on the actual (deseasonalized) wind penetration index \bar{W}_t . However, in practice, we often have access to forecasts of the wind penetration index well in advance of the actual time. Hence, in order to model this fact, we simply assume that for any time $t \geq 0$ a forecast \bar{W}_t^f of the wind penetration index \bar{W}_t becomes available at time $\max\{0, t - \Delta_f\}$, where $\Delta_f \geq 0$ is some forecast period. It is worth noting that state-of-the-art forecasting models typically yield predictions that equal the actual value plus an error term ε_t , such that $\bar{W}_t^f = \bar{W}_t + \varepsilon_t$. For simplicity, however, we assume that we have perfect forecasting models in the sense that $\bar{W}_t^f = \bar{W}_t$. In particular, this implies that \bar{W}_t is assumed to be $\mathcal{F}_{\max\{0, t - \Delta_f\}}$ -measurable for any $t \geq 0$.

Remark 3.17. While the multifactor model considered above does incorporate the dependence of wind-related variables and electricity prices, other modelling approaches can be explored. For instance, we may consider a similar model to Example 3.11, namely

$$S_t = \Lambda(t) + Z_t, \quad \text{where} \quad dZ_t = (m_t - \alpha X_t) dt + \sigma_t dB_t, \quad t \geq 0.$$

In this alternative model, the constant mean μ is replaced by a stochastic mean m_t , modelled by a Lévy-driven Ornstein-Uhlenbeck process itself. An entirely different approach based on regime-switching Markov models is presented in [Ver16].

4 The optimal battery control problem

In this chapter, we finally address the challenge of optimizing the use of a battery storage to minimize the electricity cost of a consumer through the lens of stochastic optimal control. Building on our previous chapters exploring battery degradation modelling and stochastic electricity price models, we now present an approach to derive optimal battery control strategies.

The rest of this chapter is structured as follows. In the first section, we formulate the problem as a Markov decision problem, providing the theoretical foundation for our analysis. We then introduce the Monte Carlo least square method, enabling efficient numerical solutions for a Markov decision problem. Lastly, we present the optimal battery control problem, combining the insights from Chapter 2 and 3.

4.1 Stochastic optimal control

Stochastic control problems represent an important category of stochastic optimization problems that find applications in a wide variety of fields. Roughly speaking, these problems involve decision-makers seeking to find an “optimal” decision strategy that optimizes some pre-specified cost functional. More precisely, consider a given system that can be influenced through a series of sequential decisions. Given the current system state, which could represent, for instance, the state of a battery and the current electricity price, the decision maker faces the task of selecting an action, such as whether to charge or discharge a battery, or how much electricity to buy from the market. Once this action is chosen, the system transitions to a new state. In general, this transition depends on the chosen action as well as the evolution of an underlying stochastic process. To formulate an optimization criterion, we assume that each action yields a corresponding reward or cost to the decision maker. The goal is then to find a decision strategy that maximizes the optimizes the total rewards or costs.

In what follows, we assume that the underlying stochastic process follows Markovian dynamics, meaning that past states do not impact future states. In this case, the stochastic control problem, described informally above, is also known as a Markov decision problem. Let us now give a precise definition of a so-called Markov decision model. Note, however, that we will consider a specialized setting which is different compared to what is typically found in the literature. That is, we assume that the system’s state can be separated into a deterministic control which is only influenced by the chosen actions and a stochastic process which does not depend on the chosen action. And moreover, we assume that the possible actions that can be taken are finite. This kind of setting is considered in [GHW15], see also [Bay+22]. For a more general and comprehensive treatment

of this topic, we refer the reader to [BR11].

Definition 4.1 (Markov decision model). Let $N \in \mathbb{N}$ and $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space equipped with a filtration $\mathbb{F} = (\mathcal{F}_k)_{k=0}^N$. Then, a Markov decision model with time horizon N consists of a set of components $(\mathcal{X}, (X_k)_{k=0}^N, \mathcal{Y}, \mathcal{A}, (\mathcal{A}_k)_{k=0}^{N-1}, (\varphi_k)_{k=0}^{N-1}, (r_k)_{k=0}^{N-1}, g_N)$ with the following meanings.

- $\mathcal{X} \subset \mathbb{R}^{d_X}$ denotes the state space of the system, endowed with its corresponding Borel σ -algebra $\mathcal{B}(\mathcal{X})$. Moreover, $X = (X_k)_{k=0}^N$ is an \mathbb{F} -adapted Markov process taking values in \mathcal{X} . The state X_k represents the (stochastic) information that is available to the decision-maker at time step k . We refer to X as the state process.
- $\mathcal{Y} \subset \mathbb{R}^{d_Y}$ denotes the control space of the system, endowed with its corresponding Borel σ -algebra $\mathcal{B}(\mathcal{Y})$. At each point in time $k = 0, \dots, N$, we are given a control Y_k , which we assume to be \mathcal{F}_{k-1} measurable, taking values in \mathcal{Y} . For convenience, we extend the filtration \mathbb{F} by defining $\mathcal{F}_{-1} := \{\emptyset, \Omega\}$ and $\mathcal{F}_{N+1} := \mathcal{F}_N$. We refer to $Y = (Y_k)_{k=0}^N$ as the control process.
- \mathcal{A} denotes the finite action space. $\mathcal{A}_k \subset \mathcal{Y} \times \mathcal{X} \times \mathcal{A}$ is a measurable subset of $\mathcal{Y} \times \mathcal{X} \times \mathcal{A}$ and denotes the set of possible control-state-action combinations at time step k . We assume that \mathcal{A}_k contains the graph of some measurable map $\pi_k: \mathcal{Y} \times \mathcal{X} \rightarrow \mathcal{A}$, i.e. $(y, x, \pi_k(y, x)) \in \mathcal{A}_k$ for all $(y, x) \in \mathcal{Y} \times \mathcal{X}$. Moreover, $\mathcal{A}_k(y, x) = \{a \in \mathcal{A} \mid (y, x, a) \in \mathcal{A}_k\}$ denotes the set of admissible actions in control y and state x at time step k .
- $\varphi_k: \mathcal{Y} \times \mathcal{A} \rightarrow \mathcal{Y}$, $k = 1, \dots, N$, are measurable functions which describe the update rule of the control process. Namely, if we are given a control $Y_k \in \mathcal{Y}$, a state $X_k \in \mathcal{X}$ and an action $a \in \mathcal{A}_k(Y_k, X_k)$ at time step k , for any $k = 0, \dots, N-1$, then the control is updated according to $Y_{k+1} := \varphi_{k+1}(Y_k, a)$.
- $r_k: \mathcal{A}_k \rightarrow \mathbb{R}$, $k = 0, \dots, N-1$, are measurable functions where $r_k(y, x, a)$ represents the one-stage reward of the system at time step k if the current control is y , the current state is x , and action a is taken.
- $g_N: \mathcal{Y} \times \mathcal{X} \rightarrow \mathbb{R}$ is a measurable function where $g_N(y, x)$ gives the terminal reward of the system at time N if the state is x and the control is y .

Remark 4.2. Typically, stochastic transition kernels Q_k from \mathcal{A}_k to $\mathcal{Y} \times \mathcal{X}$ are considered in the definition of a Markov decision model instead of the (Markovian) state process X and control process Y as given in Definition 4.1. By employing a stochastic transition kernel, greater flexibility is achieved in capturing the influence of the chosen action on the state transition. For more details, we refer to [BR11].

It is worth noting again that we assume that the dynamics of the state process X do not depend on the control process or the actions taken. In contrast, the control process Y clearly depends on the sequence of actions taken. Such a sequence $A = (A_0, \dots, A_{N-1})$ is called a strategy or policy on $\{0, \dots, N-1\}$. Of course, due to our constraints, not every policy is allowed and hence, we define the set of admissible policies.

Definition 4.3 (Admissible policy). A sequence of random variables $A = (A_k, \dots, A_{N-1})$ taking values in \mathcal{A} is called an admissible policy or strategy on $\{k, \dots, N-1\}$ started in $y \in \mathcal{Y}$ at time step k if it satisfies the following properties

- (i) A_ℓ is \mathcal{F}_ℓ -measurable for all $\ell = k, \dots, N-1$, and
- (ii) $A_\ell \in \mathcal{A}_\ell(Y_\ell^A, X_\ell)$ for all $\ell = k, \dots, N-1$, where Y^A is the corresponding control process, defined recursively by $Y_k^A := y$ and $Y_{\ell+1}^A := \varphi_{\ell+1}(Y_\ell^A, A_\ell)$ for $\ell = k, \dots, N-1$.

We denote the set of admissible policies on $\{k, \dots, N-1\}$ by \mathcal{P}_k .

It is easily seen that the control process Y^A , as defined in Definition 4.3, is predictable for any admissible policy $A \in \mathcal{P}_k$, i.e. Y_ℓ^A is $\mathcal{F}_{\ell-1}$ -measurable for all $\ell = k, \dots, N$. It is also worth mentioning that the control process Y^A depends on the initial control y and time step k . We may also write $Y^{A;k,y}$ if we want to highlight the dependence on the initial condition $Y_k^A = y$.

Moreover, we have that the action A_ℓ taken at time step ℓ is \mathcal{F}_ℓ -measurable, which means that A_ℓ possibly depends on the entire history of the state process up until time k . However, in what follows, we will mainly consider policies which only depend on the current state and control.

Definition 4.4 (Markov policy). A measurable mapping $\pi_k: \mathcal{Y} \times \mathcal{X} \rightarrow \mathcal{A}$ is called a decision rule at time step k if it satisfies $\pi_k(y, x) \in \mathcal{A}_k(y, x)$ for all $(y, x) \in \mathcal{Y} \times \mathcal{X}$. We denote by Π_k the set of all decision rules at time step k .

A sequence of decision rules $\pi = (\pi_k, \pi_{k+1}, \dots, \pi_{N-1})$ with $\pi_\ell \in \Pi_\ell$, $\ell = k, \dots, N-1$ is called a Markov policy on $\{k, \dots, N-1\}$. We denote the set of Markov policies on $\{k, \dots, N-1\}$ by \mathcal{MP}_k .

Remark 4.5. Given some time step $k = 0, \dots, N-1$, an initial value $y \in \mathcal{Y}$ and a Markov policy $\pi \in \mathcal{MP}_k$, we define the corresponding control process $(Y_\ell^\pi)_{\ell=k}^N$ by

$$\begin{aligned} Y_k^\pi &= y \quad \text{and,} \\ Y_{\ell+1}^\pi &= \varphi_{\ell+1}(Y_\ell^\pi, \pi_\ell(Y_\ell^\pi, X_\ell)) \quad \text{for } \ell = k, \dots, N-1. \end{aligned}$$

That is, the action at time ℓ is given by $\pi_\ell(Y_\ell^\pi, X_\ell)$ and in particular, the sequence $(\pi_\ell(Y_\ell^\pi, X_\ell))_{\ell=k}^{N-1}$ defines an admissible policy. In what follows, we write Y^π for the control process, to signify the dependence of the policy on the control. Additionally, as mentioned earlier, we may use the notation $Y^{\pi; k, y}$ if we want to highlight the dependence of the control process on the initial condition $Y_k^\pi = y$.

Let us now turn to the optimization problem. Before that, however, we need to make sure that every appearing expectation is well-defined. For this purpose, we consider a so-called upper bounding function.

Definition 4.6 (Upper bounding function). A measurable function $b: \mathcal{Y} \times \mathcal{X} \rightarrow \mathbb{R}$ is called an upper bounding function for the Markov Decision Model if there exist constants $c_r, c_g, c_b > 0$ such that for all $k = 0, 1, \dots, N - 1$, it holds

- (i) $r_k^+(y, x, a) \leq c_r b(y, x)$,
- (ii) $g_N^+(y, x) \leq c_g b(x)$ for all $x \in E$, and
- (iii) $\mathbb{E}[b(\varphi_{k+1}(y, a), X_{k+1}) \mid X_k = x] \leq c_b b(y, x)$,

for all $(y, x) \in \mathcal{Y} \times \mathcal{X}$ and $a \in \mathcal{A}_k(y, x)$.

Moreover, we call a measurable function $u: \mathcal{Y} \times \mathcal{X} \rightarrow \mathbb{R}$ bounded by b , if there exists some constant $c > 0$ such that $|u(y, x)| \leq cb(y, x)$ for all $(y, x) \in \mathcal{Y} \times \mathcal{X}$.

Note that the above definition is slightly stricter than what is typically considered for an upper bounding function, see for instance [BR11, Definition 2.4.1]. To guarantee that all appearing expectations are well-defined, we impose the existence of such an upper bounding function for the Markov decision model. However, it is worth mentioning, that in general most results that follow in this section hold under much weaker assumptions, see for example [Hin70].

Assumption 4.7. We assume the existence of an upper bounding function b for the Markov Decision Model.

We are now ready to state the optimization problem.

Definition 4.8 (Value function). For a Markov policy $\pi \in \mathcal{P}_k$, we define the objective functional by

$$\mathcal{J}_k^\pi(y, x) := \mathbb{E} \left[\sum_{\ell=k}^{N-1} r_\ell(Y_\ell^\pi, X_\ell, \pi_\ell(Y_\ell^\pi, X_\ell)) + g_N(Y_N^\pi, X_N) \mid Y_k = y, X_k = x \right], \quad (4.1)$$

for $(y, x) \in \mathcal{Y} \times \mathcal{X}$. Then, the optimization problem is given by

$$v_k(y, x) := \sup_{\pi \in \mathcal{MP}_k} \mathcal{J}_k^\pi(y, x), \quad (y, x) \in \mathcal{Y} \times \mathcal{X}. \quad (4.2)$$

We refer to v_k as the value function.

In other words, the value function $v_k(y, x)$ represents the maximum expected cumulative reward achievable with Markov policies starting from time step k , initial control y , and initial state x . It is also straight forward to see that Assumption 4.7 ensures that the objective functional and value function are well-defined, see [BR11, Proposition 2.4.2]. It is worth noting that the value function v_k is in general not guaranteed to be measurable, which causes theoretical inconveniences.

The main goal is to compute the value function $v_0(y, x)$, and to find an optimal policy $\pi^* \in \mathcal{P}_k$.

Definition 4.9 (Optimal policy). For $k = 0, \dots, N - 1$, a Markov policy $\pi^* \in \mathcal{MP}_k$ is called optimal, if it satisfies $v_k(y, x) = \mathcal{J}_k^{\pi^*}(y, x)$ for all $(y, x) \in \mathcal{Y} \times \mathcal{X}$.

It is worth recalling that for the optimization problem (4.2), we only look at Markov policies, that is, admissible policies that only depend on the current state and control instead of the complete history. At this point, one may wonder whether considering history-dependent policies could improve the maximal expected rewards. It turns out that this is not the case.

Theorem 4.10. For $k = 0, \dots, N - 1$, it holds

$$v_k(y, x) = \sup_{A=(A_k, \dots, A_{N-1}) \in \mathcal{P}_k} \mathbb{E} \left[\sum_{\ell=k}^{N-1} r_\ell(Y_\ell^A, X_\ell, A_\ell) + g_N(Y_N^A, X_N) \mid Y_k = y, X_k = x \right],$$

for all $(y, x) \in \mathcal{Y} \times \mathcal{X}$, where $(Y_\ell^A)_{\ell=k}^{N-1}$ is defined as in Definition 4.3.

Proof. This result was proven in [Hin70, Theorem 18.4]. □

4.1.1 Bellman equation

In this section, we show that the sequence of value functions $(v_k)_{k=0}^N$ satisfies the so-called Bellman equation, which is given by

$$\begin{aligned} v_N(y, x) &= g_N(y, x), \\ v_k(y, x) &= \sup_{a \in \mathcal{A}_k(y, x)} \{r_k(y, x, a) + \mathbb{E}[v_{k+1}(\varphi_{k+1}(y, a), X_{k+1}) \mid X_k = x]\} \end{aligned}$$

for $k = 0, \dots, N - 1$. We can use the Bellman equation to find the value function by iteratively solving the equation backwards, starting from the last step and working our way back. It is important to stress that at this point it is not clear whether the statement itself is well-posed, since we do not even know yet if the value function v_k is a measurable function. However, it is easy to see that the objective functional \mathcal{J}_k^π is measurable for any Markov policy $\pi \in \mathcal{MP}_k$, and can be computed recursively as follows.

Lemma 4.11 (Reward iteration). *Let $\pi = (\pi_0, \dots, \pi_{N-1}) \in \mathcal{MP}_0$ be a Markov policy. Then, $\mathcal{J}_N^\pi = g_N$, and for $k = 0, 1, \dots, N - 1$ it holds*

$$\mathcal{J}_k^\pi(y, x) = r_k(y, x, \pi_k(y, x)) + \mathbb{E} [\mathcal{J}_{k+1}^\pi(\varphi_{k+1}(y, \pi_{k+1}(y, x)), X_{k+1}) \mid X_k = x],$$

for all $(y, x) \in \mathcal{Y} \times \mathcal{X}$.

Proof. It is obvious that $\mathcal{J}_N^\pi = g_N$. Let $(y, x) \in \mathcal{Y} \times \mathcal{X}$ and $k \in \{0, 1, \dots, N - 1\}$, then we compute

$$\begin{aligned} \mathcal{J}_k^\pi(y, x) &= \mathbb{E} \left[\sum_{\ell=k}^{N-1} r_\ell(Y_\ell^{\pi;k,y}, X_\ell, \pi_\ell(X_\ell)) + g_N(Y_N^{\pi;k,y}, X_N) \mid X_k = x \right] \\ &= r_k(y, x, \pi_k(y, x)) + \mathbb{E} \left[\sum_{\ell=k+1}^{N-1} r_\ell(Y_\ell^{\pi;k,y}, X_\ell, \pi_\ell(X_\ell)) + g_N(Y_N^{\pi;k,y}, X_N) \mid X_k = x \right] \\ &= r_k(y, x, \pi_k(y, x)) + \mathbb{E} [\mathcal{J}_{k+1}^\pi(Y_{k+1}^{\pi;k,y}, X_{k+1}) \mid X_k = x] \\ &= r_k(y, x, \pi_k(y, x)) + \mathbb{E} [\mathcal{J}_{k+1}^\pi(\varphi_{k+1}(y, \pi_{k+1}(y, x)), X_{k+1}) \mid X_k = x] \end{aligned}$$

where we conditioned the inner term of the expectation on X_{k+1} using the tower property in the third equation, and used $Y_{k+1}^{\pi;k,y} = \varphi_{k+1}(\pi_{k+1}(y, x), y)$ in the fourth equation. \square

The following definition will be instrumental in finding an optimal policy.

Definition 4.12. Let $u: \mathcal{Y} \times \mathcal{X} \rightarrow \mathbb{R}$ be a measurable function. A decision rule $\pi_k^* \in \Pi_k$ is called a maximizer of u at time step k if it satisfies

$$\begin{aligned} r_k(y, x, \pi_k^*(y, x)) + \mathbb{E} [u(\varphi_{k+1}(y, \pi_k^*(y, x)), X_{k+1}) \mid X_k = x] \\ = \sup_{a \in \mathcal{A}_k(y, x)} \{r_k(y, x, a) + \mathbb{E} [u(\varphi_{k+1}(y, a), X_{k+1}) \mid X_k = x]\}, \end{aligned}$$

for all $(y, x) \in \mathcal{Y} \times \mathcal{X}$.

Remark 4.13. It is easily seen that any maximizer π_k^* of the u at time step k must satisfy

$$\pi_k^*(y, x) \in \arg \max_{a \in \mathcal{A}(y, x)} \{r_k(y, x, a) + \mathbb{E} [u(\varphi_{k+1}(y, a), X_{k+1}) \mid X_k = x]\}$$

for all $(y, x) \in \mathcal{Y} \times \mathcal{X}$.

In the next theorem we show that whenever a solution of the Bellman equation exists together with a sequence of maximizers, then this yields the solution of our optimization problem.

Theorem 4.14 (Verification Theorem). *Let $(u_k)_{k=0,\dots,N}$ be a sequence of measurable functions which satisfy the Bellman equation. Then it holds*

- (i) $u_k \geq v_k$ for $k = 0, 1, \dots, N$, and
- (ii) if π_k^* is a maximizer of u_{k+1} for $k = 0, 1, \dots, N - 1$, then $u_k = v_k$ and the Markov policy $\pi^* = (\pi_0^*, \pi_1^*, \dots, \pi_{N-1}^*)$ is optimal.

Proof. (i) We first show $u_k \geq v_k$ for $k = 0, 1, \dots, N$ by backward induction. For $k = N$, we clearly have $u_N = g_N = v_N$. For $k \in \{0, 1, \dots, N - 1\}$, suppose that $u_{k+1} \geq v_{k+1}$. Let $\pi = (\pi_0, \dots, \pi_{N-1}) \in \mathcal{MP}_k$ be some arbitrary Markov policy. Then, it is obvious that $u_{k+1} \geq v_{k+1} \geq \mathcal{J}_{k+1}^\pi$ and using Lemma 4.11 we compute

$$\begin{aligned} u_k(y, x) &= \sup_{a \in \mathcal{A}_k(y, x)} \{r_k(y, x, a) + \mathbb{E}[u_{k+1}(\varphi_{k+1}(y, a), X_{k+1}) \mid X_k = x]\} \\ &\geq r_k(y, x, \pi_k(y, x)) + \mathbb{E}[u_{k+1}(\varphi_{k+1}(y, \pi_k(y, x)), X_{k+1}) \mid X_k = x] \\ &\geq r_k(y, x, \pi_k(y, x)) + \mathbb{E}[\mathcal{J}_{k+1}^\pi(\varphi_{k+1}(y, \pi_k(y, x)), X_{k+1}) \mid X_k = x] \\ &= \mathcal{J}_k^\pi(y, x) \end{aligned}$$

for all $(y, x) \in \mathcal{Y} \times \mathcal{X}$. Taking the supremum over all Markov policies $\pi \in \mathcal{MP}_k$ yields $u_k \geq v_k$.

(ii) We prove $u_k = v_k = \mathcal{J}_k^{\pi^*}$ for $k = 0, 1, \dots, N$ again using backward induction. For $k = N$ this statement is clearly true. For $k \in \{0, 1, \dots, N - 1\}$, we suppose that $u_{k+1} = v_{k+1} = \mathcal{J}_{k+1}^{\pi^*}$. Since π_k^* is a maximizer of u_{k+1} at time step k , we obtain

$$\begin{aligned} u_k(y, x) &= \sup_{a \in \mathcal{A}_k(y, x)} \{r_k(y, x, a) + \mathbb{E}[u_{k+1}(\varphi_{k+1}(y, a), X_{k+1}) \mid X_k = x]\} \\ &= r_k(y, x, \pi_k^*(y, x)) + \mathbb{E}[u_{k+1}(\varphi_{k+1}(y, \pi_k^*(y, x)), X_{k+1}) \mid X_k = x] \\ &= r_k(y, x, \pi_k^*(y, x)) + \mathbb{E}[\mathcal{J}_{k+1}^{\pi^*}(\varphi_{k+1}(y, \pi_k^*(y, x)), X_{k+1}) \mid X_k = x] \\ &= \mathcal{J}_k^{\pi^*}(y, x) \end{aligned}$$

for all $(y, x) \in \mathcal{Y} \times \mathcal{X}$ where we used Lemma 4.11 in the last equation. As a result, we have $v_k \leq u_k = \mathcal{J}_k^{\pi^*} \leq v_k$, which finishes the proof. \square

The above theorem is sufficient for applications where a solution of the Bellman equation is obvious and the existence of maximizers clear. For example, when the state,

control, and action spaces are all finite. In our setting, the state and control space are not finite and thus, there is no clear solution of the Bellman equation. As a result, we are still left with showing that the sequence of value functions are measurable and satisfy the Bellman equation. To achieve this, the following result plays a crucial role.

Lemma 4.15. *Let $u: \mathcal{Y} \times \mathcal{X} \rightarrow \mathbb{R}$ be a measurable function bounded by b . Then, the function $w_k: \mathcal{Y} \times \mathcal{X} \rightarrow \mathbb{R}$ defined by*

$$w_k(y, x) := \sup_{a \in \mathcal{A}_k(y, x)} \{r_k(y, x, a) + \mathbb{E}[u(\varphi_{k+1}(y, a), X_{k+1}) \mid X_k = x]\}$$

is measurable and bounded by b , and there exists a maximizer $\pi_k^ \in \Pi_k$ of u at time step k , for every $k = 0, \dots, N - 1$.*

Proof. First, we define the function $\tilde{w}_k: \mathcal{Y} \times \mathcal{X} \times \mathcal{A} \rightarrow \mathbb{R}$ with

$$\tilde{w}_k(y, x, a) := r_k(y, x, a) + \mathbb{E}[u(\varphi_{k+1}(y, a), X_{k+1}) \mid X_k = x]$$

for every $k = 0, \dots, N - 1$. Since r_k and u are measurable, it is easily seen that for all $a \in \mathcal{A}$, the map $(y, x) \mapsto \tilde{w}_k(y, x, a)$ is measurable as well. In order to prove that w_k is measurable, it is sufficient to show that $\{(y, x) \in \mathcal{Y} \times \mathcal{X} \mid w_k(y, x) \geq \alpha\}$ is measurable for any $\alpha \in \mathbb{R}$. We observe

$$\begin{aligned} \{(y, x) \in \mathcal{Y} \times \mathcal{X} \mid w_k(y, x) \geq \alpha\} &= \{(y, x) \in \mathcal{Y} \times \mathcal{X} \mid \exists a \in \mathcal{A}_k(y, x): \tilde{w}_k(y, x, a) \geq \alpha\} \\ &= \text{proj}_{\mathcal{Y} \times \mathcal{X}} \{(y, x, a) \in \mathcal{A}_k \mid \tilde{w}_k(y, x, a) \geq \alpha\} \end{aligned}$$

where $\text{proj}_{\mathcal{Y} \times \mathcal{X}}$ denotes the projection of \mathcal{A}_k on $\mathcal{Y} \times \mathcal{X}$. It turns out that this set is actually measurable by a result of Kunugui and Novikov, see [HPV76] for more details. In fact, they prove that if $B \subset \mathcal{Y} \times \mathcal{X} \times \mathcal{A}$ is measurable such that $\{a \in \mathcal{A} \mid (y, x, a) \in B\}$ is compact for all $(y, x) \in \mathcal{Y} \times \mathcal{X}$, then $\text{proj}_{\mathcal{Y} \times \mathcal{X}} B$ is a measurable subset of $\mathcal{Y} \times \mathcal{X}$. It is easy to see that $\{(y, x, a) \in \mathcal{A}_k \mid \tilde{w}_k(y, x, a) \geq \alpha\}$ satisfies these conditions since \tilde{w}_k is measurable and \mathcal{A} is finite by assumption. Thus, we have shown that w_k is measurable for all $k = 0, \dots, N - 1$.

Next, we define the set $\mathcal{D}_k(y, x) = \{a \in \mathcal{A}_k(y, x) \mid \tilde{w}_k(y, x, a) = w_k(y, x)\}$. Note that this set is compact, since \mathcal{A} is finite. Moreover, $\{(y, x, a) \in \mathcal{Y} \times \mathcal{X} \times \mathcal{A} \mid a \in \mathcal{D}(y, x)\}$ is measurable because w_k and \tilde{w}_k are measurable. It now follows from the so-called Kuratowski and Ryll-Nardzewski selection theorem, see [KR65], that there exists a measurable function $\pi_k^*: \mathcal{Y} \times \mathcal{X} \rightarrow \mathcal{A}$ such that $\pi_k^*(y, x) \in \mathcal{D}(y, x)$ for all $(y, x) \in \mathcal{Y} \times \mathcal{X}$. In particular, this implies that $\tilde{w}_k(y, x, \pi_k^*(y, x)) = w_k(y, x)$, i.e. π_k^* is a maximizer of u .

We are left with showing that w_k is bounded by b . This follows easily from the

properties of the upper bounding function b , namely

$$\begin{aligned} |w_k(y, x)| &\leq \sup_{a \in \mathcal{A}_k(y, x)} \{|r_k(y, x, a)| + \mathbb{E}[|u(\varphi_{k+1}(y, a), X_{k+1})| \mid X_k = x]\} \\ &\leq \sup_{a \in \mathcal{A}_k(y, x)} \{c_r b(y, x) + \mathbb{E}[b(\varphi_{k+1}(y, a), X_{k+1}) \mid X_k = x]\} \\ &\leq \sup_{a \in \mathcal{A}_k(y, x)} \{c_r b(y, x) + c_b b(y, x)\} = (c_r + c_b)b(y, x), \end{aligned}$$

for all $(y, x) \in \mathcal{Y} \times \mathcal{X}$ and $k = 0, \dots, N - 1$, which is what we wanted to show. \square

Note that in the above proof, we have heavily relied on the assumption that \mathcal{A} is finite. However, one can show that the result holds under much weaker assumptions. For more details, we refer to [BR11, Proposition 2.4.11]. The next theorem is the main result of this section.

Theorem 4.16. *Each value function v_k is measurable, and the sequence $(v_k)_{k=0,1,\dots,N}$ satisfies the Bellman equation, i.e. for all $(y, x) \in \mathcal{Y} \times \mathcal{X}$, it holds*

$$v_N(y, x) = g_N(y, x), \text{ and}$$

$$v_k(y, x) = \sup_{a \in \mathcal{A}_k(y, x)} \{r_k(y, x, a) + \mathbb{E}[v_{k+1}(\varphi_{k+1}(y, a), X_{k+1}) \mid X_k = x]\},$$

for $k = 0, 1, \dots, N - 1$. Moreover, there exist maximizers π_k^* of v_{k+1} with $\pi_k^* \in \Pi_k$, $k = 0, 1, \dots, N - 1$, and every sequence of maximizers $\pi_k^* \in \Pi_k$ of v_{k+1} defines an optimal policy $\pi^* = (\pi_0^*, \pi_1^*, \dots, \pi_{N-1}^*)$ for the Markov Decision Problem.

Proof. We first set $u_N = g_N$. In particular, we know that g_N is measurable and bounded by the upper bounding function b . We then define the sequence of functions $(u_k)_{k=0,\dots,N}$ recursively by

$$u_N(y, x) = g_N(y, x), \text{ and}$$

$$u_k(y, x) = \sup_{a \in \mathcal{A}_k(y, x)} \{r_k(y, x, a) + \mathbb{E}[u_{k+1}(\varphi_{k+1}(y, a), X_{k+1}) \mid X_k = x]\},$$

for $k = 0, 1, \dots, N - 1$. Here, Lemma 4.15 ensures by backward induction that each u_k is well-defined, measurable, and bounded by b . In particular, $(u_k)_{k=0,\dots,N}$ satisfy the Bellman equation by construction. Moreover, Lemma 4.15 guarantees the existence of maximizers $\pi_k^* \in \Pi_k$ of u_{k+1} at time step k for every $k = 0, \dots, N - 1$. Thus, by Theorem 4.14, we have that $u_k = v_k$, i.e. $(v_k)_{k=0,\dots,N}$ is a sequence of measurable functions which satisfies the Bellman equation, and $\pi^* = (\pi_0^*, \pi_1^*, \dots, \pi_{N-1}^*)$ is an optimal policy. \square

4.2 Numerical methods

We now turn to the topic of solving the Markov decision problem (4.2) numerically. A popular approach, based on Theorem 4.16, is the so-called least squares Monte Carlo method, which we will be the focus of this section. In fact, we will consider a version that is specifically for the setting of the previous section. The main idea is to simulate a set of trajectories and then apply a dynamic programming formulation, namely the Bellman equation, to recursively compute estimates for the value function using least squares regression. For a more thorough treatment of Monte-Carlo based methods for solving optimal stopping and control problems, we refer the reader to [BS18].

Let us first recall that Theorem 4.16 induces a backward induction algorithm to solve the optimization problem. That is, we first set $v_N = g_N$ and then compute the sequence $(v_k)_{k=0}^N$ recursively using the Bellman equation. By Theorem 4.16, we know that an optimal policy is given by the sequence of maximizers of the value function. Since we assumed the action set to be finite, computing a maximizer is straightforward, see Remark 4.13. However, computing v_k from v_{k+1} is more challenging. The main difficulty here being the computation of the conditional expectation in the Bellman equation. To deal with this problem, let us first define the so-called the continuation function

$$c_k(y, x) := \mathbb{E}[v_{k+1}(y, X_{k+1}) \mid X_k = x], \quad (y, x) \in \mathcal{Y} \times \mathcal{X}, \quad (4.3)$$

for $k = 0, \dots, N - 1$. Then, the Bellman equation can be written as

$$\begin{aligned} v_N(y, x) &= g_N(y, x), \\ v_k(y, x) &= \sup_{a \in \mathcal{A}_k(y, x)} \{r_k(y, x, a) + c_k(\varphi_{k+1}(y, a), x)\}, \end{aligned}$$

for all $k = 0, \dots, N - 1$. That is, we introduce an intermediate step in which we first compute the continuation function. Subsequently, computing the value function becomes straightforward since the action set is finite. Recall that the conditional expectation $\mathbb{E}[v_{k+1}(y, X_{k+1}) \mid X_k]$ can be interpreted as the orthogonal projection in the L^2 sense from $v_{k+1}(y, X_{k+1})$ onto the linear subspace of X_k -measurable functions. In other words, the continuation function $c_k(y, \cdot)$ can be characterized as a minimizer of the mean squared distance $\mathbb{E}[|v_{k+1}(y, X_{k+1}) - f(y, X_k)|^2]$ over all measurable functions $f(y, \cdot): \mathcal{X} \rightarrow \mathbb{R}$ up to $\mathbb{P} \circ X_k^{-1}$ -almost sure equality. In general, this minimization problem cannot be solved exactly, e.g. when the joint distribution is not known precisely, or the problem is too complicated to admit a closed-form solution. Hence, in such cases, it can be approximated by minimizing an empirical mean squared distance. To be more precise, we first sample $M \in \mathbb{N}$ independently sampled trajectories $(X_k^{(m)})_{k=0}^N$, $m = 1, \dots, M$ from the underlying

(Markovian) state process $(X_k)_{k=0}^N$ and minimize

$$\frac{1}{M} \sum_{m=1}^M \left| v_{k+1} \left(y, X_{k+1}^{(m)} \right) - f(y, X_k^{(m)}) \right|^2$$

over a suitable family of measurable functions $f(y, \cdot): \mathcal{X} \rightarrow \mathbb{R}$. It is worth noting that the minimization problem at hand offers the flexibility to employ various methods, including machine learning approaches, e.g. [GBC16]. In the subsequent discussion, we will delve into the application of least squares regression as an approach to approximate the continuation function.

For this purpose, let $\{\psi_1, \dots, \psi_p\}$ with $\psi_j: \mathcal{X} \rightarrow \mathbb{R}$, $j = 1, \dots, p$ be a set of basis functions, and we write $\psi(x) := (\psi_1(x), \dots, \psi_p(x))$. We denote by $(\hat{v}_k)_{k=0}^N$ the sequence of approximated value functions and set $\hat{v}_N := g_N$. We then estimate \hat{v}_k recursively as follows. For a fixed $y \in \mathcal{Y}$ and given approximated value function \hat{v}_{k+1} , we approximate the continuation function c_k by

$$\hat{c}_k(y, x) := \sum_{j=1}^p \beta_{k,j}^{(y)} \psi_j(x), \quad (4.4)$$

where the regression coefficients $\beta_k^{(y)} = (\beta_{k,1}^{(y)}, \dots, \beta_{k,p}^{(y)})^\top$ are given by the solution to the least squares problem

$$\beta_{k,1}^{(y)}, \dots, \beta_{k,p}^{(y)} := \arg \min_{\beta_1, \dots, \beta_p} \sum_{m=1}^M \left| \hat{v}_{k+1} \left(y, X_{k+1}^{(m)} \right) - \sum_{j=1}^p \beta_j \psi_j \left(X_k^{(m)} \right) \right|^2 \quad (4.5)$$

It is well known that the solution of (4.5) gives the coefficients of the best projection in L^2 sense of the conditional expectation (4.3) onto the linear space generated by the set of basis functions ψ . In fact, (4.5) can be solved explicitly, and the solution is given by

$$\beta_k^{(y)} = (F_k^\top F_k)^{-1} F_k^\top \hat{V}_{k+1}^{(y)}, \quad (4.6)$$

where

$$F_k := \left(\psi \left(X_k^{(1)} \right), \dots, \psi \left(X_k^{(M)} \right) \right)^\top \in \mathbb{R}^{M \times p}$$

is the so-called design matrix and

$$\hat{V}_{k+1}^{(y)} := \left(\hat{v}_{k+1} \left(y, X_{k+1}^{(1)} \right), \dots, \hat{v}_{k+1} \left(y, X_{k+1}^{(M)} \right) \right)^\top \in \mathbb{R}^M$$

is referred to as the target variable. Having approximated the continuation function, the

corresponding (approximated) value function is then simply given by

$$\hat{v}_k(y, x) := \max_{a \in \mathcal{A}_k(y, x)} \{r_k(y, x, a) + \hat{c}_k(\varphi_{k+1}(y, a), x)\} \quad (4.7)$$

for $(y, x) \in \mathcal{Y} \times \mathcal{X}$.

Example 4.17 (Polynomial basis functions). It is worth mentioning that the performance of the above regression algorithm clearly depends on the choice of basis functions ψ . The class of polynomials is a popular choice, since by the Stone-Weierstrass theorem any continuous function can be approximated by a polynomial. As such, we define the following set of polynomial regression basis functions

$$\text{Poly}_d(\mathcal{X}) = \left\{ \mathcal{X} \rightarrow \mathbb{R}: (x_1, \dots, x_{d_X}) \mapsto \prod_{i=1}^{d_X} x_i^{q_i} \mid q_1, \dots, q_{d_X} \in \{0, \dots, d\} : \sum_{i=1}^{d_X} q_i \leq d \right\},$$

for every degree $d \in \mathbb{N}$.

Note that in order to approximate the continuation function (4.4), we have fixed the control y . However, we would like to evaluate the value function on any arbitrary control $y \in \mathcal{Y}$. In our current setup, this would require to perform a linear regression for each y , which is unfortunately impractical, since the control set \mathcal{Y} is in general not finite.

Remark 4.18. At this point, one might question why we are not extending the regression step to encompass the control process as well. In other words, we could approximate the conditional expectation $\mathbb{E}[\hat{v}_{k+1}(\varphi(Y_k, a), X_{k+1}) \mid Y_k = y, X_k = x]$ for all admissible actions $a \in \mathcal{A}_k(y, x)$ instead of (4.3) for fixed y . However, this approach presents challenges because we lack knowledge about the distribution of the control variable Y_k , and hence we can not accurately simulate the control process since it depends on the underlying state process X and the optimal policy which we are trying to compute.

[Bay+22] consider a similar setting, and bypass this problem by assuming that the control set \mathcal{Y} is finite. This can be achieved, for instance, by considering a suitable finite discretization $\tilde{\mathcal{Y}} \subset \mathcal{Y}$ of the control set instead. However, for this to work, we would require that the update rules satisfy $\varphi_{k+1}(y, a) \in \tilde{\mathcal{Y}}$ for all $y \in \tilde{\mathcal{Y}}$ and $a \in \bigcup_{x \in \mathcal{X}} \mathcal{A}_k(y, x)$. Unfortunately, to ensure that this condition holds, the discretization set $\tilde{\mathcal{Y}}$ typically has to be chosen quite large and grows rapidly with the number of time steps N such that our algorithm becomes infeasible.

Thus, in what follows, we will consider a different approach to solve the problem described in Remark 4.18. Namely, we also consider a suitable discretization of the control set $\tilde{\mathcal{Y}} \subset \mathcal{Y}$. Then, we perform the least squares regression as described above for every $y \in \tilde{\mathcal{Y}}$, which yields an approximated continuation function \hat{c}_k defined on $\tilde{\mathcal{Y}} \times \mathcal{X}$. In order

to obtain an approximation on $\mathcal{Y} \times \mathcal{X}$, we then simply use an interpolation method. To be more precise, we consider some interpolation function $f_{\text{interpolate}}(y; \hat{c}_k(\tilde{y}, x) : \tilde{y} \in \tilde{\mathcal{Y}})$ which interpolates in the y component and is parameterized by the values of the approximated continuation function in the discretization points. There are various interpolation methods that one could consider. A simple example is linear interpolation.

Example 4.19 (Linear interpolation). We first consider linear interpolation in one dimension, i.e. we assume $\mathcal{Y} \subset \mathbb{R}$. Our objective is to interpolate the continuation function in $(y, x) \in \mathcal{Y} \times \mathcal{X}$. By assumption, we have already approximated the value of the continuation function $\hat{c}(y, x)$ for all $(y, x) \in \tilde{\mathcal{Y}} \times \mathcal{X}$ by the least squares regression method. In the following, we assume that for each $y \in \mathcal{Y}$, there exist $y^{\text{lower}}, y^{\text{upper}} \in \tilde{\mathcal{Y}}$ such that $y^{\text{lower}} \leq y \leq y^{\text{upper}}$. Then, the maps $f_{\text{lower}}, f_{\text{upper}} : \mathcal{Y} \rightarrow \tilde{\mathcal{Y}}$ given by

$$f_{\text{lower}}(y) = \max \left\{ \tilde{y} \in \tilde{\mathcal{Y}} : \tilde{y} \leq y \right\} \quad \text{and} \quad f_{\text{upper}}(y) = \min \left\{ \tilde{y} \in \tilde{\mathcal{Y}} : \tilde{y} \geq y \right\}$$

are well-defined. In particular, we are able to define the interpolated continuation function by

$$\hat{c}_k(y, x) := \frac{\hat{c}_k(f_{\text{lower}}(y), x)(f_{\text{upper}}(y) - y) + \hat{c}_k(f_{\text{upper}}(y), x)(y - f_{\text{lower}}(y))}{f_{\text{upper}}(y) - f_{\text{lower}}(y)} \quad (4.8)$$

for all $(y, x) \in \mathcal{Y} \times \mathcal{X}$.

Let us also consider the two-dimensional case, i.e. $\mathcal{Y} \subset \mathbb{R}^2$ and we write $y = (y_1, y_2) \in \mathcal{Y}$. We present the method of bilinear interpolation. For convenience, we assume that the discretization set $\tilde{\mathcal{Y}}$ is chosen as a regular grid, i.e. we can write $\tilde{\mathcal{Y}} = \tilde{\mathcal{Y}}_1 \times \tilde{\mathcal{Y}}_2$ for some finite sets $\tilde{\mathcal{Y}}_1, \tilde{\mathcal{Y}}_2$. Similarly to the one-dimensional case, we assume that the maps $f_{\text{lower}}^{(i)}, f_{\text{upper}}^{(i)} : \mathcal{Y} \rightarrow \tilde{\mathcal{Y}}_i$, $i = 1, 2$, defined by

$$f_{\text{lower}}^{(i)}(y) = \max \left\{ \tilde{y}_i \in \tilde{\mathcal{Y}}_i : \tilde{y}_i \leq y_i \right\} \quad \text{and} \quad f_{\text{upper}}^{(i)}(y) = \min \left\{ \tilde{y}_i \in \tilde{\mathcal{Y}}_i : \tilde{y}_i \geq y_i \right\}$$

for $y = (y_1, y_2) \in \mathcal{Y}$, are well-defined. Moreover, we set

$$f_{\theta_1, \theta_2} : \mathcal{Y} \rightarrow \tilde{\mathcal{Y}}, (y_1, y_2) \mapsto \left(f_{\theta_1}^{(1)}(y_1), f_{\theta_2}^{(2)}(y_2) \right)$$

for $\theta_1, \theta_2 \in \{\text{lower}, \text{upper}\}$. The bilinear interpolation can be described as a repeated linear interpolation, that is, a linear interpolation in the first dimension followed by a linear interpolation in the second dimension. We start with the linear interpolation in the

first dimension. Namely, for any $(y, x) = ((y_1, y_2), x) \in \mathcal{Y} \times \mathcal{X}$, we set

$$\begin{aligned}\widehat{c}_k\left(\left(y_1, f_{\theta_2}^{(2)}(y)\right), x\right) &:= \\ \widehat{c}_k\left(\left(f_{\text{lower}}^{(1)}(y), f_{\theta_2}^{(2)}(y)\right), x\right)\left(f_{\text{upper}}^{(1)}(y) - y_1\right) + \widehat{c}_k\left(\left(f_{\text{upper}}^{(1)}(y), f_{\theta_2}^{(2)}(y)\right), x\right)\left(y_1 - f_{\text{lower}}^{(1)}(y)\right) \\ &\quad f_{\text{upper}}^{(1)}(y) - f_{\text{lower}}^{(1)}(y)\end{aligned}\tag{4.9}$$

for $i \in \{\text{lower}, \text{upper}\}$. We proceed by interpolating in the second dimension and obtain

$$\widehat{c}_k(y, x) := \frac{\widehat{c}_k\left(\left(y_1, f_{\text{lower}}^{(2)}(y)\right), x\right)\left(f_{\text{upper}}^{(2)}(y) - y_2\right) + \widehat{c}_k\left(\left(y_1, f_{\text{upper}}^{(2)}(y)\right), x\right)\left(y_2 - f_{\text{lower}}^{(2)}(y)\right)}{f_{\text{upper}}^{(2)}(y) - f_{\text{lower}}^{(2)}(y)}$$

for all $(y, x) \in \mathcal{Y} \times \mathcal{X}$. By inserting (4.9) into the above equation, a simple calculation yields

$$\widehat{c}_k(y, x) = \sum_{\theta_1, \theta_2 \in \{\text{lower}, \text{upper}\}} \frac{(f_{\theta_1}^{(1)}(y) - y_1)(f_{\theta_2}^{(2)}(y) - y_2)s(\theta_1, \theta_2)}{(f_{\text{upper}}^{(1)}(y) - f_{\text{lower}}^{(1)}(y))(f_{\text{upper}}^{(2)}(y) - f_{\text{lower}}^{(2)}(y))} \widehat{c}_k(f_{\theta_1, \theta_2}(y), x),$$

where $s(\theta_1, \theta_2) := \mathbb{1}_{\theta_1=\theta_2} - \mathbb{1}_{\theta_1 \neq \theta_2}$. Note that we would arrive at the same result if the interpolation is done first along the second dimension and then along the first dimension. For a treatment of the multidimensional case, we refer the reader to [WZ88].

In conclusion, the least squares Monte-Carlo approach described in this section can be summarized in the following algorithm.

Algorithm 1

```

simulate  $M \in \mathbb{N}$  sample paths  $\left(X_0^{(m)}, X_1^{(m)}, \dots, X_N^{(m)}\right)_{m=1}^M$ 
 $\widehat{v}_N \leftarrow g_N$ 
for  $k = N - 1, \dots, 0$  do
  for  $y \in \tilde{\mathcal{Y}}$  do
     $\beta_{k,1}^{(y)}, \dots, \beta_{k,p}^{(y)} \leftarrow \arg \min_{\beta_1, \dots, \beta_p} \sum_{m=1}^M \left| \widehat{v}_{k+1}\left(y, X_{k+1}^{(m)}\right) - \sum_{j=1}^p \beta_j \psi_j\left(X_k^{(m)}\right) \right|^2$ 
     $\widehat{c}_k(y, x) \leftarrow \sum_{j=1}^p \beta_{k,j}^{(y)} \psi_j(x)$ 
  end for
   $\widehat{c}_k(y, x) \leftarrow f_{\text{interpolate}}\left(y; \widehat{c}_k(\tilde{y}, x) : \tilde{y} \in \tilde{\mathcal{Y}}\right)$ 
   $\widehat{v}_k(y, x) \leftarrow \max_{a \in \mathcal{A}(y, x)} \{r_k(a, y, x) + \widehat{c}_k(\varphi_{k+1}(a, y), x)\}$ 
end for

```

Remark 4.20. (i) Notice that in the explicit solution (4.6) of the least squares problem (4.5) the matrix computation $\mathbf{F}_k := (F_k^\top F_k)^{-1} F_k^\top$ is not dependent on the control

y . This allows an optimization of the above algorithm by first precomputing \mathbf{F}_k . Then, the regression coefficients can be efficiently computed by $\beta_k^{(y)} = \mathbf{F}_k \widehat{V}_{k+1}^{(y)}$ for every $y \in \tilde{\mathcal{Y}}$.

- (ii) It is crucial to highlight that when the update rules satisfy $\varphi_{k+1}(y, a) \in \tilde{\mathcal{Y}}$ for all $y \in \tilde{\mathcal{Y}}$ and $a \in \bigcup_{x \in \mathcal{X}} \mathcal{A}_k(y, x)$, the interpolation step in Algorithm 1 becomes unnecessary and can be omitted. Many examples, such as optimal stopping problems, naturally meet this condition. Moreover, if we assume finitely many possible initial controls, we can ensure that this condition holds. Namely, due to the finite number of time steps and actions, only finitely many controls can be observed in total. However, it is worth noting that in general, the resulting discretization set $\tilde{\mathcal{Y}}$ grows exponentially with the number of time steps N , leading to a point where employing the interpolation method becomes more practical.

Having approximated the continuation functions, we immediately obtain an (estimated) optimal policy. Namely, by Theorem 4.16, an optimal policy is given by a sequence of maximizers of the value functions. Thus, $\widehat{\pi} := (\widehat{\pi}_0, \dots, \widehat{\pi}_{N-1}) \in \mathcal{MP}_0$ defined by

$$\widehat{\pi}_k(y, x) := \arg \max_{a \in \mathcal{A}_k(y, x)} \{r_k(a, y, x) + \widehat{c}_k(\varphi_{k+1}(y, a), x)\} \quad (4.10)$$

for $(y, x) \in \mathcal{Y} \times \mathcal{X}$ is our estimated optimal policy.

Remark 4.21. We can compute an estimate of the value of the approximated optimal policy (4.10) using a Monte Carlo simulation as follows. First, we sample another set of independent trajectories $(\tilde{X}_k^{(m)})_{k=0}^N$, $m = 1, \dots, M_{\text{test}}$ for $M_{\text{test}} \in \mathbb{N}$ from the underlying state process X with initial value $\tilde{X}_0 \sim X_0$. It is important that this new set of trajectories is independent of the samples $(X_k^{(m)})_{k=0}^N$, $m = 1, \dots, M$ used in the regression procedure. Then, for each $m = 1, \dots, M_{\text{test}}$, the corresponding control process $Y^{\widehat{\pi},(m)} = (Y_k^{\widehat{\pi},(m)})_{k=0}^N$ with initial control $y_0 \in \mathcal{Y}$ is given by

$$Y_0^{\widehat{\pi},(m)} = y_0 \text{ and } Y_{k+1}^{\widehat{\pi},(m)} = \varphi_{k+1} \left(Y_k^{\widehat{\pi},(m)}, \widehat{\pi}_k \left(Y_k^{\widehat{\pi},(m)}, \tilde{X}_k^{(m)} \right) \right) \text{ for } k = 0, \dots, N-1.$$

Consequently, the value of the policy $\widehat{\pi}$, that is $\mathbb{E} [\mathcal{J}_0^{\widehat{\pi}}(y_0, X_0)]$, can be estimated using a Monte Carlo simulation, namely

$$\widehat{\mathcal{J}}_0^{\widehat{\pi}}(y_0, X_0) := \frac{1}{M_{\text{test}}} \sum_{m=1}^{M_{\text{test}}} \left(\sum_{k=0}^{N-1} r_k \left(Y_k^{\widehat{\pi},(m)}, \tilde{X}_k^{(m)}, \widehat{\pi}_k \left(Y_k^{\widehat{\pi},(m)}, \tilde{X}_k^{(m)} \right) \right) + g_N \left(Y_N^{\widehat{\pi},(m)}, \tilde{X}_N^{(m)} \right) \right). \quad (4.11)$$

Note that it holds $\mathcal{J}_0^{\widehat{\pi}}(y_0, x_0) \leq v_0(y_0, x_0)$ which implies $\mathbb{E} [\mathcal{J}_0^{\widehat{\pi}}(y_0, X_0)] \leq \mathbb{E} [v_0(y_0, X_0)]$. In particular, this means that $\widehat{\mathcal{J}}_0^{\widehat{\pi}}(y_0, X_0)$ provides a lower estimate for the value of interest

$\mathbb{E}[v_0(y_0, X_0)]$. For the numerical experiments in Chapter 5, we will use these estimates in order to compare different models configurations. To be more precise, the model yielding the highest estimate $\widehat{\mathcal{J}}_0^{\widehat{\pi}}(y_0, X_0)$ (up to Monte Carlo errors) is considered the best performer, as it is closest to the true value of the control problem. In contrast, the estimate based on the approximated value function \widehat{v}_0 , i.e.

$$\frac{1}{M_{\text{test}}} \sum_{m=1}^{M_{\text{test}}} \widehat{v}_0 \left(y_0, X_0^{(m)} \right),$$

can not be used for the direct comparison of the performance of different model configurations. This is because \widehat{v}_0 is an approximation of the value function v_0 , and hence may lie above or below the true value function.

4.2.1 Computational cost and convergence analysis

Let us briefly study the computational work of Algorithm 1. For this, we assume that the solution of the least squares regression problem is computed as discussed in Remark 4.20 (i), and we use (one-dimensional) linear interpolation as described in Example 4.19. To be specific, we consider the case where $\mathcal{Y} \subset \mathbb{R}$. In order to analyse the computational cost of the algorithm, we assume that certain operations are considered to be performed at constant cost. To be more precise, we denote by

- c_X the cost of simulating from the distribution of the state process X_k ,
- c_* the cost of multiplications, additions, and other primitive operations, and
- c_f the cost of evaluating the basis functions ψ_j , the one-stage reward functions r_k , the terminal reward g_N , and the update rules φ_{k+1} .

If an expression involves several such operations, then only the most expensive constant is reported. Let us now have a look at the computational cost of each significant step of Algorithm 1.

- (i) Clearly, the simulation of the trajectories $\left(X_0^{(m)}, X_1^{(m)}, \dots, X_N^{(m)} \right)_{m=1}^M$ has a computational cost of $\text{cost}_1 = c_X M(N + 1)$.
- (ii) For each $k = 0, \dots, N-1$ and $\tilde{y} \in \tilde{\mathcal{Y}}$, we need to set up the least squares problem (4.5) which requires the computation of F_k and $V_{k+1}^{(\tilde{y})}$. In total, this has a computational cost of $\text{cost}_2 = NM \left(c_f p + \text{cost} \left((\widehat{v}_{k+1}(\tilde{y}, \cdot))_{\tilde{y} \in \tilde{\mathcal{Y}}} \right) \right)$, where $\text{cost} \left((\widehat{v}_{k+1}(\tilde{y}, \cdot))_{\tilde{y} \in \tilde{\mathcal{Y}}} \right)$ denotes the cost of evaluating $\widehat{v}_{k+1}(\tilde{y}, \cdot)$ for each $\tilde{y} \in \tilde{\mathcal{Y}}$.
- (iii) For fixed $k = 0, \dots, N-1$ and $\tilde{y} \in \tilde{\mathcal{Y}}$, we solve the least squares problem with (4.6). This has a total computational cost of $\text{cost}_3 = N|\tilde{\mathcal{Y}}|c_*Mp^2$.

Thus, we are left with analysing $\text{cost} \left((\widehat{v}_{k+1}(\tilde{y}, \cdot))_{\tilde{y} \in \tilde{\mathcal{Y}}} \right)$. For the trivial case $k = N - 1$, we know that $\widehat{v}_N = g_N$, and hence we have $\text{cost} \left((\widehat{v}_N(\tilde{y}, \cdot))_{\tilde{y} \in \tilde{\mathcal{Y}}} \right) = |\tilde{\mathcal{Y}}|c_f$. However, the computation is more involved for any other $k = 0, \dots, N - 2$. For this, it is useful to first precompute $\text{cost} \left((\widehat{c}_{k+1}(\tilde{y}, \cdot))_{\tilde{y} \in \tilde{\mathcal{Y}}} \right)$, i.e. the cost of evaluating $\widehat{c}_{k+1}(\tilde{y}, \cdot)$ for each $\tilde{y} \in \tilde{\mathcal{Y}}$. By (4.4), the cost of evaluating $\widehat{c}_{k+1}(\tilde{y}, \cdot)$ is simply given by $p(c_f + c_*)$ and thus we have

$$\text{cost} \left((\widehat{c}_{k+1}(\tilde{y}, \cdot))_{\tilde{y} \in \tilde{\mathcal{Y}}} \right) = |\tilde{\mathcal{Y}}|p(c_f + c_*).$$

Let us now fix a $\tilde{y} \in \tilde{\mathcal{Y}}$ and analyse the cost of evaluating $\widehat{v}_{k+1}(\tilde{y}, \cdot)$. By (4.7), evaluating $\widehat{v}_{k+1}(\tilde{y}, \cdot)$ requires evaluating the one-stage reward functions r_{k+1} and $\widehat{c}_{k+1}(\varphi_{k+2}(\tilde{y}, a), \cdot)$ for each admissible action a , and taking the corresponding maxima. It is important to stress that in general $\varphi_{k+2}(\tilde{y}, a) \in \mathcal{Y}$ may not be contained in $\tilde{\mathcal{Y}}$. In total, we can write

$$\text{cost} (\widehat{v}_{k+1}(\tilde{y}, \cdot)) = |\mathcal{A}| (c_f + c_* + \text{cost} (\widehat{c}_{k+1}(y, \cdot))),$$

where $\text{cost} (\widehat{c}_{k+1}(y, \cdot))$ denotes the cost of evaluating $\widehat{c}_{k+1}(y, \cdot)$ for some arbitrary $y \in \mathcal{Y}$ given the already precomputed values of the continuation function for $\tilde{\mathcal{Y}}$. By assumption, $\widehat{c}_{k+1}(y, \cdot)$ is computed using the interpolation step as given by (4.8). This requires finding the two corresponding supporting grid points of y in $\tilde{\mathcal{Y}}$, computing the corresponding coefficients, and taking the appropriate weighted sum using the precomputed values $(\widehat{c}_{k+1}(\tilde{y}, \cdot))_{\tilde{y} \in \tilde{\mathcal{Y}}}$ of the continuation function. Assuming that the grid points are already sorted, the procedure of finding the two corresponding supporting grid points of y can be implemented using a binary search. Thus, we have $\text{cost} (\widehat{c}_{k+1}(y, \cdot)) = \log |\tilde{\mathcal{Y}}|c_*$. However, it is important to emphasize that $\text{cost} (\widehat{c}_{k+1}(y, \cdot))$ heavily depends on the interpolation method used. In addition, we have mainly focused on the one-dimensional setting in this cost analysis. Nevertheless, it is worth noting that in a multidimensional setting the amount of supporting grid points grows exponentially with the dimension of d_Y of $\mathcal{Y} \subset \mathbb{R}^{d_Y}$ when using the linear interpolation method. Taking all together, we obtain

$$\begin{aligned} \text{cost} \left((\widehat{v}_{k+1}(\tilde{y}, \cdot))_{\tilde{y} \in \tilde{\mathcal{Y}}} \right) &\leq \text{cost} \left((\widehat{c}_{k+1}(\tilde{y}, \cdot))_{\tilde{y} \in \tilde{\mathcal{Y}}} \right) + |\tilde{\mathcal{Y}}||\mathcal{A}| (c_f + c_* + \text{cost} (\widehat{c}_{k+1}(y, \cdot))) \\ &= |\tilde{\mathcal{Y}}|p(c_f + c_*) + |\tilde{\mathcal{Y}}||\mathcal{A}| (c_f + c_* + \log |\tilde{\mathcal{Y}}|c_*) . \end{aligned}$$

Finally, by combining all the costs together and abandoning the difference between c_* and c_f , we obtain that the computational cost of Algorithm 1 can be bounded by

$$\text{cost} \leq \text{const}MN \left(c_X + |\tilde{\mathcal{Y}}| \left(|\mathcal{A}| \log |\tilde{\mathcal{Y}}| + p^2 \right) \right), \quad (4.12)$$

where const is a positive number independent of $|\mathcal{A}|$, $|\tilde{\mathcal{Y}}|$, p , N , and M .

Let us now analyse the convergence properties of Algorithm 1. For this purpose, we need to make some simplifying assumptions. Namely, we assume that there exists a finite subset $\tilde{\mathcal{Y}}$ of \mathcal{Y} such that the update rules satisfy $\varphi_{k+1}(y, a) \in \tilde{\mathcal{Y}}$ for all $y \in \tilde{\mathcal{Y}}$, $a \in \bigcup_{x \in \mathcal{X}} \mathcal{A}_k(y, x)$ and $k = 0, \dots, N - 1$. Then, there is no need for the interpolation step in Algorithm 1. We further make the assumption that the upper bounding function b is itself bounded by some constant B_b . In particular, this implies that the value function (4.2) is uniformly bounded by NB_b for every $k = 0, \dots, N$.

Let Ψ be a linear vector space of real-valued functions defined by $\Psi = \text{span} \{ \psi_1, \dots, \psi_p \}$, where ψ_1, \dots, ψ_p are some fixed basis functions on \mathcal{X} , e.g. polynomials. In this case $\text{VC}(\Psi) \leq p$. Moreover, let T_B denote a truncation operator at level $B = NB_b$ defined by

$$T_B f(x) = \begin{cases} f(x), & |f(x)| \leq B, \\ B \text{ sign}(f(x)), & \text{otherwise.} \end{cases}$$

We now define the sequence of the value function estimates $(\hat{v}_{k,M}(y, x))_{k=0}^N$ by

$$\hat{v}_{N,M}(y, x) = g_N(y, x), \text{ and}$$

$$\hat{v}_{k,M}(y, x) = \sup_{a \in \mathcal{A}_k(y, x)} \{r_k(y, x, a) + T_B \hat{c}_k(\varphi_{k+1}(y, a), x)\}, \quad k = 0, \dots, N - 1,$$

where

$$\hat{c}_{k,M}(y, x) = T_B \left(\arg \min_{f \in \Psi} \sum_{m=1}^M \left| \hat{v} \left(y, X_{j+1}^{(m)} \right) - f \left(X_j^{(m)} \right) \right|^2 \right) \quad (4.13)$$

represents the estimate of the continuation function based on the estimated value function and on the samples

$$\mathcal{D}_{M,k} = \left\{ \left(X_k^{(m)}, X_{k+1}^{(m)} \right), \quad m = 1, \dots, M \right\}$$

from the joint distribution of (X_k, X_{k+1}) for $k = 0, \dots, N - 1$. Note that this formulation coincides with Algorithm 1 except for the additional truncation factor T_B in (4.13). In particular, the minimizer in (4.13) exists and is given by (4.4) with coefficients (4.6).

In order to state the following convergence result, we assume that the sets $\mathcal{D}_{M,k}$ are independently sampled for each k .

Theorem 4.22. *Assume that there exists some constant $\delta > 0$ such that*

$$\inf_{f \in \Psi} \mathbb{E} [|c_k(y, X_k) - f(X_k)|^2] \leq \delta^2, \quad (4.14)$$

for every $y \in \tilde{\mathcal{Y}}$. Then, for every $k = 0, \dots, N$, it holds

$$\sqrt{\mathbb{E} \left[\sup_{y \in \tilde{\mathcal{Y}}} |\hat{v}_{k,M}(y, X_k) - v_k(y, X_k)|^2 \right]} \leq \left(cB^4 \frac{1 + \log M}{M} p + \sqrt{2}\delta \right) |\tilde{\mathcal{Y}}| \frac{((1 + \sqrt{2})|\tilde{\mathcal{Y}}|)^{N-k} - 1}{(1 + \sqrt{2})|\tilde{\mathcal{Y}}| - 1},$$

where $c > 0$ is some absolute constant.

Proof. This result was proved in [Bay+22, Corollary 6.4.]. \square

Remark 4.23. (i) The left-hand side of (4.14) is often referred to as the approximation error, and measures how well the continuation function $c_k(y, \cdot)$ can be approximated by functions of Ψ in L^2 .

(ii) Note that in Algorithm 1, we rely on only one single set of independent Monte Carlo sample paths that are repeatedly reused for all time-steps. This is contrary to the assumption made for Theorem 4.22 where the sets $\mathcal{D}_{M,k}$ are independently sampled for each k . In [Zan13], the convergence of the Longstaff-Schwartz algorithm, a regression-based approach for optimal stopping problems initially proposed in [LS01], was studied under an assumption similar to the independence $\mathcal{D}_{M,k}$ assumed here. Later, in [Zan18], it was demonstrated that using only one set of independent Monte Carlo sample paths yields the convergence rates to those in [Zan13], up to certain constants. In light of this, it is reasonable to expect that similar conclusions apply in our context. Hence, we have used a single set of M trajectories for all numerical examples in Chapter 5.

4.3 Optimal battery control

Having established a suitable battery degradation model, examined several stochastic electricity price models, and developed a theoretical framework for addressing stochastic optimal control problems along with a numerical solution method, we are now prepared to present the optimal battery control problem. As a reminder, our objective is to minimize the cost of electricity for a consumer over a fixed time horizon $T > 0$ using a battery energy storage system. To formulate this problem, we use a Markov decision model as defined in Definition 4.1.

Remark 4.24 (Minimizing Cost). Up until now, we have considered the stochastic control problem as a maximization problem, and we have interpreted $(r_k)_{k=0}^{N-1}$ and g_N as one-stage rewards and terminal reward, respectively. However, one can instead also consider one-stage costs $(c_k)_{k=0}^{N-1}$ and a terminal cost h_N . In this case, the objective is to minimize the

objective functional (4.1), i.e.

$$\inf_{\pi \in \mathcal{MP}_k} \mathbb{E} \left[\sum_{\ell=k}^{N-1} c_\ell(Y_\ell^\pi, X_\ell, \pi_\ell(Y_\ell^\pi, X_\ell)) + h_N(Y_N^\pi, X_N) \mid Y_k = y, X_k = x \right], \quad (4.15)$$

for $(y, x) \in \mathcal{Y} \times \mathcal{X}$. It is easy to see that by setting $r_k(y, x, a) := -c_k(y, x, a)$ and $g_N(y, x) := -h_N(y, x)$ this problem can be transformed into an equivalent maximization problem. In particular, all the statements so far also remain valid for the minimization problem. We merely need to replace every appropriate sup and arg max operator by the inf and arg min operator, respectively. In this case, the value function v_k is also referred to as the cost-to-go function.

In light of the previous remark, we formulate the optimal battery control problem as a cost minimization problem. To stay consistent, we employ the same notation used thus far. In particular, we write r_k and g_N for the one-stage cost and the terminal cost, respectively. We proceed by discussing each component of the Markov decision model. For this, we consider a time discretization $\mathcal{T} := \{t_k = \frac{k}{N}T \mid k = 0, \dots, N\}$ for a chosen $N \in \mathbb{N}$ and time step $\Delta t := T/N$.

As previously indicated, the state process X describes the stochastic factors of the system. In our case, this is primarily the electricity price S but also other relevant stochastic variables such as forecasted wind-related variables. For this, we can choose any stochastic model that we have presented in chapter 3. For instance, the state process can consist only of the electricity price, i.e. $X_k = S_k$, or we can also include the wind forecasted values as described in Chapter 3.3, i.e. $X_k = (S_k, \bar{W}_k, \bar{W}_{k+1}, \dots, \bar{W}_{k+K^f})$, where $K^f \in \mathbb{N}$ is some forecast period parameter. For these examples, the state space is then given by $\mathcal{X} = \mathbb{R}$ or $\mathcal{X} = \mathbb{R}^{2+K^f}$, respectively.

In practice, it is reasonable to choose the number of time steps N such that the time step Δt corresponds to a one-hour interval, since this is a typical time interval that is traded in intraday markets, as discussed in section 3.1. Note that, quarter-hour intervals are also traded on intraday market, however in section 3.2.1 we have only described how to model hourly prices. In general, we can also select the number of time steps N such that the time step Δt corresponds to several one-hour intervals, which would allow more flexibility in choosing the time discretization. For instance, in order to obtain a daily electricity price, we can simply simulate hourly prices and then take the average over the 24 hourly prices. Let us now proceed by first examining the scenario without degradation.

4.3.1 Case without degradation

The control process Y represents the state of the battery over time. In the scenario without degradation, the primary attribute of the battery state is the state of charge y^{SoC} . Remember that it quantifies the capacity relative to the original maximum capacity, which means that its value remains between 0 and 1. Consequently, we define the control space as $\mathcal{Y} = [0, 1]$. In order to solve the optimal control problem numerically, we have discussed an interpolation approach in section 4.2. For this approach, we require a discretization of the control space $\tilde{\mathcal{Y}} \subset \mathcal{Y}$. A natural choice is to consider a regular grid, i.e.

$$\tilde{\mathcal{Y}} = \{\ell/K^{\text{SoC}} \mid \ell = 0, \dots, K^{\text{SoC}}\} \subset [0, 1] \quad (4.16)$$

where $K^{\text{SoC}} \in \mathbb{N}$ is some discretization parameter.

At each point in time t_k , we must decide how much to charge or discharge the battery. This is achieved by adjusting the C-rate C_k , which we assume to be constant between the time steps. Hence, the C-rate is interpreted as the action in the Markov decision model. Up to this point, a vital factor that we have omitted is the electricity demand of the consumer. For this, we assume that the energy demand of the consumer can be described by a demand process $(D_k)_{k=0}^{N-1}$, where D_k denotes electricity current required between time t_k and t_{k+1} relative to the maximum battery capacity, similar to the C-rate. In other words, to satisfy the demand using the battery, we require a discharge C-rate of $C_k = -D_k$. For instance, if the demand is constant at $D_k \equiv 0.5$, a C-rate of $C_k = 0.5$ is necessary. In what follows, we assume that the demand process $(D_k)_{k=0}^{N-1}$ is deterministic. To satisfy the electricity demand, we have two options: utilizing the battery or purchasing electricity from the market. Let I_S be a normalization constant such that S_k/I_S denotes the price of electricity for 1 Ah. If we decide to buy electricity from the market, the amount of energy bought between time t_k and t_{k+1} , measured in Ah-throughput, is given by $I_k^{\text{buy}}\Delta t$, where I_k^{buy} represents the electricity current in that interval. Then, the cost of acquiring this energy is given by $(S_k/I_S)I_k^{\text{buy}}\Delta t$. To compare I_k^{buy} with the C-rate C_k and the demand D_k , we define $C_k^{\text{buy}} := I_k^{\text{buy}}/I_C$ where I_C is the normalization constant as defined in section 2.2.2. Then, it holds $D_k = C_k^{\text{buy}} - C_k$, i.e. the consumed energy must be supplied by the battery or purchased from the energy market, or a combination of both. In particular, for a given C-rate C_k and demand D_k , we have $C_k^{\text{buy}} = D_k + C_k$. Note that, by slight abuse of notation, we assume that the constants I_C and I_S , and the variables I_k^{buy} are all dimensionless.

In practice, every battery has a maximum charge and discharge current. Hence, in our setting the C-rate should be bounded from below and above, i.e. $C^{\min} \leq C_k \leq C^{\max}$. For simplicity, we assume $C^{\min} = -C^{\max}$ and thus $C_k \in [-C^{\max}, C^{\max}]$. In particular,

the interval $[-C^{\max}, C^{\max}]$ is a reasonable choice for the action space. However, in our definition of the Markov decision model, we assume that the action space \mathcal{A} is finite, which is why we will consider a suitable discretization instead. Namely, let

$$\mathcal{A} = \left\{ \frac{\ell}{K^{\text{SoC}} \Delta t} \mid \ell = -K^{\text{SoC}}, \dots, -1, 0, 1, \dots, K^{\text{SoC}} \right\} \cap [-C^{\max}, C^{\max}]. \quad (4.17)$$

Notably, this implies that the highest allowed (absolute) C-rate is $\frac{1}{\Delta t} \vee C^{\max}$. Moreover, we define the set of admissible actions by

$$\begin{aligned} \mathcal{A}_k(y^{\text{SoC}}, x) &= \{C_k \in \mathcal{A} \mid 0 \leq y^{\text{SoC}} + C_k \Delta t \leq 1\} \\ &= \{C_k \in \mathcal{A} \mid -y^{\text{SoC}}/\Delta t \leq C_k \leq (1 - y^{\text{SoC}})/\Delta t\} \end{aligned} \quad (4.18)$$

for all $(y^{\text{SoC}}, x) \in \mathcal{Y} \times \mathcal{X}$. This restriction ensures that the battery is not discharged beyond its available capacity and is not charged beyond its maximum capacity allowance.

Remark 4.25. Note that (4.18) implicitly assumes that the consumer is allowed to sell any excess energy back to the market. It is worth noting that we could also impose the restriction that the consumer is only allowed to purchase electricity from the market, but is not able to sell any excess energy. This assumption can be described by the condition $0 \leq C_k^{\text{buy}}$, which is equivalent to $-D_k \leq C_k$. Incorporating this into the admissible action set (4.18) yields

$$\begin{aligned} \mathcal{A}_k(y^{\text{SoC}}, x) &= \{C_k \in \mathcal{A} \mid 0 \leq y^{\text{SoC}} + C_k \Delta t \leq 1 \text{ and } -D_k \leq C_k\} \\ &= \{C_k \in \mathcal{A} \mid \max\{-y^{\text{SoC}}/\Delta t, -D_k\} \leq C_k \leq (1 - y^{\text{SoC}})/\Delta t\} \end{aligned} \quad (4.19)$$

for all $(y^{\text{SoC}}, x) \in \mathcal{Y} \times \mathcal{X}$. However, in practice, the demand D_k is typically much smaller than the highest allowed (absolute) C-rate $\frac{1}{\Delta t} \vee C^{\max}$. This means that in order to flexibly discharge the battery to satisfy the demand D_k , we require an extremely fine discretization K^{SoC} .

Note that, by (2.9), the update rule is simply given by

$$\varphi_{k+1}(y, a) = y + a \Delta t, \quad a \in \mathcal{A}_k(y, x), (y, x) \in \mathcal{Y} \times \mathcal{X}$$

for all $k = 0, \dots, N - 1$. In particular, it is easy to see that for any $a \in \mathcal{A}$ we even have $\varphi_{k+1}(y, a) \in \tilde{\mathcal{Y}}$ if $y \in \tilde{\mathcal{Y}}$. In other words, if $y_0^{\text{SoC}} \in \tilde{\mathcal{Y}}$, i.e. the initial control is in the discretized control set, then the entire control process $(Y_k^\pi)_{k=0}^N$, as defined in Remark 4.5, satisfies $Y_k^\pi \in \tilde{\mathcal{Y}}$ where $\pi \in \mathcal{MP}_0$ is some Markov policy. This property is highly desirable, as it eliminates the need for an interpolation step in the numerical Algorithm 1 as discussed in Remark 4.20 (ii).

We are left with specifying the one-stage cost functions r_k and the terminal cost g_N . For simplicity, we assume that there is no terminal cost, i.e. $g_N \equiv 0$. We define the cost function in the no degradation case by

$$r_k(Y_k, X_k, C_k) := \frac{S_k}{I_S} (D_k + C_k) I_C \Delta t + a_r |C_k| I_C \Delta t$$

for all $k = 0, \dots, N - 1$. The first term represents the cost of purchasing electricity from the market. Alternatively, we could have also considered discounted prices, i.e. $\bar{S}_k = e^{-rt_k} S_k$ instead of S_k where $r > 0$ denotes the interest rate. The second term models the running cost of utilizing the battery, such as the cost of cooling the battery to maintain a stable temperature. This cost is modelled by penalizing battery usage as measured in Ah-throughput. Here, $a_r > 0$ denotes some running cost parameter. Note that in this specific setting, where battery degradation is not considered, the one-stage cost function (4.3.1) does not depend on the control Y_k .

4.3.2 Case with degradation

Let us now turn our attention to the case involving degradation. In our previous discussion, see section 2.2.4, we explored multiple degradation models, with a focus on model **(A2)** in particular. In order to compute the degradation state η_k , we need to incorporate additional variables into the control process, specifically the normalized Ah-throughput y_k^{tot} and the average C-rate C_k^{av} . It is worth noting that y_k^{tot} can be expressed as the product of C_k^{av} and the total active battery usage time T_k^{ac} for all $k = 0, \dots, N - 1$. Consequently, we can equivalently include T_k^{ac} in our control process instead of y_k^{tot} , as both variables provide the same information when combined with C_k^{av} . Therefore, we define the control process as $Y_k = (y_k^{\text{SoC}}, T_k^{\text{ac}}, C_k^{\text{av}}, \eta_k)^{\top}$. Note that we have

$$0 \leq T_k^{\text{ac}} = \sum_{\ell=0}^{k-1} \mathbb{1}_{|C_\ell|>0} \Delta t \leq k \Delta t \leq N \Delta t = T \quad \text{and} \quad 0 \leq C_k^{\text{av}} \leq \frac{1}{\Delta t} \vee C^{\max} \quad (4.20)$$

for all $k = 0, \dots, N - 1$. In particular, the control set is given by

$$\mathcal{Y} = [0, 1] \times [0, T] \times [0, (1/\Delta t) \vee C^{\max}] \times [0, 1].$$

Recall that for the interpolation approach, we require a finite discretization of the control set $\tilde{\mathcal{Y}} \subset \mathcal{Y}$. For this, we simply consider a multidimensional regular grid, similarly to (4.16), with discretization parameters $K^{\text{SoC}}, K^{\text{ac}}, K^{\text{av}}, K^\eta \in \mathbb{N}$ for each dimension,

respectively. As discussed in Section 2.2.4, the battery variables are updated by

$$Y_{k+1} = \begin{pmatrix} y_{k+1}^{\text{SoC}} \\ T_{k+1}^{\text{ac}} \\ C_{k+1}^{\text{av}} \\ \eta_{k+1} \end{pmatrix} = \begin{pmatrix} y_k^{\text{SoC}} + C_k \Delta t \\ T_k^{\text{ac}} + \mathbb{1}_{|C_k| > 0} \Delta t \\ C_k^{\text{av}} + \frac{|C_k| \Delta t - C_k^{\text{av}} \mathbb{1}_{|C_k| > 0} \Delta t}{T_k^{\text{ac}} + \mathbb{1}_{|C_k| > 0} \Delta t} \\ \eta_k + \tilde{\eta}_{\mathbf{A2}}(C_k^{\text{av}} \cdot T_k^{\text{ac}}, C_k^{\text{av}}, |C_k|) \end{pmatrix} =: \varphi_{k+1}(Y_k, C_k),$$

where $Y_k = (y_k^{\text{SoC}}, T_k^{\text{ac}}, C_k^{\text{av}}, \eta_k)^\top$, for all $k = 0, \dots, N-1$. In particular, this specifies the update rule φ_{k+1} . In the case with degradation, the set of admissible actions (4.18) needs to be slightly adjusted, namely we have

$$\begin{aligned} \mathcal{A}_k(y^{\text{SoC}}, x) &= \{C_k \in \mathcal{A} \mid 0 \leq y^{\text{SoC}} + C_k \Delta t \leq \eta_k\} \\ &= \{C_k \in \mathcal{A} \mid -y^{\text{SoC}}/\Delta t \leq C_k \leq (\eta_k - y^{\text{SoC}})/\Delta t\} \end{aligned} \quad (4.21)$$

for all $(y^{\text{SoC}}, x) \in \mathcal{Y} \times \mathcal{X}$. This formulation ensures that the battery can not be charged beyond its degraded maximum capacity given by η_k . In case that we do not allow selling electricity back to the market, the corresponding admissible action set is given by

$$\begin{aligned} \mathcal{A}_k(y^{\text{SoC}}, x) &= \{C_k \in \mathcal{A} \mid 0 \leq y^{\text{SoC}} + C_k \Delta t \leq \eta_k \text{ and } -D_k \leq C_k\} \\ &= \{C_k \in \mathcal{A} \mid \max\{-y^{\text{SoC}}/\Delta t, -D_k\} \leq C_k \leq (\eta_k - y^{\text{SoC}})/\Delta t\} \end{aligned}$$

for all $(y^{\text{SoC}}, x) \in \mathcal{Y} \times \mathcal{X}$. Note that, in contrast to the no degradation case, it is no longer guaranteed that $\varphi_{k+1}(y, a) \in \tilde{\mathcal{Y}}$ if $y \in \tilde{\mathcal{Y}}$. As a result, the interpolation step in Algorithm 1 becomes necessary to fully approximate the continuation function as described in the section 4.2.

Similarly to the no degradation setting, we assume that there is no terminal cost, i.e. $g_N \equiv 0$. However, we need to introduce the degradation cost as an additional component to the one-stage cost function r_k . To incorporate this, we assume that the value of a battery is proportional to its (remaining) maximum capacity. To be more precise, the price of a battery with maximum capacity Q^{\max} is given by $U = a_U Q^{\max}$ where $a_U > 0$ is a parameter representing the value of 1 Ah of maximum battery capacity. Then, the initial value of our battery with an initial maximum capacity Q_0^{\max} is denoted as $U_0 = a_U Q_0^{\max}$, and its value at time t_k is given by $U_k = a_U Q_k^{\max} = a_U Q_0^{\max} \eta_k = U_0 \eta_k$. It is now easy to see that the loss of battery value due to degradation at each time step is described by $U_0 \tilde{\eta}_{\mathbf{A2}}(y_k^{\text{tot}}, C_k^{\text{av}}, |C_k|)$ for model **(A2)**. With these considerations, the one-stage cost function in the case with degradation is defined as

$$r_k(Y_k, X_k, C_k) := S_k(D_k + C_k) I_C \Delta t + a_R |C_k| I_C \Delta t - U_0 \tilde{\eta}_{\mathbf{A2}}(C_k^{\text{av}} \cdot T_k^{\text{ac}}, C_k^{\text{av}}, |C_k|) \quad (4.22)$$

for $k = 0, \dots, N - 1$. The optimal battery control problem is then given by (4.15) and we can apply the methods from the section 4.2 in order to solve the problem numerically.

It is important to note that the above formulation allows for the consideration of different degradation models beyond **(A2)** such as model **(A1)** and model **(B)**. In fact, this setup generalizes the no degradation case by simply replacing $\tilde{\eta}_{\mathbf{A2}}$ with $\tilde{\eta}_0 \equiv 0$. In what follows, we refer to the no degradation model as model **(0)**.

Remark 4.26 (Numerical considerations). (i) It is easily seen that the optimal control problem satisfies Assumption 4.7. Note, however, that for the convergence result in section 4.2.1, we assumed that the bounding function b itself is bounded by some constant. In general, this is not satisfied in our setup. Nevertheless, it is worth pointing out that electricity prices in real markets are inherently bounded from both below and above.

(ii) It is worth noting that a more efficient choice for the discretization set $\tilde{\mathcal{Y}}$ is possible. Observing (4.20), we find that the total battery usage time T_k^{ac} is bounded by $t_k = k\Delta t$ for all $k = 0, \dots, N - 1$. In our current approach, however, we have chosen $\tilde{\mathcal{Y}}$ as a multidimensional regular grid. In particular, the regular grid discretization in the T^{ac} component is chosen over the entire interval $[0, T]$. As a result, there exist controls $y \in \tilde{\mathcal{Y}}$ that cannot be reached by any control process before a certain time. This leads to unnecessary computations when we calculate the continuation function for every $y \in \tilde{\mathcal{Y}}$. To address this issue, we can utilize a normalized version of the total battery usage time, denoted as $\overline{T}_k^{\text{ac}} := T_k^{\text{ac}} / (k\Delta t) \in [0, 1]$, and leverage the time dependency of the update rule φ_{k+1} and cost function r_k to transform the normalized version back to the actual usage time. For the update rule φ_{k+1} , a simple calculation shows

$$\overline{T}_{k+1}^{\text{ac}} = \frac{T_{k+1}^{\text{ac}}}{(k+1)\Delta t} = \frac{T_k^{\text{ac}} + \mathbb{1}_{|C_k|>0}\Delta t}{(k+1)\Delta t} = \frac{k}{k+1}\overline{T}_k^{\text{ac}} + \frac{1}{k+1}\mathbb{1}_{|C_k|>0}\Delta t,$$

for all $k = 0, \dots, N - 1$. And the one-stage cost function (4.22) is instead formulated as

$$r_k(Y_k, X_k, C_k) = S_k C_k^{\text{buy}} I_C \Delta t + a_R |C_k| I_C \Delta t - U_0 \tilde{\eta}_{\mathbf{A2}} (C_k^{\text{av}} \cdot (\overline{T}_k^{\text{ac}} k \Delta t), C_k^{\text{av}}, |C_k|)$$

where $Y_k = (y_k^{\text{SoC}}, \overline{T}_k^{\text{ac}}, C_k^{\text{av}}, \eta_k)^{\top}$ for every $k = 0, \dots, N - 1$. To summarize, this approach helps avoid unnecessary computations and improves the interpolation by ensuring that the discretization set is more efficiently utilized.

5 Numerical Results

In this chapter, we study the optimal battery control problem numerically. For this purpose, we have implemented Algorithm 1 in Python. We consider various scenarios and explore the impact of different settings on the optimal control strategy. Specifically, we investigate the behaviour of the control strategy under varying battery parameter configurations and analyse the effects of degradation on its performance. We further analyse the effect of incorporating wind forecasts to the state process, as well as the influence of seasonality on the optimal control strategy. By conducting a comprehensive analysis of these diverse settings, our numerical experiments aim to provide valuable insights into the optimal battery control problem, and our findings will contribute to a comprehensive understanding of the proposed control framework's capabilities and limitations.

5.1 Degradation experiments

Throughout this section, we model the electricity price S_k by adopting the stochastic multifactor model described in Example 3.14. In particular, the state process only consists of the electricity price, i.e. $X_k = S_k$. In our analysis, we will consider price model parameters estimated by different authors. However, since these parameters are typically fitted to daily data, we need to carefully adapt them to the hourly time index setting. Namely, the (dimensionless) time for the battery degradation models in section 2.2 is stated relative to 1 h. In comparison, the time index for the stochastic electricity price models is usually relative to 24 h. Thus, a conversion is essential to ensure consistency and accuracy when incorporating the price model into our optimal battery control framework.

The seasonality component Λ is modelled by (3.11). However, (3.11) is stated in setting of a daily time index, and thus we need to slightly adapt the function by replacing the yearly seasonality term $\cos\left(\frac{c_3+2\pi t}{365}\right)$ by $\cos\left(\frac{c_3+2\pi t}{365 \cdot 24}\right)$. For the parameters we select the values $c_0 = 50.575, c_1 = -0.0134/24, c_2 = -3.027, c_3 = 8328 \cdot 24$, and refer to [VV14, Table 1] for the hourly parameters. The parameters c_0, c_1, c_2, c_3 governing the level, trend and yearly seasonality are largely inspired by [RVG21]. The values for the hourly parameters $c_{\text{Monday}}^h, \dots, c_{\text{Sunday}}^h, h = 1, \dots, 24$ are taken from [VV14]. In this study, the authors obtained these values by first removing the trend, followed by fitting the trimmed means for every hour and weekday. Note that we have adapted the parameters for the trend c_1 and the seasonality shift c_3 to fit the hourly setting, and the level c_0 to fit the specific hourly parameters from [VV14, Table 1].

Recall that the non-stationary long-term factor Z^{long} is modelled as a Lévy process. In particular, [RVG21] fit several distributions from the class of generalized hyperbolic distributions, see Remark 3.9, and find that the symmetric normal inverse Gaussian distribution

fits the data well with parameters $(\lambda = -0.5, \bar{\alpha} = 0.236, \mu = -0.005, \Sigma = 0.385, \gamma = 0)$. Since these parameters are estimated for daily increments, we assume that they correspond to the distribution of Z_{24}^{long} .

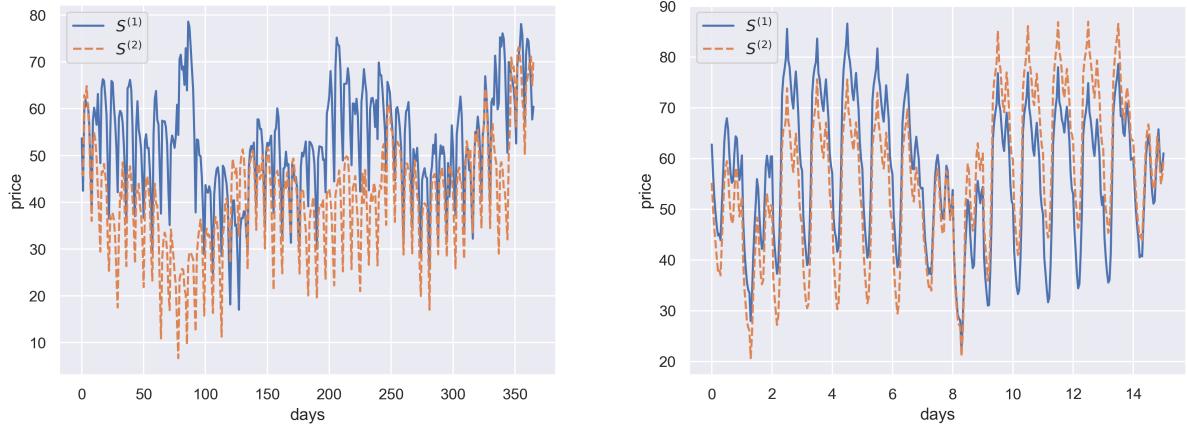
Remark 5.1. Note that when one tries to estimate the parameters of the generalized hyperbolic distribution in practice, the so-called $(\lambda, \bar{\alpha}, \mu, \Sigma, \gamma)$ -parametrization is preferred due to its ability to avoid identifiability issues. [WBL22] show that the $(\lambda, \chi, \psi, \mu, \Sigma, \gamma)$ -parametrization, as described in Remark 3.9, can be obtained from the $(\lambda, \bar{\alpha}, \mu, \Sigma, \gamma)$ -parametrization by setting

$$\psi = \bar{\alpha} \frac{K_{\lambda+1}(\bar{\alpha})}{K_\lambda(\bar{\alpha})}, \quad \chi = \frac{\bar{\alpha}^2}{\psi} = \bar{\alpha} \frac{K_\lambda(\bar{\alpha})}{K_{\lambda+1}(\bar{\alpha})},$$

where K_λ denotes the modified Bessel function of the third kind and λ, μ and γ remain the same. Note that whenever $\lambda = -0.5$, which corresponds to the normal inverse Gaussian distribution, we have that $\psi = \bar{\alpha}$ and $\chi = \bar{\alpha}$. This is because of the symmetry of the Bessel function. For a more comprehensive review of these parameterizations, we refer readers to [WBL22].

Furthermore, we model the short-term factor as a Lévy-driven CARMA(2, 1) process, see Example 3.13. For the kernel parameters, we choose $\alpha_1 = 0.0807$, $\alpha_2 = -0.9193$, $\lambda_1 = 0.0650$, $\lambda_2 = 0.7424$, which are the values given in [RVG21] obtained by the estimation algorithm described in [GKM11]. In addition, they fit 11 different generalized hyperbolic distributions to the increments of the driving Lévy process and find that the best fit is given by an asymmetric Student's-t distribution. However, for the sake of convenience, we assume that the Lévy process follows an asymmetric normal inverse Gaussian distribution, which provides a slightly worse fit but possesses the desirable property of being convolution closed, as mentioned in section 3.2.2. The specific parameters for this distribution, as provided by [RVG21, Table 1], are given by $(\lambda = -0.5, \bar{\alpha} = 0.962, \mu = 1.310, \Sigma = 7.606, \gamma = -1.342)$. Similarly to the short-term factor, these parameters describe the daily increments, i.e. L_{24} where L is background driving Lévy process.

Figure 1 shows sample paths of the electricity price process S with the above parameters. On the left, we have two independently sampled trajectories with a time horizon of 365 days and daily frequency, that is, $T = 365 \cdot 24$ and $N = 365$. The distinctive drops in prices seen in the left graph can be attributed to lower weekend prices, a characteristic pattern often observed in electricity markets. Moving to the right, we have two additional independent sample paths shown with a time horizon of 15 days and hourly frequency, i.e. $T = N = 15 \cdot 24$. In this case, a pronounced daily pattern emerges, with prices noticeably lower during nighttime hours and rising during daytime hours, which reflects the impact of demand and supply dynamics on electricity prices throughout the day.



(a) Time horizon of 365 days and daily frequency ($T = 365 \cdot 24, N = 365$)

(b) Time horizon of 15 days and hourly frequency ($T = N = 15 \cdot 24$)

Figure 1: Two sample paths of the electricity price $(S_k)_{k=0}^N$.

Before we investigate the effect of the degradation on the optimal control strategy, we initially focus on a scenario without battery degradation, that is we consider $\tilde{\eta}_0 \equiv 0$, and assume no running cost, i.e. $a_R = 0$. The electricity prices on the EEX are denoted per MWh and since we assume a constant voltage of 1 V, this results in $I_S = 10^6$. Regarding the battery parameters, we assume that the battery has a maximum capacity Q^{\max} of 10 000 Ah and hence we have $I_C = 10^4$. Moreover, we assume that the battery has a maximum C-rate of 1. Additionally, we assume a constant electricity demand from the consumer, and we denote by D^{day} the daily demand relative to the battery size Q^{\max} . The demand process is then simply given by $D_k \equiv D^{\text{day}} \Delta t / 24$ for all $k = 0 \dots, N - 1$. For instance, for $D^{\text{day}} = 1$, a fully charged battery can satisfy the demand of the consumer for exactly one day. For the basis functions, we choose the set $\text{Poly}_2(\mathcal{X})$ from Example 4.17, i.e. polynomials up to the degree 2.

We further consider a time horizon of 30 days, i.e. $T = 30 \cdot 24$. In order to obtain a time step of one hour, we can simply choose $N = T$. It is evident that a finer time discretization would yield improved performance for the corresponding optimal control strategy. Indeed, this can be observed in Table 1 where we have considered different time discretizations ranging from 1 hour to 24 hours for the time step Δt . To compare the performances, we have computed the estimate $\widehat{\mathcal{J}}_0^\pi(y_0, X_0)$ for each model as described in Remark 4.21. Since the optimal control problem is a cost minimization problem, $\widehat{\mathcal{J}}_0^\pi(y_0, X_0)$ actually represents an upper estimate of $\mathbb{E}[v_0(y_0, X_0)]$. For reference, we have also included a benchmark value. For this benchmark, we have assumed that the consumer has no battery, and thus has to satisfy the electricity demand exclusively by purchasing electricity from the market. Notably, we observe that limiting the C-rate

Δt	$\widehat{\mathcal{J}}_0^{\widehat{\pi}}(y_0, X_0)$
benchmark	15.8285 (0.0051)
24h	15.0503 (0.0050)
12h	11.8936 (0.0051)
6h	8.7794 (0.0051)
4h	7.1844 (0.0051)
2h	3.3034 (0.0051)
1h	1.2855 (0.0051)

Table 1: Comparison of upper estimates $\widehat{\mathcal{J}}_0^{\widehat{\pi}}(y_0, X_0)$ for different time discretizations. The value in the brackets denotes the $\pm 99.7\%$ Monte-Carlo error. For the daily demand, we assume $D^{\text{day}} = 1$ and for the initial control, we set $y_0 = 0$, i.e. we start with an empty battery. We have used $M = 10^4$ training sample paths and $M_{\text{test}} = 10^6$ paths for calculating $\widehat{\mathcal{J}}_0^{\widehat{\pi}}(y_0, X_0)$.

changes to once per day significantly constrains the model's performance, resulting in only marginal improvement over the benchmark. However, adopting a finer time discretization proves to be much more beneficial, leading to substantial cost savings. This can be explained by the high day-to-day price variations, as evident in Figure 1. The finer time discretization allows the model to seize opportunities and leverage these price fluctuations. In contrast, the model with a time discretization of 24 hours can only buy electricity at the average price throughout the day, missing out on these advantageous moments. This indicates that it would make sense to consider even finer time discretizations such as a quarter hourly time discretization, as these are the shortest contracts traded on intraday markets. However, in our case, our electricity price models only allow for hourly modelling and hence, in what follows, we proceed by choosing $N \in \mathbb{N}$ such that each time step corresponds to one hour, i.e. $\Delta t = 1$.

It is worth noting that for the experiments in Table 1, changing the discretization parameter K^{SoC} for the state of charge had no effect on the performance of the model. To be more precise, the optimal policy only fully charges or discharges the battery within a time step such that the only state of charge levels observed are 0 and 1. It turns out that many general optimal control problems with even continuous action and control spaces have solutions of this bang-bang type. That is, all optimal actions are taken from a finite set, usually at the boundaries of the (continuous) action sets, see also [Bay+22, Remark 2.1]. As a result, the fact that we have assumed a finite action set is no restriction for the optimal control problem in the setting without degradation.

Moreover, in Table 1 we have assumed that the consumer is allowed to sell excess energy from the battery back to the market. In view of Remark 4.25, let us also investigate the impact on the optimal control strategy if selling to the market is not permitted. In our model, this limitation is reflected by the restricted admissible action set (4.19). Note that the smallest discharge rate in the action set (4.17) is given by $1 / (K^{\text{SoC}} \Delta t)$. In particular,

by (4.19), this means that a discharge of the battery is only admissible if

$$-\frac{1}{K^{\text{SoC}} \Delta t} \geq -D_k \equiv -\frac{D^{\text{day}}}{24} \Delta t, \quad \text{which implies } \frac{24}{D^{\text{day}} \Delta t^2} \leq K^{\text{SoC}}.$$

In other words, the smaller the daily demand D^{day} , the finer the discretization parameter K^{SoC} has to be chosen. To facilitate a comparison of the optimal strategies for different daily demands, we fix K^{SoC} and select daily demand levels that are large enough such that the above condition holds. Figure 2 visually illustrates the (approximated) optimal control strategy, given by (4.10), for different daily demand scenarios. Notably,

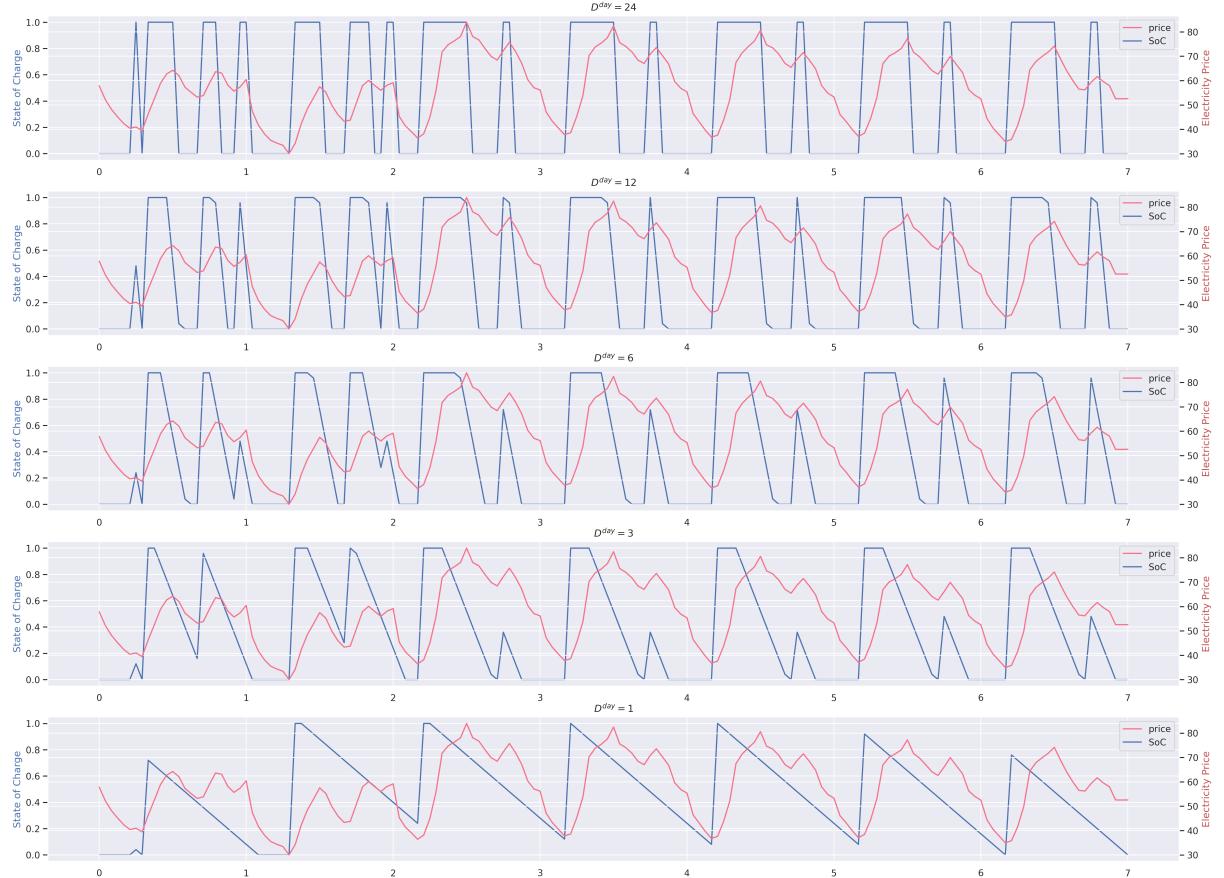


Figure 2: Comparison of optimal control strategies for one sample path for different daily demands $D^{\text{day}} = 24, 12, 6, 3, 1$ with a time horizon of 7 days and initial control $y_0 = 0$. We have used $M = 10^4$ training sample paths for computing the control strategy $\hat{\pi}$.

a straightforward calculation reveals that when the daily demand, D^{day} , is equal to 24, the hourly demand precisely matches the battery size, allowing the battery to be fully discharged within a single time step to meet the demand. As a consequence, the optimal control strategy in this case coincides with the setting where selling back to the market is allowed, resulting in the observation of bang-bang controls. In contrast, for other daily demand choices, we observe significantly lower discharge rates, as they are constrained by

the hourly demand. Interestingly, the optimal strategy predominantly leverages the daily variations in electricity prices, effectively adapting to fluctuations in the demand pattern.

Let us now delve into the impact of degradation on the optimal battery control problem by considering the battery degradation model **(A2)**. For this, we fix the degradation parameter $a_{\text{deg}} = 0.0004$ and the initial price of the battery by $U_0 = 2500$. The presence of degradation introduces regularization to the optimal control problem, which ideally leads to control strategies that are not solely of the bang-bang type. To study the degradation effect more comprehensively, we thus assume that selling energy back to the market is allowed, as this avoids the restriction on the discharge C-rate, as observed in our previous experiment. Thus, in what follows, we set $D^{\text{day}} = 1$ and consider (4.18). It is essential to emphasize that including degradation in the model has two primary effects on the optimal control problem: the loss of battery value and the limitation of the control strategy due to the maximum capacity constraint, see (4.21). For our analysis, we primarily focus on short time horizons, where the effect of the loss of maximum capacity is minimal. As a result, we make the approximation that there is no loss of maximum capacity, that is, we consider (4.18) instead of (4.21) for the admissible action set. This approximation offers an advantage in terms of computation, as it eliminates the need to track the degradation state η_k in the control process Y_k while computing the one-stage cost function r_k . This computational simplification makes the problem more manageable. Table 2 compares the performance of the model for different discretization parameters for K^{SoC} , K^{tot} and K^{ac} . We observe that choosing a sufficiently fine discretization is crucial

K^{SoC}	K^{av}	K^{ac}	$\widehat{\mathcal{J}}_0^{\pi}(y_0, X_0)$
10	5	5	11.1266 (0.0161)
10	10	5	10.6499 (0.0161)
10	10	10	10.6457 (0.0162)
10	20	20	10.6230 (0.0162)
20	10	5	9.7419 (0.0160)
20	16	8	9.6071 (0.0161)
20	20	10	9.6286 (0.0162)
25	10	5	9.4545 (0.0161)
25	16	8	9.3323 (0.0161)
25	20	5	9.3099 (0.0162)
40	10	5	9.0158 (0.0162)
50	8	5	8.8795 (0.0162)

Table 2: Comparison of different discretization parameters for the degradation model **(A2)**. We consider a time horizon of 30 days and set $y_0 = (0, 0, 0)$ for the initial control, i.e. we start with a new and empty battery. We have used $M = 10^3$ training sample paths and $M_{\text{test}} = 10^5$ paths for calculating $\widehat{\mathcal{J}}_0^{\pi}(y_0, X_0)$.

and has a significant impact on the performance. This improvement can be attributed to the reduced numerical error in the linear interpolation step in Algorithm 1. Notably, we observe that a finer discretization in the state of charge K^{SoC} has the biggest impact on the performance. This is because in our setup, increasing K^{SoC} also implicitly expands the action set (4.17), providing greater flexibility in controlling the battery. However, it is crucial to note that the computational cost grows approximately quadratically with the state of charge discretization K^{SoC} , as indicated by (4.12). On the other hand, increasing K^{ac} only leads to marginal improvements in the overall performance.

Figure 3 compares the (estimated) optimal control strategy for one sample path under the different degradation models. As expected, we observe bang-bang controls for the

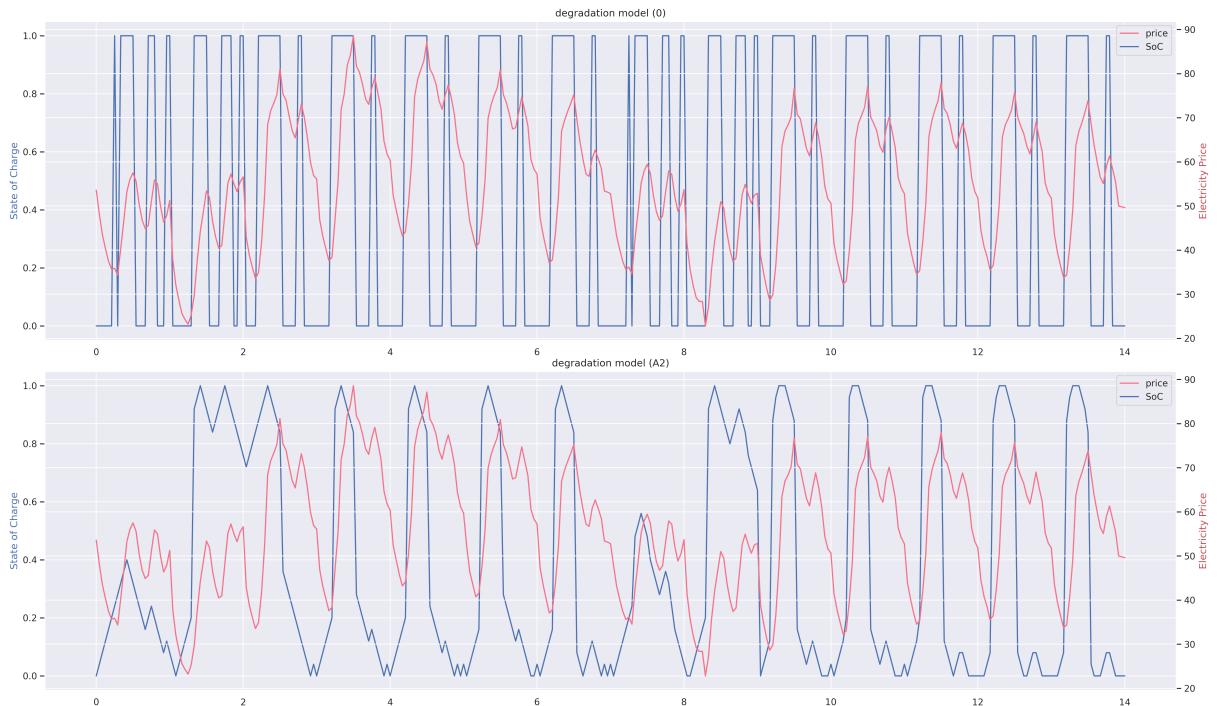


Figure 3: Comparison of the optimal control strategy for one sample path under degradation model **(0)** and **(A2)** with a time horizon of 7 days and initial control $y_0 = 0$. We have used $M = 10^3$ training sample paths and the discretization ($K^{\text{SoC}} = 25, K^{\text{av}} = 10, K^{\text{ac}} = 5$) for computing the control strategy $\hat{\pi}$.

case without degradation. In comparison, the degradation model **(A2)** regularizes the optimal control problem. That is, the optimal strategy often consists of slowly charging and discharging the battery to reduce the strain on degradation, instead of using the maximum C-rate possible.

5.2 Forecast experiments

Before we study the influence of wind forecasts on the optimal battery control problem, we aim to understand the role of seasonality in shaping the optimal control strategy. As depicted in Figure 1, day-to-day variations in electricity prices attributed to seasonality dominate the stochastic factors, i.e. the long and short-term component. Consequently, in Figure 2 and Figure 3, the optimal control strategy predominantly aligns with these deterministic daily fluctuations, leading to charging the battery during the night and discharging it during the day. As such, we might expect that if we exclude the stochastic factors in our model, we would still obtain similar control strategies and performance results. In order to explore this behaviour, we consider the electricity price model from Example 3.14, with identical parameters as specified in the previous section, but incorporating an additional coefficient $a_Z \in [0, 1]$, which allows us to control the impact of stochastic factors, namely

$$S_k = \Lambda_k + a_Z \left(Z_k^{\text{long}} + Z_k^{\text{short}} \right), \quad k = 0, \dots, N - 1.$$

For $a_Z = 1$, we retrieve the model from the previous section, while for $a_Z = 0$, the model becomes deterministic, following the seasonality function Λ . Table 3 compares the upper estimates for the cost over a time horizon of 30 days for various choices of a_Z . We observe

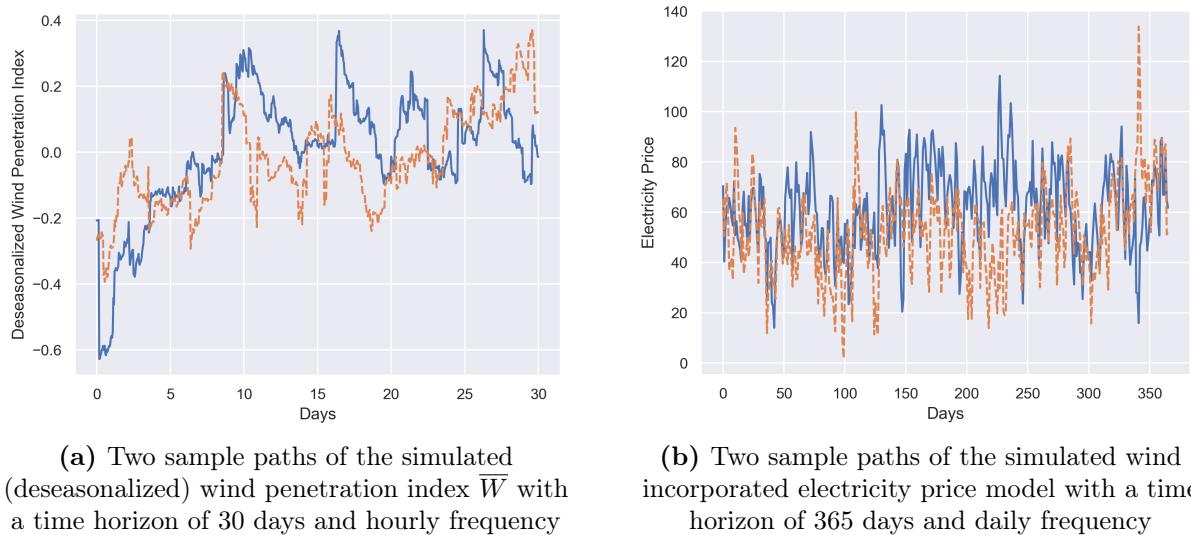
a_Z	$\widehat{\mathcal{J}}_0^{\pi}(y_0, X_0)$
1.0	9.4573 (0.0161)
0.75	9.4336 (0.0120)
0.5	9.4086 (0.0080)
0.25	9.3895 (0.0040)
0.0	9.3634 (0.0000)

Table 3: Comparison of upper estimates $\widehat{\mathcal{J}}_0^{\pi}(y_0, X_0)$ for different choices of a_Z . We considered a time horizon of 30 days and fixed $y_0 = (0, 0, 0)$ for the initial control. We have used $M = 10^3$ training sample paths and $M_{\text{test}} = 10^5$ paths and the discretization ($K^{\text{SoC}} = 25$, $K^{\text{av}} = 10$, $K^{\text{ac}} = 5$) for calculating $\widehat{\mathcal{J}}_0^{\pi}(y_0, X_0)$.

that removing the stochastic factors has little impact on the overall cost. In fact, we even achieve slightly lower values, which can be attributed to the parameters chosen for the short term factor, which cause an increase in the average price. In summary, the analysis reveals that deterministic daily fluctuations significantly influence the optimal control strategy.

In order to analyse the impact of wind forecasts on the optimal control strategy, we first need to specify a model for the wind-related variables. For this, we proceed, as discussed

in section 3.3, and model the (deseasonalized) wind penetration index $\bar{W}_k = \overline{WPI}_k$ by (3.14), that is, an Ornstein-Uhlenbeck process with a normal inverse Gaussian distributed BDLP. For the reversion rate, we choose $\theta = 0.01351$, and we assume that the distribution of the hourly increments of the BDLP is specified by the parameters ($\lambda = -0.5, \bar{\alpha} = 0.8296, \mu = 0.00218, \Sigma = 0.02458, \gamma = -0.00873$). Figure 4(a) depicts two independently simulated sample paths of the wind penetration index with the above parameters.



(a) Two sample paths of the simulated (deseasonalized) wind penetration index \bar{W} with a time horizon of 30 days and hourly frequency

(b) Two sample paths of the simulated wind incorporated electricity price model with a time horizon of 365 days and daily frequency

Figure 4: Independently simulated sample paths of the simulated wind penetration index \bar{W} and electricity price S .

As discussed in section 3.3, we extend the electricity price model from Example 3.14 by incorporating the wind penetration index into the short term factor by (3.13). For the regression coefficient, we follow [RVG21, Table 2] and choose $\beta = 74.4$. Further, we adopt the same values for the parameters of the seasonality and long term component and the CARMA(2, 1) process from the previous section. Notably, the introduction of the wind component into the electricity price model introduces additional variance into the prices. Indeed, this can be observed in Figure 4(b) which shows a plot of two independently simulated sample paths of the electricity price model with the given parameters. As highlighted in Section 4.3, we include the wind penetration index and its forecasted values in the state process denoted as $X_k = (S_k, \bar{W}_k, \bar{W}_{k+1}, \bar{W}_{k+K^f})^\top \in \mathbb{R}^{2+K^f}$, where $K^f \in \mathbb{N}$ represents the forecast period parameter, indicating the number of forecasted time steps. For this to make sense, however, we implicitly assume that \bar{W} is simulated for $N + K^f$ time steps, with a corresponding time horizon of $T + K^f \Delta t$.

In Table 4, we compare the computed upper estimates for different choices of forecast periods K^f . Note that we have included a benchmark value for reference. However,

K^f	$\widehat{\mathcal{J}}_0^\pi(y_0, X_0)$
benchmark	16.2820 (0.0239)
0	7.8728 (0.0240)
1	7.4707 (0.0239)
3	7.3899 (0.0237)
6	7.3666 (0.0237)
12	7.3664 (0.0239)
24	7.3703 (0.0238)

Table 4: Comparison of upper estimates $\widehat{\mathcal{J}}_0^\pi(y_0, X_0)$ for different choices of the forecast period parameter K^f . We considered a time horizon of 30 days and fixed $y_0 = (0, 0, 0)$ for the initial control. We have used $M = 10^3$ training sample paths and $M_{\text{test}} = 10^5$ paths and the discretization ($K^{\text{SoC}} = 25, K^{\text{av}} = 10, K^{\text{ac}} = 5$) for calculating $\widehat{\mathcal{J}}_0^\pi(y_0, X_0)$.

it should be highlighted that this value differs slightly from the benchmark presented in Table 1. This discrepancy arises due to our specific parameter selection for the wind penetration index \bar{W} , which has a negative mean, leading to slightly higher average prices. Despite that, we observe that the associated cost for $K_f = 0$, is lower than compared to the computed costs in Table 2. The case $K^f = 0$ corresponds to the setting where the system has no access to wind forecast information, but only the current wind penetration index. Nevertheless, this already provides valuable information about the electricity price which can be used to improve the control strategy, for instance due to the mean-reversion effect of the wind component. Most importantly, we observe that including a wind forecast to the state process improves the overall performance. However, it is worth mentioning that considering forecasts beyond the 3-hour horizon hardly results in further cost reduction. This finding indicates that the daily seasonality still plays a much more substantial role in shaping the optimal control strategy of the battery energy storage system. Nevertheless, the incorporation of forecasts proves to be a valuable addition in enhancing the accuracy of the battery control strategy.

6 Conclusion

In this thesis, we have developed a stochastic optimal control framework for battery energy storage systems. To accurately capture the dynamics of the system, we incorporated a realistic battery degradation model based on the average C-rate and Ah-throughput. Notably, the inclusion of battery degradation introduced a regularization effect to the optimal battery control problem, leading to more reasonable charging patterns, rather than simple bang-bang type controls. Moreover, we have discussed several stochastic electricity price models which successfully replicate essential stylized facts observed in real-world electricity markets. Among these models, the multifactor approach proposed in [RVG21] demonstrated its versatility by allowing the incorporation of additional stochastic factors, such as wind-related variables, to better model electricity prices. We found that integrating wind forecasts into the control problem improved the performance of the battery storage system. Our experiments also highlighted the significance of the daily seasonality in influencing the optimal battery control strategy. Understanding and accounting for this daily pattern is essential for achieving better control strategies. Throughout our numerical experiments, we relied on the presented least squares Monte Carlo algorithm 1 as an efficient numerical tool to solve the optimal battery control problem. In particular, it allowed us to gain a deeper understanding of the impact of various factors, such as degradation, electricity prices and forecasts, on the optimal control strategy.

Let us make a few comments on further research directions. As discussed in [Ben17], many real-world electricity price dynamics exhibit non-Markovian behaviour. This presents an exciting avenue for further research, that is, developing efficient numerical methods to handle non-Markovian stochastic optimal control problems. For instance, by using a least squares Monte Carlo approach with Markovian approximations, see [BB23]. In particular, future research can explore alternative optimization algorithms, such as reinforcement learning. Reinforcement learning algorithms have shown great promise in handling high-dimensional and continuous state and action spaces, making them suitable for optimizing the battery control problem. As an example, [KZ21] solve the optimal battery control problem using a deep Q-Network based reinforcement learning algorithm to efficiently search for the optimal control strategy. In particular, to model the degradation, they used introduced cycle-based battery degradation model according to the rainflow algorithm. Furthermore, to evaluate the effectiveness of a computed control strategy, one can employ what are known as dual methods, which can provide lower estimates for the cost. For further insights into dual methods, we refer to the work by [BS18].

In addition, it is worth pointing out that residential and commercial consumers are increasingly installing photovoltaic solar panels paired with battery storage systems. This

combination allows consumers to satisfy their demand directly from their solar panels and store excess energy in batteries for later use, thus reducing their reliance on the grid. Note that our developed framework can accommodate this setting, since we can interpret excess energy generated by solar panels as negative demand. As a result, the demand profile becomes inherently stochastic, introducing both new challenges and opportunities for the optimal battery control problem. In this context, it becomes crucial to also incorporate sun-related weather forecasts to predict the uncertain generation and demand patterns.

Overall, this thesis contributes to the ongoing research on the optimal control of battery energy storage systems by considering battery degradation, stochastic electricity prices, and weather forecasts. These considerations pave the way for the seamless integration of battery storage systems into the sustainable power grids of the future.

References

- [App09] David Applebaum. *Lévy Processes and Stochastic Calculus*. 2nd ed. Cambridge Studies in Advanced Mathematics. Cambridge University Press, 2009.
- [Bay+22] Christian Bayer et al. “Reinforced optimal control”. In: *Communications in Mathematical Sciences* 20.7 (2022), pp. 1951–1978.
- [BB23] Christian Bayer and Simon Breneis. “Markovian approximations of stochastic Volterra equations with the fractional kernel”. In: *Quantitative Finance* 23.1 (2023), pp. 53–70.
- [BBV13] Ole E. Barndorff-Nielsen, Fred Espen Benth, and Almut E. D. Veraart. “Modelling energy spot prices by volatility modulated Lévy-driven Volterra processes”. In: *Bernoulli* 19.3 (2013), pp. 803–845.
- [BDL18] Fred Espen Benth, Luca Di Persio, and Silvia Lavagnini. “Stochastic Modeling of Wind Derivatives in Energy Markets”. In: *Risks* 6.2 (2018).
- [Ben13] Fred Espen Benth. “Stochastic Volatility and Dependency in Energy Markets: Multi-Factor Modelling”. In: *Paris-Princeton Lectures on Mathematical Finance 2013*. Springer International Publishing, 2013, pp. 109–167.
- [Ben17] Mikkel Bennedsen. “A rough multi-factor model of electricity spot prices”. In: *Energy Economics* 63 (2017), pp. 301–313.
- [BLP17] Mikkel Bennedsen, Asger Lunde, and Mikko S. Pakkanen. “Hybrid scheme for Brownian semistationary processes”. In: *Finance and Stochastics* 21.4 (Oct. 2017), pp. 931–965.
- [BR11] Nicole Bäuerle and Ulrich Rieder. *Markov Decision Processes with Applications to Finance*. Universitext. Springer Berlin Heidelberg, 2011.
- [BS01] Ole E. Barndorff-Nielsen and Neil Shephard. “Non-Gaussian Ornstein-Uhlenbeck-Based Models and Some of Their Uses in Financial Economics”. In: *Journal of the Royal Statistical Society* 63.2 (2001), pp. 167–241.
- [BS18] Denis Belomestny and John Schoenmakers. *Advanced simulation-based methods for optimal stopping and control. With applications in finance*. London: Palgrave Macmillan, 2018.
- [DFG21] Thomas Deschatre, Olivier Féron, and Pierre Gruet. “A survey of electricity spot and futures price models for risk management applications”. In: *Energy Economics* 102 (2021).
- [GBC16] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT Press, 2016.

- [GHW15] Lajos Gergely Gyurkó, Ben M. Hambly, and Jan Hendrik Witte. “Monte Carlo methods via a dual approach for some discrete time stochastic control problems”. In: *Mathematical Methods of Operations Research* 81.1 (2015), pp. 109–135.
- [GKM11] Isabel García, Claudia Klüppelberg, and Gernot Müller. “Estimation of stable CARMA models with an application to electricity spot prices”. In: *Statistical Modelling* 11.5 (2011), pp. 447–470.
- [Gla13] Paul Glasserman. *Monte Carlo Methods in Financial Engineering*. Stochastic Modelling and Applied Probability. Springer New York, 2013.
- [HHM07] Ronald Huisman, Christian Huurman, and Ronald Mahieu. “Hourly electricity prices in day-ahead markets”. In: *Energy Economics* 29.2 (2007), pp. 240–248.
- [Hin70] Karl Hinderer. *Foundations of Non-stationary Dynamic Programming with Discrete Time Parameter*. Lecture Notes in Economics and Mathematical Systems. Springer Berlin Heidelberg, 1970.
- [Hir09] Guido Hirsch. “Pricing of hourly exercisable electricity swing options using different price processes”. In: *Journal of Energy Markets* 2 (2009), pp. 3–46.
- [HPV76] C. J. Himmelberg, T. Parthasarathy, and F. S. VanVleck. “Optimal Plans for Dynamic Programming Problems”. In: *Mathematics of Operations Research* 1.4 (1976), pp. 390–394.
- [KR65] Kazimierz Kuratowski and Czesław Ryll-Nardzewski. “A general theorem on selectors”. In: *Bull. Acad. Polon. Sci. Sér. Sci. Math. Astronom. Phys* 13.6 (1965), pp. 397–403.
- [KZ21] Kyung-Bin Kwon and Hao Zhu. “Reinforcement Learning-Based Optimal Battery Control Under Cycle-Based Degradation Cost”. In: *IEEE Transactions on Smart Grid* 13 (2021), pp. 4909–4917.
- [LS01] Francis A. Longstaff and Eduardo S. Schwartz. “Valuing American Options by Simulation: A Simple Least-Squares Approach”. In: *Review of Financial Studies* 14.1 (2001), pp. 113–147.
- [MT08] Thilo Meyer-Brandis and Peter Tankov. “Multi-Factor Jump-diffusion Models Of Electricity Prices”. In: *International Journal of Theoretical and Applied Finance* 11.05 (2008), pp. 503–528.
- [Pro13] Philip Protter. *Stochastic Integration and Differential Equations*. Stochastic Modelling and Applied Probability. Springer Berlin Heidelberg, 2013.

- [RVG21] Paulina A. Rowińska, Almut E.D. Veraart, and Pierre Gruet. “A multi-factor approach to modelling the impact of wind energy on electricity spot prices”. In: *Energy Economics* 104.C (2021).
- [Sch97] Eduardo S. Schwartz. “The Stochastic Behavior of Commodity Prices: Implications for Valuation and Hedging”. In: *The Journal of Finance* 52.3 (1997), pp. 923–973.
- [Sha+22] Nataliia Shamarova et al. “Review of Battery Energy Storage Systems Modeling in Microgrids with Renewables Considering Battery Degradation”. In: *Energies* 15.19 (2022).
- [SS00] Eduardo S. Schwartz and James E. Smith. “Short-Term Variations and Long-Term Dynamics in Commodity Prices”. In: *Management Science* 46 (2000), pp. 893–911.
- [TC15] Peter Tankov and Rama Cont. *Financial Modelling with Jump Processes, Second Edition*. Chapman and Hall/CRC. Taylor & Francis, 2015.
- [Ver16] Almut E. D. Veraart. “Modelling the Impact of Wind Power Production on Electricity Prices by Regime-Switching Lévy Semistationary Processes”. In: *Stochastics of Environmental and Financial Economics*. Springer International Publishing, 2016, pp. 321–340.
- [VV14] Almut E. D. Veraart and Luitgard A. M. Veraart. “Modelling Electricity Day-Ahead Prices by Multivariate Lévy Semistationary Processes”. In: *Quantitative Energy Finance: Modeling, Pricing, and Hedging in Energy and Commodity Markets*. Springer New York, 2014, pp. 157–188.
- [WBL22] Marc Weibel, Wolfgang Breymann, and David Lüthi. *ghyp: A package on generalized hyperbolic distributions*. R package version 1.6.3. 2022.
- [WZ88] Alan Weiser and Sergio E. Zarantonello. “A note on piecewise linear and multilinear table interpolation in many dimensions”. In: *Mathematics of Computation* 50.181 (1988), pp. 189–196.
- [Zan13] Daniel Z. Zanger. “Quantitative error estimates for a least-squares Monte Carlo algorithm for American option pricing”. In: *Finance and Stochastics* 17.3 (2013), pp. 503–534.
- [Zan18] Daniel Z. Zanger. “Convergence of a Least-squares Monte Carlo Algorithm for American Option Pricing with Dependent Sample Data”. In: *Mathematical Finance* 28.1 (2018), pp. 447–479.