Bachelor thesis

Stochastic optimal control problems of a diesel generator in a microgrid

Noah Scheider

submitted on the 4th October 2023

Supervisor TT-Prof. Dr. Sebastian Krumscheid
Co-Supervisor Prof. Dr. Mathias Trabs
Institute Institute for Applied and Numerical Mathematics
University Karlsruher Institute for Technology

Karlsruher Institut für Technologie

Acknowledgements

A big thanks to all the people who supported me during this work. Mainly, I want to thank Prof. Nobile from the Ecole Polytechnique de Lausanne who suggested this topic and accompanied me throughout all the months. His intuitive approach, patience and willingness to engage in countless discussions stuck with me and I am grateful for this transnational cooperation. Many thanks to Tommaso Vanzan who helped me on multiple occasions and whose kindness, dedication and cooperativeness reassured me along the project. Moreover, I highly appreciated the co-supervision by TT-Prof. Krumscheid, who also accompanied me through the thesis and gave many valuable and technical insights that helped clarify a variety of concepts.

Abstract

The thesis focuses on the challenge of efficiently utilizing a diesel generator within a grid-connected microgrid. The research examines how a microgrid, comprising a consumer, a diesel generator, a public grid, and solar panels, manages its power regulation. A stochastic model is established with the objective of minimizing the energy cost for the consumer through an optimal control policy. Equivalent numerical solutions are obtained using an appropriate Hamilton-Jacobi-Bellman equation and the Regression Monte Carlo method, both of which are implemented and assessed.

Contents

1 Introduction

Microgrids can be seen as the essential energetic subparts of mighty autonomous systems - an idea popularized through Hollywood movies like The Martian or Passengers. The soaring interest and research in these local power supply possibilities are gradually turning this science-fictional vision into reality.

A microgrid can be seen as an agglomeration of power nodes that inject or retrieve energy within a regulated system. The picturesque example above, however, describes only one type of microgrids - so-called remote microgrids[1]. These kinds of power grids operate completely detached from external energy sources and therefore need to balance the energy flux autonomously. External energy sources are sources of energy that lie outside the boundaries of the auto-regulated system. Grid-connected microgrids constitute another set of grids, differing from the previously elaborated ones by an on-/off connection to a larger energy system. A connection of this kind allows for a backup solution if local energy production is disrupted. A third class of microgrids is given by networked microgrids, where grid-connected and off-grid microgrids form larger independent systems with potentially shared nodes, e.g. joint energy plants or storages.

Contemporary civil engineering projects try to incorporate this concept of energetic autonomy. For instance, Menlo Park[2], a city in California, strives to achieve a decentralized solar-based power supply to satisfy the demand of multiple consumption entities, such as a Red Cross emergency shelter and parking lots for electric vehicles. Even though this particular example is largely driven by financial arguments, other factors favour a more general adoption of microgrids. Currently, those are primarily attributed to the escalation of conflicts, natural disasters, and economic disruptions, which contradict the escalating global need for stable energy. Technological advancements

[1] https://www.energy.gov/sites/prod/files/2018/12/f58/remote-microgrids-dan-ton.pdf
[2] https://www.microgridknowledge.com/electric-vehicles/article/11428109/

in green energy production and storage capabilities further bolster the feasibility of microgrids, providing efficient, sustainable, and resilient solutions to local demands. However, implementing microgrids comes at the cost of complexity, requiring sophisticated models for efficient power regulation within the network.

The theory dictating and quantifying this balancing procedure is called optimal control theory and has been considerably extended within the last decades. Recent research has explored a wide range of applications with diverse setups. Noticeable distinctions usually arise by considering different production sources, such as windmills, solar panels [1] or water turbines [2] and storage possibilities, such as batteries [1] or gas storages [3]. Alasseur et al. [4] deals with the coexistence of a diesel generator and a public grid. It builds a coherent model and presents a way to compute an optimal control strategy via Regression Monte Carlo (RMC), a method that has been initially developed for option pricing [5]. In the field of quantitive finance, the selling time of a financial contract is often assumed to be the controllable variable [5]. Its underlying idea of approximating conditional expectations backwards in time by a combination of Linear Regression and Monte Carlo (MC) simulations can however be generalized and paves the way for efficient computation of more general optimal control strategies. In general, Linear Regression is a method that predicts future data based on current data. A MC simulation is a sample average method that is based on the Law of large numbers and is used to approximate expected values of random variables [6]. Though RMC is a prominent method in this context, another approach is based upon the idea of solving a specific second-order, nonlinear Partial Differential Equation (PDE).

This thesis is motivated by Alasseur et al. [4] and deals with a slightly modified grid-connected microgrid: The battery is removed, and the access cost to the diesel generator is eliminated. The access cost denotes a fixed cost that has to be paid whenever the generator is turned on. The components and their interplay are illustrated in Figure 3.1. Two models are inspected, where one is the extension of the other. Their dissimilarity comes from a change in the residual demand process. The residual demand typically represents a stochastic process resulting from the difference between consumption and production in time. The seemingly easier one leads to a setting where a simple MC approach is sufficient. Its extension requires caution due to the absence of energy storage [7] and the existence of a degenerate residual demand. The residual demand is called degenerate when it is influenced by the control policy. The numerical solution requires a novel approach that

extends RMC by using Importance Sampling. At its core, this is a variance reduction technique, that is usually used to enhance the accuracy of MC approximations while reducing the variance of the generated samples [6]. In this setting, the method circumvents the theory of Backwards Stochastic Differential Equation (BSDE) [8] and hasn't been explored yet to the best of current knowledge.

The following chapters outline the research and present the findings. Chapter 2 reviews relevant results from stochastic calculus to establish the theoretical basis for the models introduced in Chapter 3 and Chapter 4. Additionally, it motivates a general set of stochastic optimal control problems and highlights the formal link between the Dynamic Programming Principle (DPP) and the Hamilton-Jacobi-Bellman equation (HJB). Chapter 3 introduces a simple microgrid model, examines it and presents two numerical solutions. A theoretical and numerical convergence analysis is carried out for the different methods. Chapter 4 extends the previous model and proposes once again two numerical solutions using the Finite Difference Method (FDM) and RMC, with a focus on the latter one. It develops the idea of Importance Sampling mentioned above and summarizes it within an algorithm. The results are discussed in Chapter 5. Chapter 6 concludes the thesis.

2 Stochastic Optimal Control Theory

This chapter provides a theoretical context for stochastic optimal control problems. It serves as a mathematical basis for the models in Chapter 3 and Chapter 4. First, key insights about stochastic calculus are summarized. A general stochastic control problem framework leading to the DPP is motivated and a formal derivation of the HJB is shown. This link allows to switch perspectives on the same problem. A discussion giving advantages and drawbacks of such a transition concludes the chapter. Standard probability concepts, such as measurability, adaptedness, σ-algebras, stopping times, martingales, their according properties and others are used in this chapter and throughout the rest of the thesis without further introduction.

2.1 Theory of Stochastic Calculus

Essential definitions used to introduce the theory are taken from Subsection 1.1.1 in Pham [9]. Let $(\Omega, \mathcal{F}, \mathcal{F}_t, \mathbb{P})$ be a filtered probability space with $\mathcal{F}$ being the natural filtration to $\{\mathcal{F}_s\}_{0\leq s\leq T}$. T denotes a finite time horizon and $0 < T < \infty$. $\mathcal{F}_t$ is a filtration at time t, that is generated by a d-dimensional continuous Brownian motion $\{W_s\}_{0\leq s\leq t}$ with $0 \leq t \leq T$, i.e. $\mathcal{F}_t := \sigma(W_s)$ for $0 \leq s \leq t$. $\mathcal{F}_t$ is the smallest σ-field such that W_s is measurable for $s \in [0, t]$. W_s is a progressively measurable process, as $(s, \omega) \mapsto W_s(\omega)$ is measurable on $[0, t] \times \Omega$ for $\forall \omega \in \Omega$ equipped with the product σ-field $\mathcal{B}([0, t]) \otimes \mathcal{F}_t$. A Stochastic Differential Equation (SDE) is a dynamic of the form

$$dX_s = b(s, X_s)ds + \sigma(s, X_s)dW_s \tag{2.1}$$

starting in $X_0 = x_0 \in \mathbb{R}^n$, with a drift function $b : [0, T] \times \mathbb{R}^n \to \mathbb{R}^n$ and diffusion term $\sigma : [0, T] \times \mathbb{R}^n \to \mathbb{R}^n \times \mathbb{R}^d$. According to Definition 6.2 in Chapter 1 of Yong et al. [10] a continuous process X_t is a strong solution to (2.1), if

$$\int_0^t |b(s, X_s)| + |\sigma(s, X_s)|^2 ds < \infty, \ \forall t \in [0, T]$$

$$X_t = x + \int_0^t b(s, X_s) ds + \int_0^t \sigma(s, X_s) dW_s \tag{2.2}$$

$|\cdot|$ on the drift function b denotes the sum of the elementwise absolute values, that is $|b(s, X_s)| = \sum_{i=1}^n |b_i(s, X_s)|, \ \forall s \in [0, T]$. Similarly, $|\cdot|$ on the diffusion term σ stands for $|\sigma(s, X_s)|^2 = \sum_{i,j=1}^{n,d} |\sigma_{i,j}(s, X_s)|^2, \ \forall s \in [0, T]$. Theorem 6.3 in Chapter 1 of Yong et al. [10] shows that (2.1) admits a unique, strong solution, if

$$|b(s, x) - b(s, y)| + |\sigma(s, x) - \sigma(s, y)| \leq L|x - y|, \ \forall s \in [0, t] \tag{2.3}$$

$$\int_0^t \left(|b(s, 0)| + |\sigma(s, 0)|\right)^2 ds < \infty \tag{2.4}$$

The solution is unique in the sense, that any two strong solutions coincide almost surely along their trajectory. There are more conditions that induce different types of solutions. A brief overview of the prerequisites can be found in Cherny et al. [11].

Consider a control process $a : [0, T] \times \mathbb{R}^n \to A \subset\subset \mathbb{R}^m$ that is indirectly random in the sense that $a_s(\omega) = a(s, X_s(\omega))$ for $\omega \in \Omega$. The double inclusion $\subset\subset$ denotes, that A is compactly contained in $\mathbb{R}^m$. Moreover, the mapping $s \mapsto a_s$ is said to be in $C([0, T], \mathbb{R}^n)$. Therefore, it is progressively measurable due to Lemma 6.1 in Yong et al. [10], as it is clearly càdlàg, i.e. right-continuous with left limits everywhere and adapted. These kinds of control processes are called Markovian controls [12]. Given a, (2.1) is extended such that

$$dX_s = b(s, X_s, a_s) ds + \sigma(s, X_s, a_s) dW_s \tag{2.5}$$

with initial value $X_0 = x_0 \in \mathbb{R}^n$. Definition 6.15 in Chapter 1 of Yong et al. [10] says that an $\{\mathcal{F}_t\}$-adapted continuous process $X_t, \ t \geq 0$ is a solution of (2.5), if

$$\int_0^t |b(s, X_s(\omega), a_s(\omega))| + |\sigma(s, X_s(\omega), a_s(\omega))|^2 ds < \infty, \tag{2.6}$$

$$X_t(\omega) = x_0 + \int_0^t b(s, X_s(\omega), a_s(\omega)) ds + \int_0^t \sigma(s, X_s(\omega), a_s(\omega)) dW_s(\omega) \tag{2.7}$$

for $t \geq 0$ a.s. The notation for b and σ is used from Subsection 1.3.1 in Pham [9]. Theorem 1.3.15 in Pham [9] states, that (2.5) admits a unique and strong solution X_t if b and σ satisfy the following two conditions. There $\exists K > 0$ such that for $\forall x, y \in \mathbb{R}^n,\ \alpha \in A,\ s \in [0, t]$

$$|b(s, x, \alpha) - b(s, y, \alpha)| + |\sigma(s, x, \alpha) - \sigma(s, y, \alpha)| \leq K|x - y| \tag{2.8}$$

$$|b(s, x, \alpha)| + |\sigma(s, x, \alpha)| \leq \kappa_{s,\alpha} + K|x| \tag{2.9}$$

$\kappa_{s,\alpha}$ is a real-valued process in $L^2([0, T])$. Theorem 1.3.15 in Pham [9] indicates that (2.7) can be defined in a way that it has a unique and strong solution $\{X_s^{t,x_t}\}_{t \leq s \leq T}$ with initial value $x_t \in \mathbb{R}^n$ at time $t > 0$. The same theorem reveals, that for all $T > t$, there exists $C_T > 0$ such that

$$\mathbb{E}\left[\sup_{t \leq s \leq T} |X_s|^p\right] \leq C_T(1 + |x_t|^p),\ p \geq 1$$

2.2 Motivating Stochastic Control Theory

This section motivates a general framework and introduces expressions used throughout the rest of the thesis. Moreover, it provides context for the derivation of the HJB in (2.3). Its mathematical assumptions are based upon Pham [9] if not stated otherwise.

Consider a player or a consumer named Jim within a grid-connected microgrid. He requires access to power at all times and constantly consumes a certain amount of energy $x \in \mathbb{R}^n$. Therefore, he is forced to pay according to a Borel-measurable function f, called the cost function. $f : [0, \infty) \times \mathbb{R}^n \to \mathbb{R}, (s, x) \mapsto f(s, x)$ dictates the cost of energy at any time $s \in [0, \infty)$. Hereafter, the player is able to deal with his current situation by controlling the amount and source of energy. The cost function f is changed to $f : [0, \infty) \times \mathbb{R}^n \times A \to \mathbb{R}$. It is assumed to be measurable and of quadratic growth. This means

$$|f(s, x, \alpha)| \leq C(1 + |x|^2 + |b(s, 0, \alpha)|^2 + |\sigma(s, 0, \alpha)|^2)$$

$\forall (s, x, \alpha) \in [0, \infty) \times \mathbb{R}^n \times A$. Jim's current situation can be described as $(0, x, \alpha)$. He is now interested in the total cost at some final time T, e.g. the end of a month or a year. He wants to reduce his expenses by controlling the source of energy to the best of his capability. His demand and his local energy supply are cyclical due to changes in loads during day and night[1].

[1] https://www.energy.gov/eere/articles/confronting-duck-curve-how-address-over-generation-solar-energy

Both are of stochastic nature and their difference is denoted by X_s. For a stable prediction, all possible outcomes are taken into account. The naïve cost functional is defined as

$$\mathbb{E}\left[\int_0^T f(s, X_s^{0,x}, a_s)ds\right]$$

and reads as the expected value of the accumulated cost processes based on all possible residual demands and their corresponding control strategy a_s along a time period $[0, T]$. Further assumptions are required to make the expression well-defined. A terminal fee is common in these kinds of scenarios and either stems from the microgrid set-up or the access to the public grid [1, 4]. Alasseur et al. [4] suggests linking the terminal fee to a scenario, where a battery has been rented and needs to be returned with the same level of charge. This final condition is important for the latter derivation of the HJB in Section 2.3, where the terminal fee turns into the terminal condition for the corresponding backwards PDE. It is given by $g(X_T^{0,x})$ for a measurable function $g : \mathbb{R}^n \to \mathbb{R}$ and needs to be of quadratic growth, i.e. $|g(x)| \leq C(1 + |x|^2)$ for $\forall x \in \mathbb{R}^n$. In general, the growth conditions on f and g are imposed to keep a reasonable growth rate such that (2.11) is well-defined. Moreover, the strategy a of the consumer has to be reasonable. $\mathcal{A}(0, x)$ denotes the space for the control strategies starting in time $t = 0$ and state x such that

$$\mathbb{E}\left[\int_0^T |f(s, X_s^{0,x}, a_s)|ds\right] < \infty \tag{2.10}$$

This enables to introduce the cost functional J as

$$J(0, x, a) := \mathbb{E}\left[\int_0^T f(s, X_s^{0,x}, a_s)ds + g(X_T^{0,x})\right] \tag{2.11}$$

It is also possible to think of J as a gain functional, especially in settings, where profits are maximized. The cost functional can be decomposed into two parts. Therefore, let $\theta \in [0, T]$, then

$$J(0, x, a) = \mathbb{E}\left[\int_0^\theta f(s, X_s^{0,x}, a_s)ds + J(\theta, X_\theta^{0,x}, a^\theta)\right], \tag{2.12}$$

where $a^\theta \in \mathcal{A}(\theta, X_\theta^{0,x})$. This identity uses the result $X_s^{0,x} = X_s^{\theta, X_\theta^{0,x}}$ a.s. from Theorem 1.3.15 in Pham [9] and the Law of iterated conditional expectations. The player wants to choose the best $a \in \mathcal{A}(0, x)$ that minimizes his total cost.

This assumption leads to the value function

$$v(0,x) := \sup_{a \in \mathcal{A}(0,x)} \mathbb{E}\left[\int_0^T f(s, X_s^{0,x}, a_s) + g(X_T^{0,x}) ds\right] \tag{2.13}$$

The supremum is taken here, as the previously introduced costs are negative. Maximizing the value function then corresponds to minimizing expenses. The question of how to determine an optimal strategy a^* arises. Several possibilities are presented throughout this work and another equivalent representation for (2.13) is shown in the next section.

2.3 Hamilton-Jacobi-Bellman equation

This section turns the previously stated value function (2.13) into a nonlinear, second-order PDE. This is a profound result and is therefore included in this work. The DPP is introduced and lays the origin for the derivation of the HJB. It is further used in Chapter 4 to come up with a numerical solution to v. Until now, (2.11) and (2.13) were only introduced for $t = 0$, but $t \in [0,T]$ in general.

Theorem 2.1 (Dynamic Programming Principle [9]).
Let $(t,x) \in [0,T] \times \mathbb{R}^n$. Then

$$\begin{aligned} v(t,x) &= \sup_{a \in \mathcal{A}(t,x)} \sup_{\theta \in \mathcal{T}_{t,T}} \mathbb{E}\left[\int_t^\theta f(s, X_s^{t,x}, a_s) ds + v(\theta, X_\theta^{t,x})\right] \\ &= \sup_{a \in \mathcal{A}(t,x)} \inf_{\theta \in \mathcal{T}_{t,T}} \mathbb{E}\left[\int_t^\theta f(s, X_s^{t,x}, a_s) ds + v(\theta, X_\theta^{t,x})\right], \end{aligned}$$

where $\mathcal{T}_{t,T}$ is the set of stopping times valued in $[t,T]$. Another but slightly stronger version is given by

(i) For $\forall \varepsilon > 0, \exists a \in \mathcal{A}(t,x)$ such that for $\forall \theta \in \mathcal{T}_{t,T}$

$$v(t,x) - \varepsilon \leq \mathbb{E}\left[\int_t^\theta f(s, X_s^{t,x}, a_s) ds + v(\theta, X_\theta^{t,x})\right] \tag{2.14}$$

(ii) For $\forall a \in \mathcal{A}(t,x)$ and for $\forall \theta \in \mathcal{T}_{t,T}$

$$v(t,x) \geq \mathbb{E}\left[\int_t^\theta f(s, X_s^{t,x}, a_s) ds + v(\theta, X_\theta^{t,x})\right] \tag{2.15}$$

The first two equations for v in Theorem 2.1 imply

$$v(t,x) = \sup_{a \in \mathcal{A}(t,x)} \mathbb{E}\left[\int_t^\theta f(s, X_s^{t,x}, a_s)ds + v(\theta, X_\theta^{t,x})\right] \tag{2.16}$$

for all $\theta \in \mathcal{T}_{t,T}$. This suggests that the value function can be split into two parts at any time θ in $[t,T]$. The first one is a functional based on the cost function, whereas the second part is a future version of the value function itself. This enables the application to discrete-time settings, where any time step in the discretized time interval $[t,\dots,T]$ can be identified with a constant stopping time.

The proof of Theorem 2.1 shows the two following inequalities

$$\begin{aligned} v(t,x) &\geq \sup_{a \in \mathcal{A}(t,x)} \sup_{\theta \in \mathcal{T}_{t,T}} \mathbb{E}\left[\int_t^\theta f(s, X_s^{t,x}, a_s)ds + v(\theta, X_\theta^{t,x})\right] \\ v(t,x) &\leq \sup_{a \in \mathcal{A}(t,x)} \inf_{\theta \in \mathcal{T}_{t,T}} \mathbb{E}\left[\int_t^\theta f(s, X_s^{t,x}, a_s)ds + v(\theta, X_\theta^{t,x})\right], \end{aligned}$$

by using the slightly stronger version (2.14), (2.15) of the DPP. A sketch following the rigorous proof in Pham [9] is presented.

For the first inequality, the proofs starts by finding a control strategy and stopping time satisfying $v(\cdot,\cdot) - \varepsilon \leq J(\cdot,\cdot,\cdot)$ for a fixed ϵ. There is no connection to (2.14) so far. This holds by definition of the value function and equality remains only for $\varepsilon = 0$ and $a = \arg\sup_{a \in \mathcal{A}(\cdot,\cdot)} J(\cdot,\cdot,a)$, which exists due to $a \in \mathcal{A}(\cdot,\cdot)$. Another control strategy a is defined such that it changes at the previously fixed stopping time. The cost functional J is decomposed upon this stopping time as done in (2.12). The part inside of the expected values then satisfies the previously stated ε-inequality. (2.14) is now applicable and shows the first inequality.

The second inequality starts with the decomposed version of J as done in (2.12). The inner cost functional is substituted by the value function, which induces an inequality that holds for $\forall a \in \mathcal{A}(t,x)$ and $\forall \theta \in \mathcal{T}_{t,T}$ and consequently for the infimum in $\mathcal{T}_{t,T}$. Adding the supremum in front of the infimum turns the initial cost functional, that is the LHS of (2.12) into the value function, proving the inequality.

The DPP enables the derivation of the HJB. It is presented as in Pham [9]. Consider $\theta = t + h$ and a control strategy that is constant in an interval

$[t, t+h]$, i.e. $a_s \equiv \alpha \in A,\ \forall s \in [t, t+h]$. The inequality of (2.15) suggests

$$v(t,x) \geq \mathbb{E}\left[\int_t^{t+h} f(s, X_s^{t,x}, \alpha)ds + v(t+h, X_{t+h}^{t,x})\right] \tag{2.17}$$

For $v \in C^{1,2}$, Itô's formula dictates between t and $t+h$

$$\begin{aligned} v(t+h, X_{t+h}^{t,x}) = v(t,x) &+ \int_t^{t+h} \left(\frac{\partial}{\partial t}v + \mathcal{L}^\alpha v\right)(s, X_s^{t,x})ds \\ &+ \int_t^{t+h} (D_x v(s, X_s^{t,x}))^T \sigma(s, X_s^{t,x}, \alpha) dW_s, \end{aligned} \tag{2.18}$$

where

$$\mathcal{L}^\alpha v(s,x) = b(s,x,\alpha) \cdot D_x v(s,x) + \frac{1}{2}\mathrm{tr}\left(\sigma(s,x,\alpha)\sigma(s,x,\alpha)^T D_x^2 v(s,x)\right)$$

The last term in (2.18) needs to be examined further because of its stochastic increments. By continuity of $X_s^{t,x}$ and $D_x v$ in time and the Lipschitz condition for σ

$$\int_t^{t+h} |(D_x v)^T \sigma|^2 ds \leq 2\int_t^{t+h} |(D_x v)^T|^2 ds + 2\int_t^{t+h} |\sigma|^2 ds < \infty$$

holds a.s. If the leftmost expression in the inequality chain is well-defined within its expected value, $(D_x v)^T \sigma$ is in $\mathcal{H}^2([0,T])$, where

$$\mathcal{H}^2([0,T]) := \{h|\ h \text{ measurable, adapted and } \mathbb{E}\left[\int_0^T h^2(\omega, s)ds\right] < \infty\}$$

This observation would turn the stochastic integral in (2.18) into a martingale, implying

$$\mathbb{E}\left[\int_t^{t+h} (D_x v(s, X_s^{t,x}))^T \sigma(s, X_s^{t,x}, \alpha) dW_s\right] = 0$$

The issue is that X_t is not necessarily bounded preventing the application of the Dominated Convergence Theorem (DOM). By its construction, it can be shown that for any $\epsilon > 0$ it is possible to choose a sufficiently large compact domain such that the probability of containing X_t is greater than $1-\epsilon$ for h arbitrarily small. The idea is supported by the Gaussian increments whose probability density decays exponentially fast and indicates that continuously transformed, bounded trajectories on a compact domain remain bounded. Thus, concluding the existence of $(D_x v)^T \sigma$ within $\mathcal{H}^2([0,T])$. This technique was suggested by TT-Prof. Dr. Krumscheid.

Another way leading to the mean zero result is to choose a different stopping time than $\theta = t+h$. One, that also depends on h and stops when X_t grows too fast as done in Proposition 3.4 in Touzi et al. [13]. This preserves the boundedness of the integrand and consequently its existence within $\mathcal{H}^2([0,T])$. The first argument was given because it is needed later in Section 3.5.2.

Substituting (2.18) into (2.17) then yields

$$0 \geq \mathbb{E}\left[\int_t^{t+h} \left(\frac{\partial}{\partial t}v + \mathcal{L}^\alpha v\right)(s, X_s^{t,x}) + f(s, X_s^{t,x}, \alpha)ds\right]$$

Dividing by h and applying the Mean-value theorem for definite integrals returns

$$\begin{aligned} 0 &\geq \frac{\partial}{\partial t}v(t,x) + \mathcal{L}^\alpha v(t,x) + f(t,x,\alpha) \\ \implies -\frac{\partial}{\partial t}v(t,x) &- \sup_{\alpha \in A}\left[\mathcal{L}^\alpha v(t,x) + f(t,x,\alpha)\right] \geq 0 \end{aligned}$$

If a^* in $\mathcal{A}(t,x)$ is the optimal control process, (2.17) turns into equality. By repeating the same steps with $\{X_s^{t,x}\}_{t\leq s\leq T}^*$ being the solution to

$$dX_s = b(s, X_s, a_s^*)ds + \sigma(s, X_s, a_s^*)dW_s$$

the last inequality turns into equality and provides

$$-\frac{\partial}{\partial t}v(t,x) - \mathcal{L}^{\alpha^*}v(t,x) + f(t,x,\alpha^*) = 0 \tag{2.19}$$

Therefore, v satisfies

$$-\frac{\partial}{\partial s}v(s,x) - \sup_{\alpha\in A}\left[\mathcal{L}^\alpha v(s,x) + f(s,x,\alpha)\right] = 0, \ \forall (s,x) \in [0,T)\times\mathbb{R}^n \tag{2.20}$$

Theorem 3.5.2 in Pham [9] shows that the value function v solves (2.20) with a few assumptions. The optimization problem (2.13) is therefore equivalent to finding a solution to (2.20). Whether a transition to the PDE formulation is reasonable, depends on the specific case at hand. A method to compute a numerical solution to the HJB is proposed in Chapter 3 and Chapter 4. To conclude this section, note that the smoothness assumption on v is existential for the derivation to be possible. In practice, this condition is oftentimes not met (cf. Chapter 3) or difficult to prove. If $v \in C^{1,2}$ can't be shown, it is possible to apply the concept of viscosity solutions [9]. These only require local boundedness on v and quadratic growth on f, i.e.

$$|f(t,x,\alpha)| \leq C(1+|x|^2) + \kappa(\alpha), \ \forall (t,x,\alpha) \in [0,T]\times\mathbb{R}^n \times A$$

for some positive constant C and positive function $\kappa : A \to \mathbb{R}_{\geq 0}$. A further discussion is deployed to Chapter 4 in Pham [9].

2.4 Computing the Value Function

Section 2.3 showed a new perspective for the optimal control setting established in Section 2.2. This section discusses which approach to use for different circumstances. The main parameter is the dimension of $\{X_s\}_{t\leq s\leq T}$.

Subsection 5.1.3 in Evans [14] features a naïve approach to the design of an optimal control policy in a deterministic setting. In Subsection 7.5.2 it is then extended to the stochastic one. The approach suggests, that given a solution v on $[t,T]\times\mathbb{R}^n$ it is possible to compute

$$\hat{a}(s,x) = \arg\max_{\alpha\in A} \left[\mathcal{L}^{\alpha} v(s,x) + f(s,x,\alpha)\right],\ \forall (s,x)\in[t,T]\times\mathbb{R}^n$$

This allows to set $dX_s^* = b(s,X_s,\hat{a}_s)ds + \sigma(s,X_s,\hat{a}_s)dW_s$ for some initial condition. Provided dX_s^*, the optimal control policy is then of the form $a_s^* = \hat{a}(s,X_s^*)$. To compute v, several aspects determine the appropriate method. A major one is the dimensionality n of the underlying random process $\{X_s\}_{t\leq s\leq T}$. For dimension one or two it is reasonable to transition to the PDE formulation of (2.13) and aim for an explicit or implicit FDM to solve (2.20). At this point the influence of a on $\{X_s\}_{t\leq s\leq T}$ is neglectable.

Increasing d however, diminishes the computational efficiency of FDM. The reason is that FDM provides a solution to an entire Domain of Interest (DoI), even though in reality only a certain point within this area needs to be considered, as was explained in Section 2.2. The DoI is a domain that contains a relevant range for the microgrid. Examples for an appropriate DoI are given in Chapter 3 and Chapter 4. In this setting, it is reasonable to solve (2.13). Recall the multi-dimensional dynamic of $\{X_s\}_{t\leq s\leq T}$.

$$dX_s = b(s,X_s,a_s)ds + \sigma(s,X_s,a_s)dW_s,\ \forall s\in[t,T]$$

for some initial condition $X_t = x$. Its individual components are given by

$$dX_s^i = b_i(s,X_s,a_s)ds + \sum_{j=1}^{d}\sigma_{i,j}(s,X_s,a_s)dW_s^j,\ 1\leq i\leq n$$

with initial condition $X_t^i = x_i$. The question of whether and how a acts on $\{X_s\}_{t\leq s\leq T}$ needs to be addressed and three possibilities are considered in the following.

- If a doesn't influence the underlying process $\{X_s\}_{t\leq s\leq T}$ but only the cost function f, a time-discretized version of (2.13) can be approximated by MC. This is possible, becazse $\{X_s\}_{t\leq s\leq T}$ can be sampled

explicitely. Moreover, the optimal control policy is given by the maximization of the cost function, i.e. $a(s,x)^* = \arg\sup_{\alpha \in A} f(s,x,\alpha)$, $\forall (s,x) \in [t,T] \times \mathbb{R}$. The value function is then given by the average of the integral, summing up the cost functions along the sampled trajectories. Chapter 3 illustrates this case.

- When a only acts on deterministic components of $\{X_s^i\}_{t \leq s \leq T}$, there exist $\sigma_{i,j}(s, X_s, a_s) \equiv 0$ for $\forall j = 1, \dots, d$. These deterministic elements usually describe storage dynamics, such as for a battery or for gas storage and are solved by a RMC approach (cf. [3, 4]). This setting is known as an inventory-control problem. In a time-discretized version of (2.13) with finite horizon, the DPP is applied, splitting the value function at the penultimate time-step. The optimal control policy is then computed backwards in time until the initial point (t,x) is reached. Now, all components of $\{X_s\}_{t \leq s \leq T}$ are sampled forwards in time and a mean-average idea summing over the costs along each trajectory yields an approximation to $v(t,x)$. Different variants for the backwards in time method exist and are examined in Balata et al. [7].

- When a purely acts on stochastic processes, the previous approach is no longer eligible. The residual demand is now degenerate and current approaches require the use of BSDEs [15]. An overview of their numerical implementation so far is given in Chessari et al. [8]. The overall idea is to solve a BSDE and map it to the value function. It is laid out in the Introduction of Kharroubi et al. [15]. Chapter 4 deals with the case, where a acts on the drift part of a one-dimensional residual demand process and presents a way to circumvent the field of BSDEs. It seems reasonable to extend the idea to the case, where a also acts on the diffusion part of the residual demand X_t.

3 A Simple Microgrid model

Research has explored many different applications and a few of them have been introduced in Chapter 1. This chapter deals with a simple grid-connected microgrid, motivated by Alasseur et al. [4]. It is called simple because it provides a practical understanding of the earlier introduced theory and because is a restriction for the model presented in Chapter 4. The chapter starts by introducing the model and its components. A formal problem setting is established and different possibilities for the cost function are presented. It considers the case, where the underlying residual demand process loses its diffusion term and is consequently deterministic inducing an analytical solution. Two numerical solutions follow for a non-zero diffusion, and their convergence is investigated afterwards on a theoretical and numerical level.

3.1 Model Description

The model consists of four components, including a household consumer, solar panels for renewable energy production, a diesel generator, and a power grid connection. These components work together to manage energy consumption and generation. The difference between energy consumed by the household and the renewable energy produced with solar panels is called residual demand. It is denoted by X_s and displays the energy surplus or deficiency at time s. Consequently, consumption is firstly covered by solar power before other energy sources are accessed. Moreover, the micro-grid contains a diesel generator that provides energy $d_s \in [0, d_{max}]$ at time s and is controlled by the consumer. Diesel power comes with a price $p_s \geq 0$, whose tradeoff with the cost $c_s \geq 0$ of the energy g_s from the public grid needs to be examined. The model operates within a finite time horizon $0 < T < \infty$ to be consistent with Chapter 2. Figure 3.1 portrays the model.

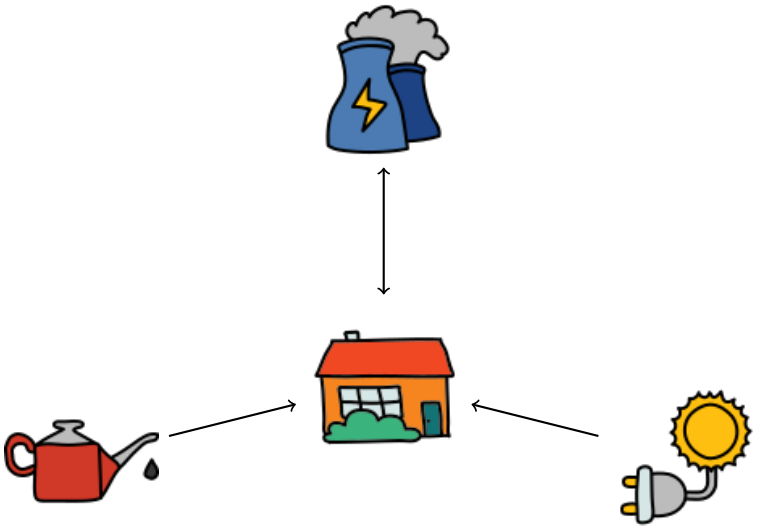

Figure 3.1: Illustration of the microgrid model and the relationships between its components [1]

The arrows represent the energy flux within the microgrid. The bidirectional edge in Figure 3.1 connecting the household to the public grid is justified by the possibility of selling energy to the public grid in case of energy surplus.

3.1.1 Residual Demand

Consumption and production are stochastic. The power balance equation is given by

$$X_s = cons - prod = d_s + g_s$$

The residual demand can always be met through the diesel generator, power from the public grid, or a combination of both. Scenarios, where the consumption can't be fully covered, i.e. $X_s > d_s + g_s$ are called blackouts [4] and are omitted. It is reasonable to consider X_s as a mean-reverting process with a cyclical mean process capturing the energy trend during day and night. Its dynamic is given by

$$dX_s = \mu(\Lambda_s - X_s)ds + \sigma dW_s \tag{3.1}$$

for a cyclical process Λ_s, mean-reverting speed $\mu > 0$, diffusion constant $\sigma > 0$ and some initial condition x at time t. It is a slightly modified Ornstein–Uhlenbeck process and X_t is the only state variable in the microgrid.

[1] Icons by https://www.icons8.com/

Note, that (3.1) is a linear SDE with an explicit solution (cf. Theorem 6.14 in Yong et al. [10]).

3.1.2 Diesel Generator

The diesel generator contributes to the power balance equation and is the only controlled variable in this model. It captures the amount of energy injected by the generator and is limited according to the microgrid, with $d_s \in D := [0, d_{max}]$ for all $s \in [0, T]$. The diesel generator needs $\rho(d_s) := \frac{(d_s-6)^3+6^3+d_s}{10}$ liters of crude oil at price $p_s \geq 0$ to produce d_s amount of power. Hence, the consumer pays $p_s\rho(d_s)$ for any diesel energy d_s generated at time s.

3.1.3 Price Processes

The price forecasts, c_s for grid energy and p_s for crude oil are deterministic. Their forecast is modelled by a Geometric Brownian motion (GBM) [16] of the form

$$dY_s = bY_s ds + \sigma Y_s dW_s$$

with $Y_s = \{c_s, p_s\}$ and are thus functions of time. In practice, they are generated by two particular non-negative realizations of the GBM. c_s and p_s can also be modelled as random processes but in that case, they enter as additional state variables into the later problem framework and increase the previously discussed dimensionality of X_s.

3.2 Problem Setting

This section integrates the components introduced in Section 3.1, formulates the value function and derives its corresponding HJB.

Let $(\Omega, \mathcal{F}, \mathcal{F}_t, \mathbb{P})$ for $t \in [0, T]$ be a filtered probability space with a filtration $\{\mathcal{F}_s\}_{0 \leq s \leq T}$ generated by a one-dimensional Brownian motion $\{W_s\}_{0 \leq s \leq T}$. The cost function f_γ is determined by

$$f_\gamma : [0, T] \times \mathbb{R} \times D \to \mathbb{R}, \ (s, x, \delta) \mapsto f_\gamma(s, x, \delta) = -p_s\rho(\delta) - c_s\gamma(\delta, x) \quad (3.2)$$

with several variations for $\gamma : D \times \mathbb{R} \to \mathbb{R}$:

$$\gamma(\delta, x) = \begin{cases} x - \delta, & \text{Linear} \\ (x-\delta)^2, & \text{Squared} \\ (x-\delta)_+, & \text{Maximum} \\ |x - \delta|, & \text{Absolute} \end{cases}$$

For each variant (3.2) satisfies the quadratic growth condition of Section 2.2. They have different interpretations and the first two versions are inspected in Section 3.2.1 and Section 3.2.2. d_s is a Markovian control of the form $d_s := d(s, X_s)$, where

$$d : [0, T] \times \mathbb{R} \to D, \ (s, x) \mapsto d(s, x)$$

Note, that d with arguments takes values in D, i.e. $d(\cdot, \cdot) = \delta \in [0, d_{max}]$, whereas d without arguments refers to the entire process in the space of admissible control strategies $\mathcal{D}(0, x)$. This space corresponds to $\mathcal{A}(0, x)$ in (2.10). The notation is changed to emphasize that the action space now only consists of diesel regulation. (3.1) admits a strong solution X_s that is pathwise unique, because (2.3) and (2.4) are satisfied. A trivial terminal condition 0 is imposed and an initial value (t, x) is fixed for t in $[0, T]$ and x in $\mathbb{R}$ arbitrary. The value function is consequently given by

$$v(t, x) = \sup_{d \in \mathcal{D}(t,x)} \mathbb{E}\left[\int_t^T -p_s\rho(d_s) - c_s\gamma(d_s, X_s)ds | X_t = x\right] \tag{3.3}$$

As the residual demand within a microgrid has physical limitations, it is reasonable to consider $v(t, x)$ only on a compact subspace $[0, T] \times [a, b] \subset\subset [0, T] \times \mathbb{R}$. Realizations of the residual demand might lie outside the compact subspace, due to the construction of $\{X_s\}_{t \le s \le T}$. Assuming, that $\partial[a, b]$ is a manifold of class $C^{(3)}$,

$$\begin{cases} -\frac{\partial}{\partial s}v(s, x) - \sup_{\delta \in D}\left(\mu(\Lambda_s - x)\frac{\partial}{\partial x}v(s, x) + \frac{\sigma^2}{2}\frac{\partial^2}{\partial x^2}v(s, x) + f_\gamma(s, x, \delta)\right) = 0 \\ v(T, x) = 0, \forall x \in [a, b] \end{cases} \tag{3.4}$$

has a unique solution $v \in C^{1,2}([0, T] \times [a, b])$ by Theorem 4.3 in Fleming et al. [17]. The theorem only holds for γ linear or quadratic. The optimal control policy is given by

$$\begin{aligned} d^* &= \arg\max_{\delta \in [0, d_{max}]}\left(\mu(\Lambda_s - x)\frac{\partial}{\partial x}v(s, x) + \frac{\sigma^2}{2}\frac{\partial^2}{\partial x^2}v(s, x) + f_\gamma(s, x, \delta)\right) \\ &= \arg\max_{\delta \in [0, d_{max}]} -p_s\rho(\delta) - c_s\gamma(\delta, x) \end{aligned} \tag{3.5}$$

for $\forall(s,x) \in [0,T] \times \mathbb{R}$. Due to the explicit form of ρ and γ, (3.5) is an optimization problem of a $C^\infty(D)$ function on a compact domain. Therefore, the maximum exists and is well-defined [18]. The optimal control policy is computed analytically in the following subsections.

3.2.1 Linear Cost Function

This version of γ is suitable when the microgrid can sell excess energy to the public grid.

Claim 3.1. *With $\gamma(\delta,x) = x - \delta$, the solution to (3.5) is given by*

$$d^*_{lin}(s,x) = \arg\max_{\delta\in\tilde{D}} f_\gamma(s,x,\delta)\mathbb{1}_{\{10c_s \geq p_s > 0\}} + d_{max}\mathbb{1}_{\{p_s=0\}}$$

with $(s,x) \in [0,T]\times[a,b]$ and $\tilde{D} := \{0, \min\{\sqrt{\frac{10\frac{c_s}{p_s}-1}{3}} + 6, d_{max}\}\}$.

Proof. Consider $f_\gamma(s,x,\delta) = -p_s\frac{(\delta-6)^3+6^3+\delta}{10} - c_s(x-\delta)$. As $f_\gamma \in C^\infty(D)$, it is continuously differentiable in δ and

$$\frac{\partial}{\partial\delta} f_\gamma(s,x,\delta) = -p_t\frac{3(\delta-6)^2+1}{10} + c_t \tag{3.6}$$

Case $p_s = 0$:
Then $\frac{\partial}{\partial\delta} f_\gamma(s,x,\delta) = c_s \geq 0$ and therefore f_γ must be increasing. Thus the maximizing argument is $d^*_{lin}(s,x) = d_{max}$. This validates the intuition, that if the price for diesel is 0, the consumer exploits the resource to minimize expenses and makes profits by selling energy surplus back to the public grid. The converse for $c_s = 0$ holds too.

Case $c_s = 0$:
Then $\frac{\partial}{\partial\delta} f_\gamma(s,x,\delta) = -p_s\frac{3(\delta-6)^2+1}{10} \leq 0$ and therefore f_γ must be decreasing. Hence, $d^*_{lin}(s,x) = 0$.

Case $10c_s < p_s$:
This leads to $\frac{\partial}{\partial\delta} f_\gamma(s,x,\delta) < -p_s\frac{3(\delta-6)^2+1}{10} + \frac{p_s}{10} = -p_s\frac{3(\delta-6)^2}{10} \leq 0$. Therefore, $f_\gamma(s,x,\delta)$ is strictly decreasing and $d^*_{lin}(s,x) = 0$.

Case $10c_s \geq p_s > 0$:
Setting (3.6) equal to zero and solving with respect to δ yields

$$\delta_\pm = \pm\sqrt{\frac{10\frac{c_s}{p_s}-1}{3}} + 6$$

Because ρ is cubic and has a "$-x^3$" behaviour only δ_+ is considered. If $\delta_+ \in D$, it is a local maximum as

$$\frac{\partial^2}{\partial \delta^2} f_\gamma(s,x,\delta)|_{\delta=\delta_+} = -p_s \frac{6}{10}(\delta - 6)|_{\delta=\delta_+} = -p_s \frac{6}{10}\sqrt{\frac{10\frac{c_s}{p_s} - 1}{3}} \leq 0$$

Another particularity in this case is $10c_s = p_s$. This results in an inflection point at $\delta = 6$, transitioning from concave upwards to concave downwards due to a change in sign. Overall,

$$d^*_{lin}(s,x) = \arg\max_{\delta \in \tilde{D}} f_\gamma(s,x,\delta), \;\; \tilde{D} := \{0, \min\{\sqrt{\frac{10\frac{c_s}{p_s} - 1}{3}} + 6, d_{max}\}\}$$

taking into account the possibility $\delta_+ \notin D$.

□

3.2.2 Quadratic Cost Function

This version is suitable when the primary source for meeting residual demand is the diesel generator, but using power from the grid is acceptable in specific situations.

Claim 3.2. *Given $\gamma(\delta, x) = (x - \delta)^2$, the solution to (3.5) is*

$$\begin{aligned} d^*_{quad}(s,x) =& f_\gamma(s,x,\delta^*)\mathbb{1}_{\{(36p_s - 20c_s)^2 \geq 12p_s(109p_s - 20c_s) \;\wedge\; p_s \neq 0\}} \\ &+ \min\{\max\{x,0\}, d_{max}\}\mathbb{1}_{\{p_s = 0\}} \end{aligned}$$

with $(s,x) \in [0,T] \times [a,b]$ and
$\delta^* := \min\{\max\{\frac{36p_s - 20c_s + \sqrt{(36p_s - 20c_s)^2 - 12p_s(109p_s - 20c_s x)}}{6p_s}, d_{max}\}, 0\}$.

Proof. Clearly $f_\gamma(s,x,\delta) = -p_s \frac{(\delta-6)^3 + 6^3 + \delta}{10} - c_s(x - \delta)^2$ is in $C^\infty(D)$. Hence,

$$\frac{\partial}{\partial \delta} f_\gamma(s,x,\delta) = -p_s \frac{3(\delta - 6)^2 + 1}{10} + 2c_s(x - \delta) \tag{3.7}$$

Case $p_s = 0$:
Then $f_\gamma(s,x,\delta) = -c_s(x-\delta)^2$ and $\frac{\partial}{\partial \delta} f_\gamma(s,x,\delta) = 2c_s(x-\delta)$. Thus $d^*_{quad}(s,x) = \min\{\max\{x,0\}, d_{max}\}$.

Case $(36p_s - 20c_s)^2 \geq 12p_s(109p_s - 20c_s x)$:
Setting (3.7) equal to zero and solving with respect to δ yields

$$\begin{aligned} & 3p_s\delta^2 - (36p_s - 20c_s)\delta + 109p_s - 20c_s x = 0 \\ \implies \delta_\pm = & \frac{36p_s - 20c_s \pm \sqrt{(36p_s - 20c_s)^2 - 12p_s(109p_s - 20c_s x)}}{6p_s} \end{aligned}$$

Due to the cubic beviour of ρ only δ_+ is considered and the maximimzing argument is $d^*_{quad}(s,x) = f_\gamma(s,x,\delta^*)$, with $\delta^* := \min\{\max\{\delta_+, 0\}, d_{max}\}$.

Case $(36p_s - 20c_s)^2 < 12p_s(109p_s - 20c_s x)$:
There aren't any local minima/maxima and $f_\gamma(s,x,\delta)$ must be decreasing, which induces $d^*_{quad}(s,x) = 0$. This case implicitly contains $c_s = 0$, as $0 < 36^2 p_s^2 = 1296 p_s^2 < 1308 p_s^2 = 12 \cdot 109 p_s^2$

□

3.3 Theoretical Solution to Hamilton-Jacobi-Bellman equation

So far, (3.1) has only been defined for a non-zero diffusion constant σ. In the limit case $\sigma = 0$ an exact solution for the value function can be reproduced and the corresponding approach is given in the following. Note, that the stochastic diffusion part vanishes, which gives rise to a deterministic problem. The following result is presented to broaden the view of stochastic optimal control problems and to observe that certain deterministic control problems can be modelled as restrictions to stochastic control problems. Given the pointwise optimal control δ^* of the optimal control policy d^*, (3.4) takes the form

$$\begin{cases} \frac{\partial}{\partial s} v(s,x) + \mu(\Lambda_s - x)\frac{\partial}{\partial x} v(s,x) + f_\gamma(s,x,\delta^*) = 0 \\ v(T,x) = 0 \end{cases}$$

for different γ with $(s,x) \in [0,T] \times [a,b]$. The drift function satisfies an Ordinary Differential Equation (ODE) given by $y : [t,T] \to \mathbb{R}$:

$$\begin{cases} \dot{y}(s) = \mu(\Lambda_s - x) \\ y(t) = x \end{cases}$$

Define $\tilde{v}(s) := v(s, y(s)),\ s \in [t, T]$. Assuming its differentiability in s yields

$$\begin{aligned}\frac{\partial}{\partial s}\tilde{v}(s) &= \frac{\partial}{\partial s}v(s, y(s)) + \frac{\partial}{\partial x}v(s, y(s))\dot{y}(s) \\ &= \frac{\partial}{\partial s}v(s, y(s)) + \mu(\Lambda_s - x)\frac{\partial}{\partial x}v(s, y(s)) \\ &= -f_\gamma(s, y(s), \delta^*)\end{aligned}$$

$\tilde{v}(T) = v(T, y(T)) = 0$ and the Fundamental Theorem of Calculus lead to

$$\tilde{v}(s) = v(s, y(s)) = \int_t^T f_\gamma(s, y(s), d(s, y(s)))ds$$

The idea isn't carried on; instead $\sigma > 0$ is examined.

3.4 Numerical Solutions to Hamilton-Jacobi-Bellman equation

Two approaches are examined to compute a numerical approximation to the value function (3.3). The first one uses FDM leading to the problem, that the boundary conditions at a and b are unknown. However, extending the DoI circumvents the issue at the cost of higher computational time. The second approach is based on a MC simulation. Therefore, numerous trajectories of the mean-reverting process X_t are sampled and the average of the approximated integrals is taken. A theoretical convergence analysis is performed, suggesting convergence rates and discussing their validity.

3.4.1 Finite Difference Method

The microgrid introduced in Section 3.4 only deals with one space variable. Sofar, (3.4) is an extended Advection-Diffusion equation [19], which is usually defined as

$$\begin{cases}\frac{\partial}{\partial s}v(s, x) + \varepsilon u\frac{\partial}{\partial x}v(s, x) - \xi\frac{\partial^2}{\partial x^2}v(s, x) = f(s, x) \\ v(0, x) = g(x), \forall x \in [a, b] \\ v(s, a) = h_a(s), v(s, b) = h_b(s)\ \forall s \in [0, T]\end{cases}$$

for some functions g, h_a, h_b that prescribe the boundary conditions on a truncated space. The PDE is defined forward in time. ε and u are physical parameters. They aren't essential for the thesis and consequently aren't further

elaborated. In Section 3.2, $v(T,x)=0$ is imposed and $\xi := -\frac{\sigma^2}{2} < 0$. Recall (3.4)

$$\begin{cases} \frac{\partial}{\partial s}v(s,x) + \mu(\Lambda_s - x)\frac{\partial}{\partial x}v(s,x) + \frac{\sigma^2}{2}\frac{\partial^2}{\partial x^2}v(s,x) = -\sup_{\delta\in D} f_\gamma(s,x,\delta) \\ v(T,x) = 0, \forall x \in [a,b] \end{cases}$$

Time and space are discretized for computational purposes. Therefore, let $t_n = n\Delta t$ for $\forall n \in \{0,\dots,N\}$ and $x_i = x_0 + ih,\ \forall i \in \{0,\dots,I\}$ such that $0 = t_0 < t_1 < \cdots < t_{N-1} < t_N = T$ and $a = x_0 < x_1 < \cdots < x_{I-1} < x_I = b$ for $\{a,b\} \in \mathbb{R}$. Δt denotes the time step and h denotes the spatial step. The numerical approximation of $v(t_n,x_i)$ is denoted by $\hat{v}_i^n$. Occasionally, the subscript $[\cdot]_{t_n}$ for time-dependent processes is changed to $[\cdot]_n$. The Forward Euler method (FE) as in Quarteroni et al. [20] is applied to (3.4) with a central finite difference approach on $\frac{\partial}{\partial x}$ and $\frac{\partial^2}{\partial x^2}$. The FE requires to discretize the HJB formulation as follows.

$$\begin{aligned} \frac{\hat{v}_i^{n+1} - \hat{v}_i^n}{\Delta t} =& -\frac{\sigma^2}{2}\frac{\hat{v}_{i+1}^n - 2\hat{v}_i^n + \hat{v}_{i-1}^n}{h^2} - \mu(\Lambda_n - x_i)\frac{\hat{v}_{i+1}^n - \hat{v}_{i-1}^n}{2h} \\ & - f_\gamma(t_n, x_i, d^*(t_n,x_i)) \\ \hat{v}_i^{n+1} =& \hat{v}_i^n - \frac{\Delta t\sigma^2}{2h^2}(\hat{v}_{i+1}^n - 2\hat{v}_i^n + \hat{v}_{i-1}^n) - \frac{\Delta t\mu(\Lambda_n - x_i)}{2h}(\hat{v}_{i+1}^n - \hat{v}_{i-1}^n) \\ & - \Delta t f_\gamma(t_n, x_i, d^*(t_n,x_i)) \end{aligned}$$

for $\forall n \in \{0,\dots,N\}$ and $\forall i \in \{0,\dots,I\}$. The corresponding matrix representation is provided by

$$A := \begin{pmatrix} -2 & 1 & \\ 1 & \ddots & 1 \\ & 1 & -2 \end{pmatrix},\ C_n := \mu \begin{pmatrix} \Lambda_n - x_1 & & \\ & \ddots & \\ & & \Lambda_n - x_{I-1} \end{pmatrix} \begin{pmatrix} 0 & 1 & \\ -1 & \ddots & 1 \\ & -1 & 0 \end{pmatrix}$$

$$\begin{aligned} \begin{pmatrix} \hat{v}_1^{n+1} \\ \vdots \\ \hat{v}_{I-1}^{n+1} \end{pmatrix} =& \underbrace{\left(\mathbb{1} - \frac{\Delta t\sigma^2}{2h^2}A - \frac{\Delta t}{2h}C_n\right)}_{B:=} \begin{pmatrix} \hat{v}_1^n \\ \vdots \\ \hat{v}_{I-1}^n \end{pmatrix} \\ & - \Delta t \begin{pmatrix} f_\gamma(t_n, x_1, d^*(t_n,x_1)) + \left(\frac{\sigma^2}{2h^2} - \mu\frac{\Lambda_n - x_1}{2h}\right)\hat{v}_0^n \\ \vdots \qquad\qquad + \qquad 0 \\ f_\gamma(t_n, x_{I-1}, d^*(t_n,x_{I-1})) + \left(\frac{\sigma^2}{2h^2} + \mu\frac{\Lambda_n - x_{I-1}}{2h}\right)\hat{v}_I^n \end{pmatrix} \end{aligned} \tag{3.8}$$

Note, that the boundary conditions of the solution v, i.e. $v(t_n, a)$ and $v(t_n, b)$, $\forall n = 0, \ldots, N$ are unknown and that there is no reasonable way to guess them [1]. The FE turns into an implicit approach backwards in time. Therefore, a system of linear equations needs to be solved at every time step dissolving the explicit dependency. Any error from the forcing term is then passed onto the next space layer. This suggests, that errors on the boundary propagate throughout the implementation returning a potentially inaccurate approximation of v. Such an issue could be solved by an explicit method whose domain is extended so that the error doesn't touch the DoI. This idea works due to the banded characteristic of A and C_n inferring that any point (t_k, x_i) inside the DoI only needs data from the value function in $(t_{k+1}, x_{i-1}), (t_{k+1}, x_i), (t_{k+1}, x_{i+1})$. The discussion is clarified in Figure 3.2.

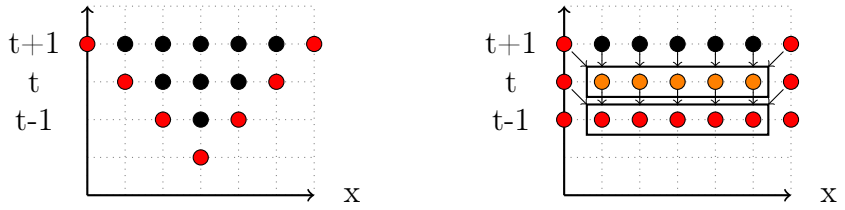

Figure 3.2: Error propagation of explicit FDM (left) and implicit FDM (right). The red points allude to inaccurate approximations to v, whereas the black points indicate accurate ones

The explicit method seems to be a reasonable solution to the problem induced by the unknown boundaries until the stability criteria are taken into account. In the context of numerical analysis, stability criteria are constraints that need to be satisfied during the implementation of numerical processes to guarantee reliable results. They are important for the accuracy and correctness of numerical simulations. Failure to meet them can lead to numerical instability, divergence, or oscillations in the results [20]. In this context, stability is granted when the discretization sizes h and Δt are small enough. Chapter 9 in Quarteroni et al. [20] states, that oscillations vanish in steady-state convection-diffusion problems if

$$h < \frac{\sigma^2}{\Lambda_n - x_i}$$

$\forall n = 0, \ldots, N, i = 0, \ldots, I$. In the same book, a classical heat equation remains stable if

$$\Delta t < \frac{h^2}{\sigma^2}$$

Therefore, it seems reasonable to consider both criteria simultaneously for the discretization of (3.4). The quadratic dependency of Δt on h and the fact, that the extended domain needs to be chosen large enough, make this approach computationally inefficient.
However, the error propagation sketched for the implicit method in Figure 3.2 is not as global as the figure suggests. It propagates locally on the boundaries. This discovery was made by empirical testing and no rigorous argument is presented. Therefore, extending the DoI and applying the implicit method is sufficient for an accurate approximation to v. Even though the procedure is unconditionally stable [20], the local Peclet number can't be neglected and is elaborated later. In short, it is a measure that measures how much the drift term prevails over the diffusive one. It hints, at whether the approximation to the value function by a FDM might exhibit oscillations [20]. The so-called upwind scheme circumvents this issue [20] by using a forward difference approach for $\frac{\partial}{\partial x}$.
Overall, the previous discussion suggests a pseudo-algorithm of the form

Algorithm 1: Implicit Finite Difference Method backwards in time

Discretize $[0, T] \times [\tilde{a}, \tilde{b}]$
Generate $p_{t_n}, c_{t_n}, \Lambda_{t_n}, \ \forall n = 0, \dots, N$
for $t_n \in [T, \dots, t_1]$ **do**
  Evaluate $f_\gamma(t_{n-1}, x_i, d^*(t_{n-1}, x_i))$ on $x_i \in \{x_1, \dots, x_{\tilde{I}-1}\}$
  Update C_{n-1}
  Compute $(\hat{v}_1^{n-1}, \dots, \hat{v}_{\tilde{I}-1}^{n-1})^T$ with $\hat{v}_0^n, \hat{v}_{\tilde{I}}^n = 0, \forall n = 0, \dots, N$
end

$\tilde{I}$ is the number of discretizations for $[\tilde{a}, \tilde{b}] := [2a, 2b]$ with the same discretization size h. $[\tilde{a}, \tilde{b}]$ is a cover of $[a, b]$ as $a < 0$ and $b > 0$ in the numerical implementation. If the upwind scheme is applied, C_n changes to

$$C_n = \mu \begin{pmatrix} \Lambda_n - x_1 & & \\ & \ddots & \\ & & \Lambda_n - x_{I-1} \end{pmatrix} \begin{pmatrix} -1 & 1 & \\ & \ddots & 1 \\ & & -1 \end{pmatrix}$$

with a corresponding modification to (3.8). For the complexity note, at every time step - $N+1$ in total - two vectors of length $I-1$ are added and a system of linear equations is solved. Algorithm 1 was implemented employing sparse arrays to enhance the computational efficiency. The theoretical convergence for known boundary terms is carried out in the following.

Claim 3.3. *Given $v \in C^{2,4}$, Algorithm 1 has convergence order*

(i) $\mathcal{O}(\Delta t) + \mathcal{O}(h^2)$, using a central finite difference approximation for $\frac{\partial}{\partial x}$

(ii) $\mathcal{O}(\Delta t) + \mathcal{O}(h)$*, using a forward difference approximation for* $\frac{\partial}{\partial x}$

Discussion. Suppose $v \in C^{2,4}$. Taylor expansion then yields

$$v(s, x+h) = v(s,x) + \frac{\partial}{\partial x} v(s,x)h + \frac{\partial^2}{\partial x^2} v(s,x)\frac{h^2}{2} + \frac{\partial^3}{\partial x^3} v(s,x)\frac{h^3}{6}\mathcal{O}(h^4)$$
$$v(s, x-h) = v(s,x) - \frac{\partial}{\partial x} v(s,x)h + \frac{\partial^2}{\partial x^2} v(s,x)\frac{h^2}{2} - \frac{\partial^3}{\partial x^3} v(s,x)\frac{h^3}{6} + \mathcal{O}(h^4)$$

for $\forall (s,x) \in [0,T] \times \mathbb{R}$. The first space derivative is given by

$$\frac{v(s,x+h) - v(s,x-h)}{2h} = \frac{\partial}{\partial x} v(s,x) + \mathcal{O}(h^2)$$

for a central finite difference approximation approach and by

$$\frac{v(s,x+h) - v(s,x)}{h} = \frac{\partial}{\partial x} v(s,x) + \mathcal{O}(h)$$

for a forward approximation approach on $\frac{\partial}{\partial x} v$. An analogous argument is made for $\frac{\partial}{\partial s} v$. The second-space derivative is of the form

$$\frac{v(s,x+h) - 2v(s,x) + v(s,x-h)}{h^2} = \frac{\partial^2}{\partial x^2} v(s,x) + \mathcal{O}(h^2)$$

Until now the convergence rates only hold pointwise. By starting with an exact terminal condition, every iteration of Algorithm 1 drags an error of $\mathcal{O}(\Delta t) + \mathcal{O}(h^2)$ or $\mathcal{O}(\Delta t) + \mathcal{O}(h)$ in case of the upwind scheme, giving rise to the claim. □

Overall, the upwind scheme prevents oscillations but comes with the prize of slower convergence in space. The claim has assumed higher regularity on v. Increasing smoothness is common in these kinds of numerical convergence settings, where it is desirable to apply Taylor-Expansion to get hold of the potential convergence behaviour [21]. Hence, the above convergence rates need to be treated with caution when solving an (3.4) via Algorithm 1

3.4.2 Monte Carlo

The idea requires three distinct discretizations. Indeed, given d^* for different γ an approximation to

$$v(t_n, x_i) = \mathbb{E}\left[\int_{t_n}^{T} -p_s \rho(d_s^*) - c_s \gamma(d_s^*, X_s) ds | X_{t_n} = x_i\right]$$

is needed. Therefore, $\mathbb{E}[\cdot]$ is approximated by a standard MC simulation. The Left Rectangle Approximation Method (LRAM) is chosen to discretize $\int_{t_n}^{T}$, because of the terminal condition. The dynamics of the residual demand (3.1) are simulated with Euler Maruyama (EM). Even though an exact solution exist, EM is used to be consistent with the implementation in Chapter 4. Overall, these choices lead to

$$\hat{v}_i^n := \frac{1}{M} \sum_{m=1}^{M} \sum_{k=n}^{N-1} \left(-p_{t_k} \rho(d_{t_k}^*) - c_{t_k} \gamma(d_{t_k}^*, X_{t_k}^{(m)}) \right) \Delta t \tag{3.9}$$

given $X_{t_n}^{(m)} = x_i \; \forall m = 1, \ldots, M$. Recall, that $d_{t_k}^*$ denotes the optimal control decision in time t_k for $X_{t_k}^{(m)}$, $m = 1, \ldots, M$. Note the abuse of notation for $\hat{v}_i^n$ with respect to Section 3.5.1. The explicit form of (3.9) translates into Algorithm 2.

Algorithm 2: Monte Carlo

```
Discretize [0, T] × [a, b]
Generate p_{t_n}, c_{t_n}, Λ_{t_n}, ∀n = 0, ..., N
for t_n ∈ [0, ..., T] do
    for x_i ∈ [a, ..., b] do
        Generate M samples of
          {X_{t_k}^{(m)} | X_{t_n}^{(m)} = x_i}_{n≤k≤N} ∀m = 1, ..., M
        Evaluate value function
            v̂_i^n := (1/M) Σ_{m=1}^{M} Σ_{k=n}^{N-1} ( -p_{t_k} ρ(d*_{t_k}) - c_{t_k} γ(d*_{t_k}, X_{t_k}^{(m)}) ) Δt
    end
end
```

The overall complexity results in the amount of time steps $N + 1$ times the amount of space points $I + 1$. Moreover, M trajectories have to be sampled at every grid point (t_n, x_i) from t_n until the final time T. The optimal control policy has to be computed at every point of the trajectories. The cost function is then evaluated with the optimal control policies and trajectories. Its average yields an approximation to the value function in a specific point (t_n, x_i) of the DoI.

The inner for-loop might be questionable because in practice v only needs to be computed for a specific space state. Hence, iterating over the whole DoI increases the complexity. Suitable vectorization doesn't circumvent this issue entirely. However, this increase in complexity allows to asses the two

methods across the whole DoI. Its pointwise convergence is quantified in the following claim.

Claim 3.4. *Let* $v(t_n, x_i) = \mathbb{E}\left[\int_{t_n}^{T} -p_s\rho(d_s^*) - c_s\gamma(d_s^*, X_s)ds | X_{t_n} = x_i\right]$ *and* $\hat{v}_i^n := \frac{1}{M}\sum_{m=1}^{M}\sum_{k=n}^{N-1}\left(-p_{t_k}\rho(d_{t_k}^*) - c_{t_k}\gamma(d_{t_k}^*, X_{t_k}^{(m)})\right)\Delta t$ *with* $X_{t_n}^{(m)} = x_i$ *for* $\forall m = 1, \ldots, M$. *Additionally, let* $i \in [0, \ldots, I]$ *and* $n \in [0, \ldots, N]$ *arbitrarily. Then*

$$|v(t_n, x_i) - \hat{v}_i^n| \leq \mathcal{O}(\Delta t^p) + \mathcal{O}(\frac{1}{\sqrt{M}})$$

The convergence rate $p \in \mathbb{R}_{>0}$ *is yet to be determined.*

Discussion. The idea is to split the three discretizations and examine the convergence separately. Define

$$\hat{v}(t_n, x_i) := \mathbb{E}\left[\int_{t_n}^{T} -p_{t_s}\rho(d_{t_s}^*) - c_{t_s}\gamma(d_{t_s}^*, X_{t_s})ds\right]$$
$$v_i^n := \mathbb{E}\left[\sum_{k=n}^{N-1}\left(-p_{t_k}\rho(d_{t_k}^*) - c_{t_k}\gamma(d_{t_k}^*, X_{t_k})\right)\Delta t\right]$$

The subscript t_s in X_{t_s} of the first definition $\hat{v}(t_n, x_i)$ is used to indicate that X_{t_s} is a continuous approximation to X_s, for instance a linearly interpolated version of X_{t_k} as done in Lecture 9 of Blumenthal [22]. Therefore, both processes coincide in $t_s = t_k$. For interpretation purposes, t_s is equivalent to $s, \forall s \in [t_n, T]$ but denoted differently to clarify the previous distinction. The $\triangle$-inequality yields

$$|v(t_n, x_i) - \hat{v}_i^n| \leq |v(t_n, x_i) - \hat{v}(t_n, x_i)| + |\hat{v}(t_n, x_i) - v_i^n| + |v_i^n - \hat{v}_i^n|$$

The first one is the difference between two expected values with different random variables yielding a weak error [22].
Note, that the drift and diffusion term of the residual demand (3.1) satisfy the growth condition of Lemma 2 in Lecture 9 of Blumenthal [22] and as a consequence $\{X_{t_s}\}_{t_n \leq t_s \leq T} \to \{\hat{X}_s\}_{t_n \leq s \leq T}$ in $L^2([t_n, T])$ with order 1, where $\hat{X}_s = \sum_{k=n}^{N} X_{t_{k-1}} \mathbb{1}_{(t_{k-1}, t_k]}(s)$. The newly introduced process $\{\hat{X}_s\}_{t_n \leq s \leq T}$ is a formal notation for the step process given by the solution of the EM method. In Lecture 10 of Blumenthal [22] the classical strong convergence of EM with order 1/2 is proved, i.e. $\{\hat{X}_s\}_{t_n \leq s \leq T} \to \{X_s\}_{t_n \leq s \leq T}$ in $L^2([t_n, T])$. To apply this result Hölder continuity with $\alpha = 1/2$ of the mean-reverting process is

required.

A real-valued function g is said to be Hölder continuous with exponent $0 < \alpha \leq 1$ on a d-dimensional Euclidean space S [23], when there exist $C > 0$, such that for $\forall x \neq y$ and $x, y \in S$

$$|g(x) - g(y)| \leq C||x - y||_2^{\alpha}$$

The mean-reverting process is $\Lambda_s = 6\sin(\pi s)$. The Mean value theorem and the differentiability of Λ_s in s lead to

$$\frac{|\Lambda_r - \Lambda_s|}{|r-s|^{1/2}} \leq \frac{|\frac{\partial}{\partial u}\Lambda_u||r-s|}{|r-s|^{1/2}} \leq 6\pi|r-s|^{1/2} \leq 6\pi T^{1/2},$$

where $s \neq r$ and $s, r \in [0, T]$. u is between s and r. The last inequality comes from the increasing behaviour of the square root function. With the Hölder condition validated, Lecture 10 of Blumenthal [22] allows to deduce

$$\mathbb{E}[(X_s - X_{t_s})^2] \leq 2\mathbb{E}[(X_s - \hat{X}_s)^2] + 2\mathbb{E}[(X_{t_s} - \hat{X}_s)^2] \to 0$$

for $\Delta t \searrow 0$ with order $1/2$. This result implies convergence in distribution, hence $X_{t_s} \xrightarrow{\mathcal{D}} X_s$ [24]. Denote

$$\psi_{n,i} : C_{n,i} \to \mathbb{R},\ y \mapsto \psi_{n,i}(y) = \int_{t_n}^{T} -p_s\rho(d^*(s, y_s)) - c_s\gamma(d^*(s, y_s), y_s)ds,$$

where $C_{n,i} := \{y \in C[t_n, T] |\ y_{t_n} = x_i\}$ contains all continuous processes y in $[t_n, T]$ with start in x_i. Assume that $\psi_{n,i}$ is a continuous functional in functional analysis sense. The Continuous Mapping Theorem [24] states

$$X_{t_s} \xrightarrow{\mathcal{D}} X_s \iff \mathbb{E}\left[g(X_{t_s})\right] \to \mathbb{E}\left[g(X_s)\right], \text{ for every continuous, bounded functional } g$$

$\psi_{n,i}$ needs to be identified with g. It is done by declaring it as bounded on a compact space large enough following the same argument as in the derivation of the HJB in Section 2.3. Hence,

$$|v(t_n, x_i) - \hat{v}(t_n, x_i)| = |\mathbb{E}\left[\psi_{n,i}(\{X_s^{t_n,x_i}\}_{t_n \leq s \leq T})\right] - \mathbb{E}\left[\psi_{n,i}(\{X_{t_s}^{t_n,x_i}\}_{t_n \leq t_s \leq T})\right]| \to 0$$

for $\Delta t \searrow 0$ at an unknown convergence rate p.

The middle term $|\hat{v}(t_n, x_i) - v_i^n|$ conveys the discretization error from the integral approximation. Note, that

$$\begin{aligned} & |\int_{t_n}^{T} f_\gamma(t_s, X_{t_s}, d_{t_s})ds - \sum_{k=n}^{N-1} f_\gamma(t_k, X_{t_k}, d_{t_k})\Delta t| \\ =& |\sum_{k=n}^{N-1} \left(\int_{t_k}^{t_{k+1}} f_\gamma(t_s, X_{t_s}, d_{t_s})ds - f_\gamma(t_k, X_{t_k}, d_{t_k})\Delta t \right)| \\ =& |\sum_{k=n}^{N-1} \left(\int_{t_k}^{t_{k+1}} f_\gamma(t_s, X_{t_s}, d_{t_s}) - f_\gamma(t_k, X_{t_k}, d_{t_k})ds \right)| \end{aligned} \tag{3.10}$$

Define

$$\begin{aligned} \vartheta_{\Delta t} : [t_n, T] \times \Omega \to \mathbb{R},\ (t_s, \omega) \mapsto \vartheta_{\Delta t}(t_s, \omega) = f_\gamma(t_s, X_{t_s}(\omega), d_{t_s}) \\ \vartheta : [t_n, T] \times \Omega \to \mathbb{R},\ (s, \omega) \mapsto \vartheta_{\Delta t}(s, \omega) = f_\gamma(s, X_s(\omega), d_s) \end{aligned}$$

The dependency of $\vartheta_{\Delta t}$ on Δt comes from the discretization of the EM scheme. Assume, that ϑ is Hölder continuous in the first argument. Then $\vartheta_{\Delta t}$ is also Hölder continuous for a $\alpha_{\Delta t,\omega} \in (0, 1]$ with Δt arbitrary small. Indeed, define the Hölder norm for $\vartheta_{\Delta t}$ and ϑ as follows

$$\begin{aligned} ||\vartheta_\Delta||_{H^{\alpha_{\Delta,\omega}}} &:= \sup_{\substack{t_s, \in [t_n, T],\ t_k, \in [t_n, \ldots, T] \\ t_s \neq t_k}} \frac{|\vartheta_{\Delta t}(t_s, \omega) - \vartheta_{\Delta t}(t_k, \omega)|}{|t_s - t_k|^{\alpha,\omega}} = C_{\vartheta_{\Delta t},\omega} < \infty \text{ a.s.} \\ ||\vartheta||_{H^{\alpha\omega}} &:= \sup_{\substack{s,t, \in [t_n, T] \\ s \neq t}} \frac{|\vartheta(s, \omega) - \vartheta(t, \omega)|}{|s - t|^{\alpha,\omega}} < \infty \text{ a.s.} \end{aligned}$$

The previous statement then holds because of

$$||\vartheta_\Delta||_{H^{\alpha_{\Delta,\omega}}} \leq ||\vartheta||_{H^{\alpha,\omega}}$$

$C_{\vartheta_{\Delta t},\omega}$ depends on Δt in the sense that $t_k \in [t_n, \ldots, T]$. Therefore, refining time leads to a change in the bounding constant.
Applying the Hölder characteristic of $\vartheta_{\Delta t}$ to (3.10) for Δt arbitrary small

leads to

$$\begin{aligned}
&C_{\vartheta_{\Delta t,\omega}} \left| \sum_{k=n}^{N-1} \left(\int_{t_k}^{t_{k+1}} (t_s - t_k)^{\alpha_{\Delta t,\omega}} ds \right) \right| \\
=&\frac{C_{\vartheta_{\Delta t,\omega}}}{\alpha_{\Delta t,\omega}+1} \left| \sum_{k=n}^{N-1} (t_{k+1} - t_k)^{\alpha_{\Delta t,\omega}+1} \right| \\
=&\frac{C_{\vartheta_{\Delta t,\omega}}}{\alpha_{\Delta t,\omega}+1} \frac{1}{N^{\alpha_{\Delta t,\omega}+1}} (N-1-n) \\
\leq&\frac{||\vartheta||_{H^{\alpha,\omega}}}{\alpha_{\Delta t,\omega}+1} \frac{1}{N^{\alpha_{\Delta t,\omega}+1}} (N-1-n) \text{ a.s.}
\end{aligned}$$

As $||\vartheta||_{H^{\alpha,\omega}}$ is supposed to be bounded a.s., it is possible to apply the DOM, which leads to $|\hat{v}(t_n, x_i) - v_i^n| \to 0$ for $\Delta t \searrow 0$. The overall convergence rate p is yet to be determined.

The rightmost term is a standard piece of theory for MC simulations and yields convergence rate $\mathcal{O}(\frac{1}{\sqrt{M}})$ [6]. □

3.5 Numerical Analysis

The section presents numerical results illustrating the problem setting elaborated in Section 3.2. Moreover, it inspects the practical convergence rates of Algorithm 1 and Algorithm 2. As the error is compared across the whole domain of interest, a reasonable uniform error measure is introduced beforehand. Two different types of convergence settings are considered. In the first one, the algorithm is performed on a known function and the exact error can be computed along the grid for different discretization sizes. In the second set-up, the algorithms are run for the quadratic version of the cost function (3.2). As the exact solution to (3.4) isn't available, a reference solution is computed on an "overkilling" discretization and compared to solutions obtained with coarser discretization sizes.

The results obtained in this section and in Chapter 5 have been computed with the files contained in https://git.scc.kit.edu/uuyog/bachelor-thesis.

Illustrating the formal framework The final time is set to $T = 2$, which leads to a time interval $[0, 2]$. The final time is chosen such that the mean process of the residual demand runs for one cycle. The state space is $[-6, 6]$. The overall DoI is then given by $[0, 2] \times [-6, 6]$. The compact control space

is given by $D = [0, 7]$. The different processes are simulated by EM with the following parameters:

- Public energy price forecast $\begin{cases} dc_s = 0.08 c_s ds + 0.3 c_s dW_s \\ c_o = 1 \end{cases}$
- Crude oil price forecast $\begin{cases} dp_s = 0.08 p_s ds + 0.8 p_s dW_s \\ p_o = 1.25 \end{cases}$
- Residual demand $\begin{cases} dX_s = (6 \sin(\pi s) - X_s)\, ds + \sigma dW_s \\ X_t = x \end{cases}$

The mean-reverting speed for the residual demand is therefore 1. The diffusion constant σ varies for different simulations. The cost processes are run beforehand and need to be positive at all times. A random seed fixes the forecasts and the MC simulations. Moreover, the cost processes are consistently simulated for $\Delta t = 0.001$ and sliced accordingly for coarser time discretizations. Algorithm 1 and Algorithm 2 give two approximations to the value function (3.3) and their results are displayed in Figure 3.3.

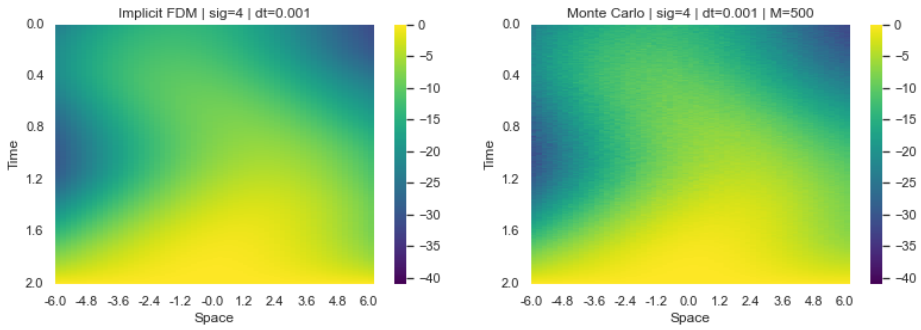

Figure 3.3: Heatmap of the value function given by implicit FDM (left) and MC simulation (right), with $\sigma = 4$ and discretization sizes $\Delta t = 0.001, h = 0.2$. The MC simulation used 500 trajectories for every point in $[0, \dots, 2] \times [-6, \dots, 6]$

The difference of the two approximations is 0.4693 in the $||\cdot||_{gl}$ error norm. This norm is defined and explained in the subsequent paragraph. Clearly, the value function is negative everywhere, as the integrated cost function itself is negative. Its particular shape comes from the residual demand, which constantly reverts to the mean process $6 \cdot \sin(\pi s)$, $s \in [0, 2]$. For most parts, the value function seems to increase in time and attains the maximum in the

terminal condition 0. In the negative space part, however, it first decreases and then increases once the time 1 is surpassed. Note, that $\sigma = 4$ is a high diffusion constant compared to the value of the mean process. The behaviour is influenced by the optimal control policy, which is presented in Figure 3.4.

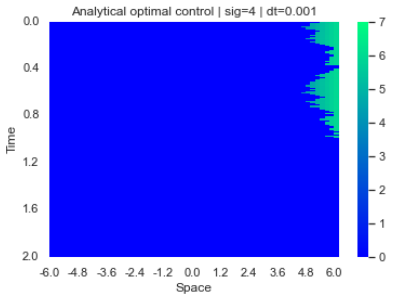

Figure 3.4: Heatmap of analytical optimal control policy computed as in 3.2.2 for $\sigma = 4$ and discretization sizes $\Delta t = 0.001, h = 0.2$

The control policy in Figure 3.4 is predominantly 0, especially in the negative space domain. The latter observation comes from the fact, that $-p_s\rho(\delta) - c_s(x-\delta)^2 \leq -c_s x^2$, $\forall s \in [0,2]$ and $\delta \in [0,7]$. In the positive part of the DoI, the control policy jumps very fast to the maximal diesel production, suggesting that the control process is not continuous. The optimal control policy exclusively depends on the price forecasts and changes once they change. The fixed price processes of the model are illustrated in Figure 3.5. The same figure features an example trajectory for the residual demand process and its corresponding control policy.

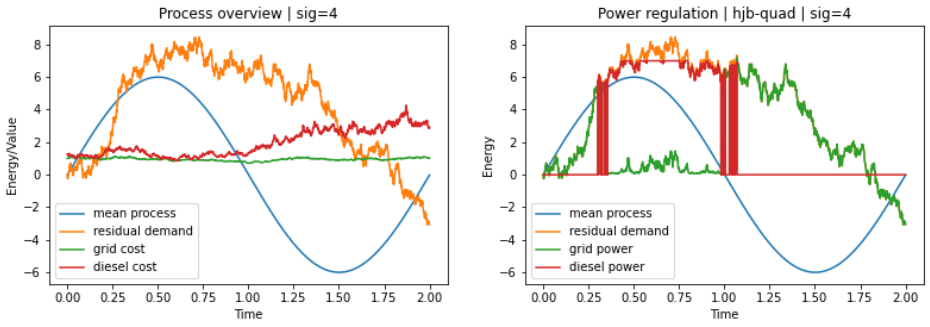

Figure 3.5: Overview of the involved processes (left) and the power regulation within the microgrid (right) for $\sigma = 4$ and discretization size $= 0.001$. The price processes have to be interpreted on a price scale, whereas the mean process and the residual demand are some kind of energy units

The y-scale for the processes on the left of Figure 3.5 has two different units of measurement. The mean process and the residual demand process are measured in energy, whereas the grid cost and diesel cost are measured with a price tag. The different amounts of diffusion within the cost processes at equal drift rates let them diverge. On the right plot, the residual demand is covered by the grid power if the consumer only requires a moderate amount of energy. The diesel generator is turned on in high-energy deficiency settings. It seems that the power provided by the generator comes in thrusts instead of continuous trends.

The section continues by looking at the convergence results for Algorithm 1 and Algorithm 2. Therefore, an error measure is needed and a global one is proposed. It measures the error of the approximations of the value function evaluated on a two-dimensional grid representing the discretized DoI. This allows to quantify the accuracy of the Algorithms across the whole DoI at the cost of pointwise dilution.

Introducing a global error norm Claim 3.3 and Claim 3.4 have given pointwise convergence rates across the whole domain of interest $[0, \dots, T] \times [a, \dots, b]$. An error measure, that captures the error along the DoI is the square root of the two-dimensional Mean squared error (MSE) and is defined as

$$|| \cdot ||_{gl} := \frac{|| \cdot ||_F}{\sqrt{(N+1)(I+1)}},$$

where $|| \cdot ||_F$ is the Frobenius norm, i.e. the L^2-norm for a matrix $(B_{n,i})_{n,i=0}^{N,I}$

$$||M||_F = \sqrt{\sum_{n,i=0}^{N,I} B_{n,i}^2}$$

$||\cdot||_{gl}$ is a norm, as it is subadditive, positive definite and satisfies the absolute homogeneity condition. Note that, $|| \cdot ||_{gl}$ depends on Δt and h. Those parameters are omitted in the notation as they are provided by the specific context at hand. If the true value function v is known, the error of its approximation $\hat{v}$ along the discretized DoI is computed as

$$||\hat{v} - v||_{gl}$$

In this notation, the argument of $|| \cdot ||_{gl}$ is a matrix with shape $[0, \dots, 2] \times [a, \dots, b]$.. Assuming that (3.4) can't be solved analytically, v is substituted

by a best-cast solution $\hat{v}_f$, that uses a finer discretization than $\hat{v}$. This approach is reasonable due to

$$||\hat{v} - \hat{v}_f||_{gl} \leq ||\hat{v} - v||_{gl} + ||v - \hat{v}_f||_{gl}$$

The latter term of the RHS is smaller such that $||\hat{v} - v||_{gl}$ is the main contributor to the error. The two cases are denoted by "intra=False" and by "intra=True" in the practical convergence analysis. The notation refers to the comparison between approximations. Note, that the convergence in $||\cdot||_{gl}$ is bounded by the pointwise convergence in $|\cdot|$ with Δ-inquality

$$\begin{aligned}
||B||_{gl} =& \sqrt{\frac{\sum_{n,i=0}^{N,I} B_{n,i}^2}{(N+1)(I+1)}} \\
\leq& \sqrt{\frac{\left(\sum_{n,i=0}^{N,I} |B_{n,i}|\right)^2}{(N+1)(I+1)}} \\
\leq& \frac{(N+1)(I+1)(\mathcal{O}(\Delta t) + \mathcal{O}(h) + \mathcal{O}(1/\sqrt{M}))}{(N+1)(I+1)} \\
=& \mathcal{O}(\Delta t) + \mathcal{O}(h) + \mathcal{O}(1/\sqrt{M})
\end{aligned}$$

The convergence plots in Section 3.5.1 and Section 3.5.2 always consider a setting where the true value function is known and a setting where the value function corresponds to (3.3) in Section 3.2 with $\gamma(\delta, x) = (x - \delta)^2$. The known value function is $v_0(s, x) = (T - s)\cos(x)|_{T=2}$ and called zero-function, referring to its value in $s = 2$.

3.5.1 Convergence Finite Difference Method

Claim 3.3 yielded the theoretical convergence rates for a value function with higher regularity and known boundary values. The practical convergence rates of the FDM are only examined with the upwind scheme to avoid bias coming from oscillations. To compute the zero-function with FDM the cost function needs to be substituted with an adequate expression, also called forcing term. Because v_0 is smooth, its derivatives are well-defined. The forcing term is therefore given by

$$\begin{aligned}
f(s, x) &= -\frac{\partial}{\partial s} v_0(s, x) - \mu(\Lambda_s - x)\frac{\partial}{\partial x} v_0(s, x) - \frac{\sigma^2}{2}\frac{\partial^2}{\partial x^2} v_0(s, x) \\
&= \cos(x) + \mu(\Lambda_s - x)(T - s)\sin(x) + \frac{\sigma^2}{2}(T - s)\cos(x)
\end{aligned}$$

and $v_0(T, x) = 0,\ \forall x \in [\tilde{a}, \tilde{b}]$. Moreover, σ is reduced to 1 and Figure 3.6 shows the convergence of Algorithm 1 in space. Note, that the y-axis measuring the error does not coincide for different plots.

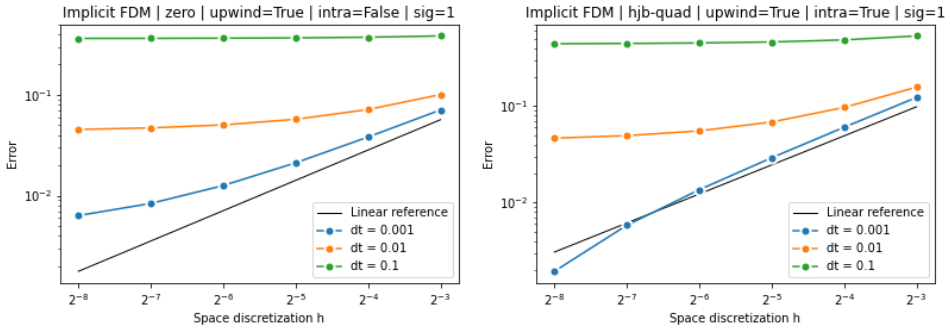

Figure 3.6: Error convergence of spatial step h with known verification function v_0 (left) and unknown value function with squared cost term $f_\gamma(s, x, \delta) = -p_s\rho(\delta) - c_s(x - \delta)^2$ (right). The upwind scheme is turned on and the diffusion constant is $\sigma = 1$. The best-case solution in the intra=True case is computed with $\Delta t = 0.001$ and $h = 2^{-9}$

The error progression in both plots of Figure 3.6 is measured for different time discretization sizes Δt, i.e. $0.1, 0.01$ and 0.001. The error doesn't seem to converge for the coarsest time discretization $\Delta t = 0.1$, as it stays on the same error level. For $\Delta t = 0.01$ the error only diminishes for the first couple of discretization sizes h and stagnates afterwards. This behaviour shows that the error induced by the time discretization is still too high and prevents further error reduction with finer h. For $\Delta t = 0.001$, the error keeps getting smaller at a linear rate for most discretization sizes. In the left plot, it starts stagnating at the end, whereas it turns slightly over-linear in the right one. The latter observation can be reasoned by the fact that the simulation converges to its best-case approximation given by $h = 2^{-9}$. The stagnation effect can also be observed when the Algorithm is examined regarding its convergence in time. Figure 3.7 fixes different discretization sizes h and refines time gradually.

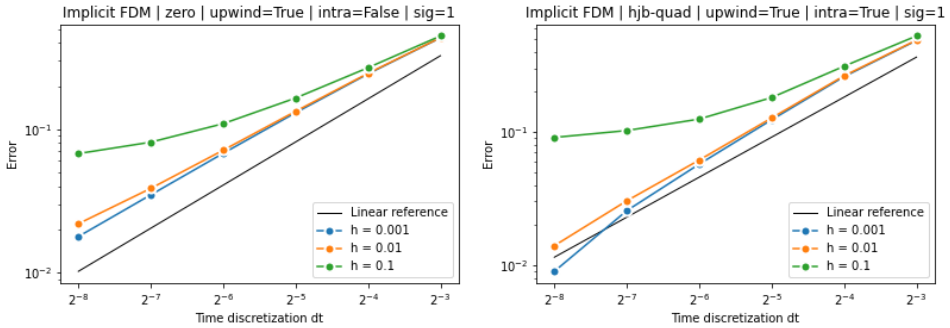

Figure 3.7: Error convergence of time step Δt of the verification function v_0 (left) and of the value function with squared cost term $f_\gamma(s, x, \delta) = -p_s\rho(\delta) - c_s(x-\delta)^2$ (right). The upwind scheme is turned on and the diffusion constant is $\sigma = 1$. The best-case solution in the intra=True case is computed with $h = 0.001$ and $\Delta t = 2^{-9}$

The error in Figure 3.7 only seems to start stagnating for $h = 0.1$. Most of the graphs stick to the linear reference slope, which was the convergence rate suggested by Claim 3.3. Moreover, when Δt is coarse, the different fixed discretization sizes h don't seem to influence the error as much as in Figure 3.6. The slightly over-linear convergence on the right plot has the same explanation given previously. Consequently, Claim 3.3 seems to have predicted the convergence order accurately, even though the proof assumed higher regularity on the value function and known boundary data.

3.5.2 Convergence Monte Carlo

Due to the randomness in Algorithm 2, it seems reasonable to run every error approximation multiple times to quantify the probability for each of them. This information could be used to determine a variance estimate for the error or a confidence interval, which quantifies the probability of the error. However, only one error approximation is performed and justified by the global nature of $||\cdot||_{gl}$ across the whole DoI. Note, that discretizing space is not a source of error in Algorithm 2. Therefore, it is kept at $h = 0.1$ throughout all subsequent simulations in this subsection. Figure 3.8 shows the convergence behaviour for a fixed time and space discretization with varying numbers of MC simulations.

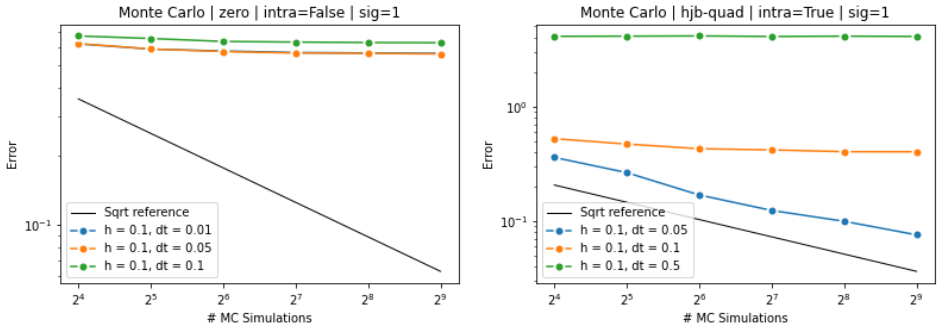

Figure 3.8: Error convergence with respect to the number of MC simulations of the verification function v_0 (left) and the value function with squared cost term $f_\gamma(s, x, \delta) = -p_s\rho(\delta) - c_s(x - \delta)^2$ (right). The diffusion constant is $\sigma = 1$ and the best-case solution in the intra=True case is computed with $\Delta t = 0.05, h = 0.1$ and $M = 2^{10}$

Claim 3.4 dictates the rate of convergence $\mathcal{O}(\frac{1}{\sqrt{M}})$ and is matched on the right plot in Figure 3.8 with $h = 0.1$ and $\Delta t = 0.05$ for the first few M. It starts stagnating after a certain number of simulations. The coarser discretizations lead much faster to a motionless state of the error. The left plot raises some doubts as practically no convergence takes place. Refining the fixed discretization leads to the same results. It turns out, that the convergence behaviour improves for higher σ. In Figure 3.10, the same calculations are performed for $\sigma = 3$ instead of $\sigma = 1$. Also, higher σ leads to a more even-spaced convergence for the same parameters of the left plot in Figure 3.9. This observation is left as an open question mark. The righthand plot in Figure 3.9 seems to converge with a linear convergence rate. Recall, that the convergence rate p in time for Algorithm 2 was not given.

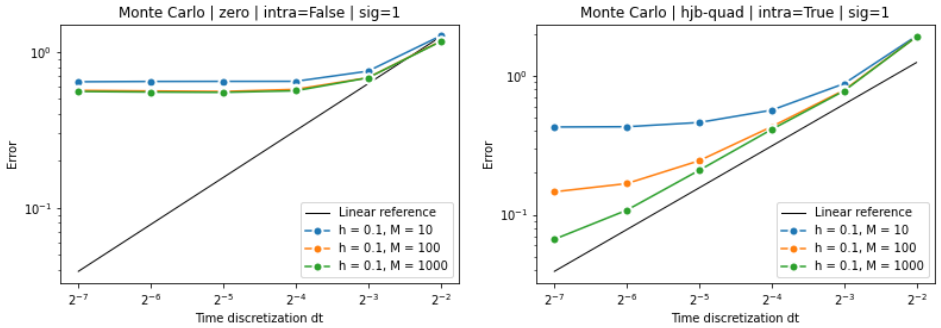

Figure 3.9: Error convergence in time of the verification function v_0 (left) and of the value function with squared cost term $f_\gamma(s,x,\delta) = -p_s\rho(\delta) - c_s(x-\delta)^2$ (right). The diffusion constant is $\sigma = 1$ and the best-case solution in the intra=True case is computed with $\Delta t = 2^{-8}, h = 0.1$ and $M = 1000$

The error on the right plot of Figure 3.9 starts stagnating for different fixed parameters, in this case the number of MC simulations. The linear convergence in time seems validated once more by the right plot in Figure 3.10.

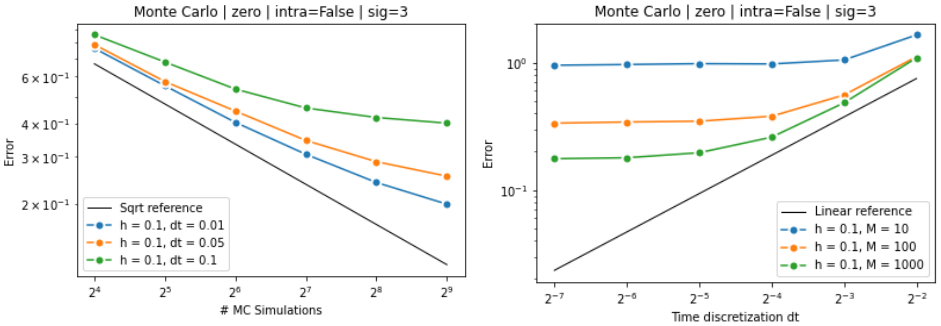

Figure 3.10: Error convergence of the left plots of Figure 3.8 and Figure 3.9 with diffusion constant $\sigma = 3$

4 An Intermediate Microgrid model

This chapter discusses an extended version of the model set up in Section 3.2. It starts by elaborating on its difference and presents an updated formal framework, invoking once again two numerical solutions. The first solution is a FDM, whereas the second one is based on RMC. The chapter focuses on the latter one motivating its feasibility and presents an algorithm that uses Importance Sampling to cope with the updated residual demand. The numerical results are discussed in Chapter 5.

Consider

$$dX_s^d = \left(\mu(\Lambda_s - X_s^d) - \lambda(X_s^d - d_s)_+\right) ds + \sigma dW_s$$

with $(\cdot)_+ := \max(\cdot, 0)$ and $\lambda > 0$. The superscript d shows the dependence of the residual demand on the control variable. $-\lambda(X_s^d - d_s)_+$ has to be interpreted as a penalization term that forces the consumer to reduce his consumption by a linear factor of λ once the amount of energy generated by the diesel generator is exceeded. As seen before, rough drift and diffusion terms are unfavourable for the existence of a smooth value function which is necessary to legitimate the HJB derivation. The Softplus or SmoothReLu function is a smooth approximation to $\max(\cdot, 0)$ and is occasionally used in Neural Networks [25]. It is defined as follows

$$z \mapsto \ln\left(1 + e^z\right), z \in \mathbb{R}$$

The previously introduced dynamics are modified in a way that preserves the penalization characteristic and allows to get back to the model presented in Chapter 3 for $\lambda = 0$. The new dynamics are thus given by

$$\begin{aligned} dX_s^d &= \left(\mu(\Lambda_s - X_s^d) + \ln(2) - \ln\left(1 + e^{\lambda(X_s^d - 1 - d_s)}\right)\right) dt + \sigma dW_s \\ &= \left(\mu(\Lambda_s - X_s^d) - \ln\left(\frac{1 + e^{\lambda(X_s^d - 1 - d_s)}}{2}\right)\right) ds + \sigma dW_s \end{aligned}$$

The -1 in $\lambda(X_s^d - 1 - d_s)$ is used for a closer approximation to the max function. Underlying processes that are influenced by the optimal control policy are called degenerate and it turns out that the optimal control policy now depends on the value function itself. Therefore, it can't be computed analytically any longer and alters the numerical approach. Throughout this chapter, only the quadratic version of γ is considered for the cost function (3.2).

4.1 Problem Setting

This section updates the problem setting of Chapter 3 and checks the assumptions justifying the existence of a strong solution to X_s. The following SDE formalizes the previously introduced underlying stochastic process.

$$\begin{cases} dX_s^d = \left(\mu(\Lambda_s - X_s^d) - \ln\left(\frac{1+e^{\lambda(X_s^d - 1 - d_s)}}{2}\right)\right) ds + \sigma dW_s \\ X_t = x \end{cases} \tag{4.1}$$

The drift and diffusion functions are defined as follows

$$b_\lambda : [0,T] \times \mathbb{R} \times D \to \mathbb{R}, (s,x,\delta) \mapsto \mu(\Lambda_s - x) - \ln\left(\frac{1 + e^{\lambda(x-1-\delta)}}{2}\right)$$
$$\sigma : [0,T] \times \mathbb{R} \times D \to \mathbb{R}_{>0}, (s,x,\delta) \mapsto \sigma$$

They are measurable due to continuity. Furthermore, b_λ and σ satisfy (2.8) and (2.9) for $\forall x, y \in \mathbb{R}, \delta \in D$ and $s \in [t,T]$:

$$\begin{aligned} & |b_\lambda(s,x,\delta) - b_\lambda(s,y,\delta)| + |\sigma(s,x,\delta) - \sigma(s,y,\delta)| \\ = \; & |\mu(\Lambda_s - x) + \ln(2) - \ln\left(1 + e^{\lambda(x-1-\delta)}\right) - \mu(\Lambda_s - y) - \ln(2) + \ln\left(1 + e^{\lambda(y-1-\delta)}\right)| \\ \leq \; & |\mu(x-y)| + |\ln\left(1 + e^{\lambda(x-1-\delta)}\right) - \ln\left(1 + e^{\lambda(y-1-\delta)}\right)| \\ \leq \; & (\mu+1)|x-y| \end{aligned}$$

due to $\frac{\partial}{\partial z}\ln(1+e^z) = e^z/(1+e^z) \in (0,1), \forall z \in \mathbb{R}$, where the Lipschitz constant is 1 by the Mean-value theorem. Moreover,

$$\begin{aligned} & |b_\lambda(s,x,\delta)| + |\sigma(s,x,\delta)| \\ = \; & |\mu(\Lambda_s - x) + \ln(2) - \ln\left(1 + e^{\lambda(x-1-\delta)}\right)| + \sigma \\ \leq \; & \mu|\Lambda_s| + \mu|x| + \ln(2) + 1 + \lambda|x| + \lambda(\delta+1) + \sigma \\ = \; & \underbrace{\mu|\Lambda_s| + \ln(2) + 1 + \lambda(\delta+1) + \sigma}_{\kappa_{s,\delta} :=} + \underbrace{(\mu+\lambda)}_{K:=}|x| \end{aligned}$$

$\forall x \in \mathbb{R}, \delta \in D$ and $s \in [t,T]$. The inequality was made possible by $\ln(1+e^z) \leq 1+|z|$ for $\forall z \in \mathbb{R}$. Clearly, $\kappa_{s,\delta}$ is in $L^2([t,T])$ for $\forall \delta \in D$. Consequently, (4.1) has a strong solution that is unique according to (2.7). The value function is defined as

$$v(t,x) = \sup_{d \in \mathcal{D}(t,x)} \mathbb{E}\left[\int_t^T -p_s\rho(d_s) - c_s(X_s^d - d_s)^2 ds | X_t^d = x\right] \tag{4.2}$$

When approximating $\int_t^T$ via LRAM, (4.2) turns into

$$v_i^n = \sup_{d \in \mathcal{D}(t_n,x_i)} \mathbb{E}\left[\sum_{k=n}^{N-1}(-p_{t_k}\rho(d_{t_k}) - c_{t_k}(X_{t_k}^d - d_{t_k})^2)\Delta t | X_{t_n}^d = x_i\right] \tag{4.3}$$

$\{X_{t_k}^d | X_{t_n} = x_i\}_{t_n \leq t_k \leq T}$ is the numerical solution to (4.1) given by EM

$$X_{t_{k+1}}^d = X_{t_k}^d + \left(\mu(\Lambda_{t_k} - X_{t_k}^d) - \ln\left(\frac{1+e^{\lambda(X_{t_k}^d - 1 - d_{t_k})}}{2}\right)\right)\Delta t + \sigma\sqrt{\Delta t}\, W_{t_k}, \tag{4.4}$$

where $W_{t_k} \overset{iid}{\sim} \mathcal{N}(0,1),\ \forall k = n, \ldots, N$.

4.2 Finite Difference Method

The previous framework lead again to a non-linear, second-order PDE. This section briefly discusses the different components of the PDE and presents a low-level Pseudocode for a numerical solution. Consider

$$\begin{cases} -\frac{\partial}{\partial s}v(s,x) - \mu(\Lambda_s - x)\frac{\partial}{\partial x}v(s,x) - \frac{\sigma^2}{2}\frac{\partial^2}{\partial x^2}v(s,x) \\ -\sup_{\delta \in [0,d_{max}]}\left(-\ln\left(\frac{1+e^{\lambda(x-1-\delta)}}{2}\right)\frac{\partial}{\partial x}v(s,x) - p_s\rho(\delta) - c_s(x-\delta)^2\right) = 0 \\ v(T,x) = 0 \end{cases} \tag{4.5}$$

Assuming that $\partial[a,b]$ is a manifold of class $C^{(3)}$, (4.5) has a unique solution $v \in C^{1,2}([0,T] \times [a,b])$ by the same argument as in Section 3.2. This results enables to identify (4.5) as the HJB to (4.2). As stated earlier, the value function is now part of the optimal control policy and the policy optimization is given by

$$d^*(s,x) = \arg\sup_{\delta \in [0,d_{max}]}\left(-\ln\left(\frac{1+e^{\lambda(x-1-\delta)}}{2}\right)\frac{\partial}{\partial x}v(s,x) - p_s\rho(\delta) - c_s(x-\delta)^2\right) \tag{4.6}$$

The complexity of the optimizing argument prevents the computation of an explicit form for $d^*(s,x)$ and a transition to a numerical optimization solves the issue. Grid optimization is applied but first, $\frac{\partial}{\partial x}v(s,x)$ needs to be examined further. An approximation to this term is made possible by switching from an implicit FDM to an explicit FDM [1]. Once, a valid approximation to $\frac{\partial}{\partial x}v$ exists, it is possible to compute v backwards in time. The stability criteria and higher computation time elaborated earlier have to be taken into account. A sketch for the implementation is laid out in the following.

Step 1 - Numerical Approximation of $\frac{\partial}{\partial x}v$ Different possibilities to approximate the derivative of a function were shown earlier and their concept is used again. Consider $\frac{\partial}{\partial x}v(t_n,x_i) \approx \frac{\partial}{\partial x}v_i^n = \frac{v_{i+1}^n - v_{i-1}^n}{2h}$ and $\frac{\partial}{\partial x}v(t_n,\tilde{a}) \approx \frac{\partial}{\partial x}v_0^n = \frac{v_1^n - v_0^n}{h}$, $\frac{\partial}{\partial x}v(t_n,\tilde{b}) \approx \frac{\partial}{\partial x}v_{\tilde{I}}^n = \frac{v_{\tilde{I}}^n - v_{\tilde{I}-1}^n}{h}$ at the boundary. The matrix form is given by

$$\begin{pmatrix} \frac{\partial}{\partial x}v(t_n,\tilde{a}) \\ \vdots \\ \frac{\partial}{\partial x}v(t_n,x_i) \\ \vdots \\ \frac{\partial}{\partial x}v(t_n,\tilde{b}) \end{pmatrix} \approx \frac{1}{h} \begin{pmatrix} -1 & 1 & & & \\ -1/2 & 0 & 1/2 & & \\ & & \ddots & & \\ & & -1/2 & 0 & 1/2 \\ & & & -1 & 1 \end{pmatrix} \begin{pmatrix} v_0^n \\ \vdots \\ v_i^n \\ \vdots \\ v_{\tilde{I}}^n \end{pmatrix}$$

The approximation for $\frac{\partial}{\partial x}v_i^n$, only works for a Backward Euler method (BE).

Step 2 - Compute Optimal policy With approximations from Step 1 compute the optimal control in $(t_k,\cdot)$ for $k = n,\dots,T$ via an optimization method on (4.6). This provides $d^*(t_n,x_i), \forall i = 0,\dots,\tilde{I}$.

Step 3 - Compute layers with forcing term The cost function or forcing term can be computed with the information obtained in Step 2. Repeating Step 1 to Step 3 backwards in time computes an approximation of the value function to a broader domain $[\tilde{a},\dots,\tilde{b}]$ as explained in Section 3.5.1. The approximation on the actual DoI is extracted afterwards.

Sparse arrays once again enhance the computational efficiency throughout the implementation. This numerical solution provides once again an overload of information. However, it remains computationally efficient as the state dimension is 1. The implemented method serves as a comparison possibility for the RMC method introduced in the next section.

4.3 Regression Monte Carlo

This section forms the thesis's core and uses the DPP presented earlier introducing the concept of continuation values. The quintessential part of the RMC is to accurately approximate them by mirroring the grid-discretization approach proposed in Alasseur et al. [4] via Importance Sampling.

Theorem 2.1 with constant stopping time $\theta = t_{n+1}$ and (2.16) state

$$\begin{aligned} v(t_n, x_i) &= \sup_{d \in \mathcal{D}(t_n, x_i)} \mathbb{E}\left[\int_{t_n}^{T} -p_s\rho(d_s) - c_s(X_s^d - d_s)^2 ds | X_{t_n}^d = x_i\right] \\ &\overset{2.1}{=} \sup_{d \in \mathcal{D}(t_n, x_i)} \mathbb{E}\left[\int_{t_n}^{t_{n+1}} -p_s\rho(d_s) - c_s(X_s^d - d_s)^2 ds + v(t_{n+1}, X_{t_{n+1}}^d) | X_{t_n}^d = x_i\right] \end{aligned}$$

Now, $\sup_{d \in \mathcal{D}(t_n, x_i)}$ in front of the expectation depends on the control stratgy $\sup_{d \in \mathcal{D}(t_{n+1}, X_{t_{n+1}}^d)}$, within $v(t_{n+1}, X_{t_{n+1}}^d)$. Note also, that the control strategy within $v(t_{n+1}, X_{t_{n+1}}^d)$ implicitly contains the control strategy in front of the expectation, except for the time-continuous controls exerted during $[t_n, t_{n+1})$. As the problem is discretized in time, it induces a strategy that only acts at times $t_n, t_{n+1}, \ldots, t_{N-1}$. Because $X_{t_n}^d$ is given by x_i, the cost function is deterministic at time t_n. The optimization problem is therefore decomposed into the optimal control δ at time t_n and all subsequent optimal controls at time $t_{n+1}, \ldots, T$.

$$\begin{aligned} v(t_n, x_i) &\approx \sup_{d \in \mathcal{D}(t_n, x_i)} \mathbb{E}\left[(-p_{t_n}\rho(d_{t_n}) - c_{t_n}(X_{t_n}^d - d_{t_n})^2)\Delta t + v(t_{n+1}, X_{t_{n+1}}^d) | X_{t_n}^d = x_i\right] \\ &\approx \max_{\delta \in [0, d_{max}]} (-p_{t_n}\rho(\delta) - c_{t_n}(x_i - \delta)^2)\Delta t + \mathcal{C}(t_n, x_i; \delta), \end{aligned}$$

where the last term is defined as the continuation value

$$\mathcal{C}(t_n, x_i; \delta) := \mathbb{E}[v(t_{n+1}, X_{t_{n+1}}^{\delta}) | X_{t_n}^d = x_i]$$

Applying, the same idea to $v(t_{n+1}, X_{t_{n+1}}^{\delta})$ iteratively and making use of the terminal condition $v(t_N, \cdot) = 0$, suggests a numerical backwards in time solution. The notational change from $X_{t_{n+1}}^d$ to $X_{t_{n+1}}^{\delta}$ originates from the fact, that $X_{t_{n+1}}^d$ is only influenced by one parameter $\delta := d_{t_n} \in [0, d_{max}]$. The continuation value is the expected value of all future costs in $[t_{n+1}, \ldots, T]$ once a decision was made in time t_n. Note, that $\mathcal{C}(t_{N-1}, \cdot; \delta) = 0$ for the terminal value $v(T, \cdot) = 0$. The optimization problem of (4.3) therefore consists of a one-step maximization problem and an approximation problem of $\mathcal{C}(\cdot, \cdot; \delta)$ for δ in D. It turns out, that

> if the conditional expectation is an element of the [Hilbert space] L^2 [...], it has a countable orthonormal basis and the conditional expectation can be represented as a linear function of the elements of the basis. – Longstaff et al. [5]

Chapter 2 in Longstaff et al. [5] initially introduces this approach for the accurate pricing of a financial contract, where the idea is to sell it in time t_k, if its current value exceeds its continuation value. $\mathcal{C}(\cdot, \cdot; \delta)$ is in L^2 as shown in Lemma 4.1 and with the previous quote it can be written as

$$\mathcal{C}(s, x; \delta) = \sum_{q=1}^{\infty} \alpha_{q,\delta}^{s} \tilde{\phi}_q(x),$$

where $\tilde{\phi}_q(\cdot),\ q \in \mathbb{N}$ are the basis functions. Projecting $\mathcal{C}(s, x; \delta)$ into a basis spanned by only Q instead of an infinite number of basis functions yields

$$\mathcal{C}_Q(s, x; \delta) := \sum_{q=1}^{Q} \alpha_{q,\delta}^{s} \phi_q(x) = \alpha_{\delta}^{s} \cdot \phi(x)$$

Note, that the number Q and the type of basis functions ϕ_q are essential for an accurate approximation to $\mathcal{C}(s, x; \delta)$. It is considered later in more detail. The question of how to compute α_{δ}^{s} remains for now. The theory above motivates the following claim.

Claim 4.1. *The second moment of the value function v and the conditional value are bounded and* $\mathbb{E}\left[\mathcal{C}(t_k, X_{t_k}^d; \delta)^2\right] \leq \mathbb{E}\left[v(t_{k+1}, X_{t_{k+1}}^d)^2\right] < \infty$.

Proof. Theorem 1.3.15 in Pham [9] states, that all moments of X_t^d are finite. Moreover, Theorem 3.5.2 in Pham [9] proclaims a quadratic growth condition on v. Hence, there exists a constant $C > 0$ such that

$$\mathbb{E}\left[v(t_{k+1}, X_{t_{k+1}}^d)^2\right] \leq \mathbb{E}\left[C(1 + |X_{t_{k+1}}^d|^2)^2\right] < \infty$$

The inequality for the conditional value is given by

$$\begin{aligned} \mathbb{E}\left[\mathcal{C}(t_k, X_{t_k}^d; \delta)^2\right] &= \mathbb{E}\left[\mathbb{E}\left[v(t_{k+1}, X_{t_{k+1}}^{\delta})|X_{t_k}^d\right]^2\right] \\ \leq \mathbb{E}\left[\mathbb{E}\left[v(t_{k+1}, X_{t_{k+1}}^{\delta})^2|X_{t_k}^d\right]\right] &= \mathbb{E}\left[v(t_{k+1}, X_{t_{k+1}}^d)^2\right] < \infty \end{aligned}$$

and stems from Jensen's Inequality for Conditional Expectations [16]. □

Theorem 4.1.15 in Durrett [26] - also known as the Hilbert Projection Theorem - is now applicable and states

$$\mathbb{E}\left[\left(v(t_{k+1}, X^{\delta}_{t_{k+1}}) - \mathcal{C}(t_k, X^d_{t_k}; \delta)\right)^2\right] = \min_{\varrho(X_{t_k}) \in L^2(\mathcal{F}_{t_k})} \mathbb{E}\left[\left(v(t_{k+1}, X^{\delta}_{t_{k+1}}) - \varrho(X_{t_k})\right)^2\right], \tag{4.7}$$

where $L^2(\mathcal{F}_s) := \{\varrho : \mathbb{R} \to \mathbb{R} | \varrho(X_s) \in \mathcal{F}_s \text{ and } \mathbb{E}\left[\varrho(X_s)^2\right] < \infty\}$ is a closed subset of $L^2(\mathcal{F})$ for $s \in [t, T]$. Recall, that $\mathcal{F}_s$ were the σ-algebras generated by the Brownian motions up to time s. Replacing the true conditional expectation by a projection into a space only spanned by Q basis function suggests the approximation

$$\mathbb{E}\left[\left(v(t_{k+1}, X^{\delta}_{t_{k+1}}) - \mathcal{C}(t_k, X^d_{t_k}; \delta)\right)^2\right] \approx \mathbb{E}\left[\left(v(t_{k+1}, X^{\delta}_{t_{k+1}}) - \mathcal{C}_Q(t_k, X^d_{t_k}; \delta)\right)^2\right]$$

Computing $\alpha^{t_k}_{\delta}$ in the RHS via Linear Regression gives the method its name. Initially, the method required an approximation to the conditional value and not the value function. The Lemma 4.1 clarifies this ambiguity.

Lemma 4.1. *Let v be as in (4.2). Then*

$$\arg\min_{\alpha \in \mathbb{R}^Q} \mathbb{E}\left[\left(v(t_{k+1}, X^{\delta}_{t_{k+1}}) - \alpha \cdot \phi(X^d_{t_k})\right)^2\right] = \arg\min_{\alpha \in \mathbb{R}^Q} \mathbb{E}\left[\left(\mathcal{C}(t_k, X^d_{t_k}; \delta) - \alpha \cdot \phi(X^d_{t_k})\right)^2\right],$$

where $\alpha \cdot \phi(x) = \sum_{q=1}^{Q} \alpha_q \phi_q(x)$ with $\alpha \in \mathbb{R}^Q$.

Proof.

$$\begin{aligned}
&\arg\min_{\alpha \in \mathbb{R}^Q} \mathbb{E}\left[\left(V(t_{k+1}, X^{\delta}_{t_{k+1}}) - \alpha \cdot \phi(X^d_{t_k})\right)^2\right] \qquad (4.8)\\
&= \arg\min_{\alpha \in \mathbb{R}^Q} \mathbb{E}\left[\mathbb{E}\left[\left(V(t_{k+1}, X^{\delta}_{t_{k+1}}) - \alpha \cdot \phi(X^d_{t_k})\right)^2 | X^d_{t_k}\right]\right]\\
&= \arg\min_{\alpha \in \mathbb{R}^Q} \mathbb{E}\left[\mathbb{E}\left[\left(V(t_{k+1}, X^{\delta}_{t_{k+1}}) - \mathcal{C}(t_k, X^d_{t_k}; \delta) + \mathcal{C}(t_k, X^d_{t_k}; \delta) - \alpha \cdot \phi(X^d_{t_k})\right)^2 | X^d_{t_k}\right]\right]\\
&= \arg\min_{\alpha \in \mathbb{R}^Q} \mathbb{E}\left[\mathbb{E}\left[\left(V(t_{k+1}, X^{\delta}_{t_{k+1}}) - \mathcal{C}(t_k, X^d_{t_k}; \delta)\right)^2 + \left(\mathcal{C}(t_k, X^d_{t_k}; \delta) - \alpha \cdot \phi(X^d_{t_k})\right)^2 | X^d_{t_k}\right]\right]\\
&\quad + 2\mathbb{E}\left[\mathbb{E}\left[\left(V(t_{k+1}, X^{\delta}_{t_{k+1}}) - \mathcal{C}(t_k, X^d_{t_k}; \delta)\right)\left(\mathcal{C}(t_k, X^d_{t_k}; \delta) - \alpha \cdot \phi(X^d_{t_k})\right) | X^d_{t_k}\right]\right] =: *
\end{aligned}$$

The first term is independent of α and doesn't influence the minimal argument arg min. For the cross-product note, that

$$\begin{aligned}
&2\mathbb{E}\left[\mathbb{E}\left[\left(V(t_{k+1}, X^{\delta}_{t_{k+1}}) - \mathcal{C}(t_k, X^d_{t_k}; \delta)\right)\left(\mathcal{C}(t_k, X^d_{t_k}; \delta) - \alpha \cdot \phi(X^d_{t_k})\right) | X^d_{t_k}\right]\right] \\
=&2\mathbb{E}\left[\mathbb{E}\left[\left(V(t_{k+1}, X^{\delta}_{t_{k+1}}) - \mathcal{C}(t_k, X^d_{t_k}; \delta)\right) | X^d_{t_k}\right]\left(\mathcal{C}(t_k, X^d_{t_k}; \delta) - \alpha \cdot \phi(X^d_{t_k})\right)\right] \\
=&2\mathbb{E}\left[\left(\mathbb{E}\left[V(t_{k+1}, X^{\delta}_{t_{k+1}}) | X^d_{t_k}\right] - \mathcal{C}(t_k, X^d_{t_k}; \delta)\right)\left(\mathcal{C}(t_k, X^d_{t_k}; \delta) - \alpha \cdot \phi(X^d_{t_k})\right)\right] = 0
\end{aligned}$$

Both observations lead to

$$\begin{aligned}
* &= \underset{\alpha \in \mathbb{R}^Q}{\arg\min}\, \mathbb{E}\left[\mathbb{E}\left[\left(\mathcal{C}(t_k, X^d_{t_k}; \delta) - \alpha \cdot \phi(X^d_{t_k})\right)^2 | X^d_{t_k}\right]\right] \\
&= \underset{\alpha \in \mathbb{R}^Q}{\arg\min}\, \mathbb{E}\left[\left(\mathcal{C}(t_k, X^d_{t_k}; \delta) - \alpha \cdot \phi(X^d_{t_k})\right)^2\right]
\end{aligned}$$

□

In order to compute α numerically, a MC simulation with M trajectories can be performed [5], i.e.

$$\hat{\alpha}^{t_k}_{\delta} = \underset{\alpha \in \mathbb{R}^Q}{\arg\min}\, \frac{1}{M} \sum_{m=1}^{M} \left(v(t_{k+1}, X^{\delta,(m)}_{t_{k+1}}) - \alpha \cdot \phi(X^{d,(m)}_{t_k})\right)^2$$

The fitted values of the Regression are of the form

$$\hat{\mathcal{C}}_Q(t_k, x; \delta) = \sum_{q=1}^{Q} \hat{\alpha}^{t_k}_{\delta} \phi_q(x)$$

For notational clarity, Q in $\hat{\mathcal{C}}_Q$ is dropped throughout the rest of the paper. By now, it remains unclear how to sample $\{X^d_{t_k} | X^d_{t_n} = x_i\}_{t_n \le t_k \le T}$ for the MC simulation. They still depend on the unknown control $d \in \mathcal{D}(t_n, x)$ and can't be sampled as done in Chapter 3. Providing a naïve control strategy $\hat{d}$, however, circumvents this difficulty at the expense of many inaccurate underlying processes $\{X^{\hat{d}}_{t_k} | X^{\hat{d}}_{t_n} = x_i\}_{t_n \le t_k \le T}$. By the stochastic dynamic of $\{X^d_s | X^d_s = x_i\}_{t \le s \le T}$ in (4.1), every trajectory point of $\{X^{\hat{d}}_{t_k} | X^{\hat{d}}_{t_n} = x_i\}_{t_n \le t_k \le T}$ is reachable. Therefore, regressing according to (4.8) enables to find the optimal control policy along every naïvely sampled trajectory point. Performing a RMC on these naïve trajectories leads to an approximation of the value function and control policy that is only accurate within the domain spanned by the naïve trajectories but inaccurate outside this area.
The idea is to resample these inaccurate trajectories with a given control policy, that is locally accurate and improve the trajectories iteratively such

that they resemble more and more $\{X_{t_k}^{d^*} | X_{t_n}^{d^*} = x_i\}_{t_n \leq t_k \leq T}$, where d^* is the optimal control policy in $\mathcal{D}(t_n, x)$. The "level" of iteration is called l in the following and Figure 4.1 illustrates the previous discussion.

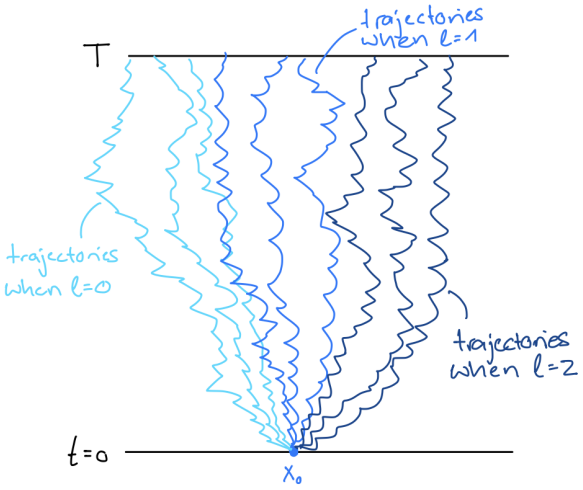

Figure 4.1: Resampling strategy for trajectories with different levels of accuracy $l = 0, 1, 2$ for the control policy

When $l = 0$, initial trajectories are sampled and the control policy is only accurate within their domain. Those naïve trajectories are represented by the light-blue ones in Figure 4.1. With increasing l, the domain for the accurate control policy extends to the domain spanned by the blue and dark-blue trajectories. The procedure could be initiated by a naïve control strategy $d_l \equiv 0$ for $l = 0$. The discussion suggests to introduce l into (4.8), which leads to

$$\alpha_{\delta,l}^{t_k} = \arg\min_{\alpha \in \mathbb{R}^Q} \mathbb{E}\left[\left(v(t_{k+1}, X_{t_{k+1}}^{\delta}) - \alpha \cdot \phi(X_{t_k}^{d_l})\right)^2\right]$$

In this form, $\alpha_{\delta,l}^{t_k}$ could be approximated via a MC simulation for different δ. However, a change of measure idea suggested by Prof. Nobile gives the possibility to reduce the amount of MC simulation:

Denote by $\pi(\cdot, \cdot)$ the joint probability density function (pdf) of $X_{t_k}^{d_l}$ and $X_{t_{k+1}}^{\delta}$. Its existence is assumed for now. Then

$$
\begin{aligned}
\alpha_{\delta,l}^{t_k} &= \arg\min_{\alpha\in\mathbb{R}^Q} \mathbb{E}\left[\left(v(t_{k+1}, X_{t_{k+1}}^{\delta}) - \alpha\cdot\phi(X_{t_k}^{d_l})\right)^2\right] \\
&= \arg\min_{\alpha\in\mathbb{R}^Q} \int_{\mathbb{R}\times\mathbb{R}} \left(v(t_{k+1}, y) - \alpha\cdot\phi(x)\right)^2 \pi_{X_{t_k}^{d_l}, X_{t_{k+1}}^{\delta}}(x,y)d(x,y) \\
&= \arg\min_{\alpha\in\mathbb{R}^Q} \int_{\mathbb{R}}\int_{\mathbb{R}} \left(v(t_{k+1}, y) - \alpha\cdot\phi(x)\right)^2 \pi_{X_{t_{k+1}}^{\delta}|X_{t_k}^{d_l}}(y|x)\pi_{X_{t_k}^{d_l}}(x)dydx \\
&= \arg\min_{\alpha\in\mathbb{R}^Q} \int_{\mathbb{R}}\int_{\mathbb{R}} \left(v(t_{k+1}, y) - \alpha\cdot\phi(x)\right)^2 \frac{\pi_{X_{t_{k+1}}^{\delta}|X_{t_k}^{d_l}}(y|x)\pi_{X_{t_k}^{d_l}}(x)}{\pi_{X_{t_{k+1}}^{d_l}|X_{t_k}^{d_l}}(y|x)\pi_{X_{t_k}^{d_l}}(x)} \pi_{X_{t_k}^{d_l}, X_{t_{k+1}}^{d_l}}(y,x)dydx \\
&= \arg\min_{\alpha\in\mathbb{R}^Q} \int_{\mathbb{R}}\int_{\mathbb{R}} \left(v(t_{k+1}, y) - \alpha\cdot\phi(x)\right)^2 \frac{\pi_{X_{t_{k+1}}^{\delta}|X_{t_k}^{d_l}}(y|x)}{\pi_{X_{t_{k+1}}^{d_l}|X_{t_k}^{d_l}}(y|x)} \pi_{X_{t_{k+1}}^{d_l}, X_{t_k}^{d_l}}(x,y)dydx \\
&= \arg\min_{\alpha\in\mathbb{R}^Q} \mathbb{E}\left[\left(v(t_{k+1}, X_{t_{k+1}}^{d_l}) - \alpha\cdot\phi(X_{t_k}^{d_l})\right)^2 \frac{\pi_{X_{t_{k+1}}^{\delta}|X_{t_k}^{d_l}}(X_{t_{k+1}}^{d_l}|X_{t_k}^{d_l})}{\pi_{X_{t_{k+1}}^{d_l}|X_{t_k}^{d_l}}(X_{t_{k+1}}^{d_l}|X_{t_k}^{d_l})}\right]
\end{aligned}
$$

Bayes Rules as in Tsitsiklis [27] and Fubini's theorem entered the third identity. The numerical approximation of the regression parameters is given by

$$
\hat{\alpha}_{\delta,l}^{t_k} = \arg\min_{\alpha\in\mathbb{R}^Q} \frac{1}{M}\sum_{m=1}^{M} \left(v(t_{k+1}, X_{t_{k+1}}^{d_l,(m)}) - \alpha\cdot\phi(X_{t_k}^{d_l,(m)})\right)^2 \frac{\pi_{X_{t_{k+1}}^{\delta}|X_{t_k}^{d_l}}(X_{t_{k+1}}^{d_l,(m)}|X_{t_k}^{d_l,(m)})}{\pi_{X_{t_{k+1}}^{d_l,(m)}|X_{t_k}^{d_l,(m)}}(X_{t_{k+1}}^{d_l,(m)}|X_{t_k}^{d_l,(m)})}
$$

and the amount of MC simulation is reduced to one per time step if the regression parameters can be made explicitly dependent on δ. Gaussian increments are used in the EM scheme of (4.4). Consequently, $\pi_{X_{t_{k+1}}^{\delta}|X_{t_k}^{d_l}}(\cdot|x)$ is the pdf of

$$
X_{t_{k+1}}^{\delta}|X_{t_k}^{d_l} = x \sim \mathcal{N}\left(x + \left(\mu(\Lambda_{t_k} - x) - \ln\left(\frac{1+e^{\lambda(x-1-\delta)}}{2}\right)\right)\Delta t, \sigma^2\Delta t\right)
$$

The theoretical convergence of the previous heuristic derivation is not inspected. Instead, Algorithm 3 is presented incorporating the details from the previous discussion within a computationally friendly frame.

Algorithm 3: Regression Monte Carlo

input : Initial values $x_1, \dots, x_p$ and $d_l|_{l=0} \equiv 0$
Sample M times $\{X_{t_k}^{d_l}|X_0^{d_l} = x_1\}_{0\le t_k\le T}, \dots, \{X_{t_k}^{d_l}|X_0^{d_l} = x_p\}_{0\le t_k\le T}$
while *criteria* $l < L$ **do**
  for $t_k \in [t_{N-1}, \dots, 0]$ **do**
    Initiate $v(t_N, X_{t_N}^{d_l,(m)}) = 0$ for all previously sampled trajectories
    Compute regression parameters

$$\hat{\alpha}_{\delta,l}^{t_k} = \underset{\alpha\in\mathbb{R}^Q}{\arg\min} \sum_{j=1}^{p} \left(\frac{1}{M} \sum_{m=1}^{M} \left[v_\delta(t_{k+1}, X_{t_{k+1}}^{d_l,(m)}) - \alpha \cdot \phi_\delta(X_{t_k}^{d_l,(m)}) \right]^2 |X_0^{d_l,(m)} = x_j \right)$$

$$\text{with } v_\delta(t_{k+1}, X_{t_{k+1}}^{d_l,(m)}) = v(t_{k+1}, X_{t_{k+1}}^{d_l,(m)}) \sqrt{\frac{\pi_{X_{t_{k+1}}^{\delta}|X_{t_k}^{d_l}}(X_{t_{k+1}}^{d_l,(m)}|X_{t_k}^{d_l,(m)})}{\pi_{X_{t_{k+1}}^{d_l}|X_{t_k}^{d_l}}(X_{t_{k+1}}^{d_l,(m)}|X_{t_k}^{d_l,(m)})}}$$

$$\text{and } \phi_\delta(X_{t_k}^{d_l,(m)}) = \phi(X_{t_k}^{d_l,(m)}) \sqrt{\frac{\pi_{X_{t_{k+1}}^{\delta}|X_{t_k}^{d_l}}(X_{t_{k+1}}^{d_l,(m)}|X_{t_k}^{d_l,(m)})}{\pi_{X_{t_{k+1}}^{d_l}|X_{t_k}^{d_l}}(X_{t_{k+1}}^{d_l,(m)}|X_{t_k}^{d_l,(m)})}}$$

$$\hat{\mathcal{C}}(t_k, X_{t_k}^{d_l,(m)}; \delta) = \hat{\alpha}_{\delta,l}^{t_k} \cdot \phi(X_{t_k}^{d_l,(m)})$$

    This provides data for the optimal control and value function

$$\delta^* = \underset{\delta\in[0,d_{max}]}{\arg\max} \left(-p_{t_k}\rho(\delta) - c_{t_k}(X_{t_k}^{d_l,(m)} - \delta)^2 \right) \Delta t + \hat{\mathcal{C}}(t_k, X_{t_k}^{d_l,(m)}; \delta)$$

$$v(t_k, X_{t_k}^{d_l,(m)}) = \left(-p_{t_k}\rho(\delta^*) - c_{t_k}(X_{t_k}^{d_l,(m)} - \delta^*)^2 \right) \Delta t + \hat{\mathcal{C}}(t_k, X_{t_k}^{d_l,(m)}; \delta^*)$$

  end
  if $l + 1 \neq L$ **then**
    Resample trajectories in the second line with improved control using regression parameters at every iteration

$$\hat{\mathcal{C}}(t_k, \cdot; \delta) = \hat{\alpha}_{\delta,l}^{t_k} \cdot \phi(\cdot)$$

$$d^*(t_k, \cdot) = \underset{\delta\in[0,d_{max}]}{\arg\max} \left(-p_{t_k}\rho(\delta) - c_{t_k}(\cdot - \delta)^2 \right) \Delta t + \hat{\mathcal{C}}(t_k, \cdot; \delta)$$

    Increase l by 1
  end
end
Specify DoI, i.e. $t_n = 0, \dots, T$ and $x_i = a, \dots, b$
Compute v on DoI for some optimal δ^* using $\alpha_{\delta,l}^{t_k}$, i.e.

$$v(t_n, x_i) = \left(-p_{t_n}\rho(\delta^*) - c_{t_n}(x_i - \delta^*)^2 \right) \Delta t + \hat{\mathcal{C}}(t_n, x_i; \delta^*)$$

The following example is intended to clarify the if-statement inAlgorithm 3 and constructs a solution to (4.4) for $N = 3$ when $l = 1$.

Example 4.1. *Consider the time steps $t_0 = 0, t_1 = 1, t_2 = 2 = T$ and the initial value $x \in \mathbb{R}$. Let $W_0, W_1 \overset{iid}{\sim} \mathcal{N}(0,1)$ and $l = 1$. d_0 is given by a simple control strategy, e.g. $d_0 = 0$. As the initial values don't change, d_0 can be used at every level l for the first iteration of the residual demand trajectories. W_0 leads a.s. to a previously untouched point $X_1 \in \mathbb{R}$.*

$$X_1 = x + \left(\mu(\Lambda_0 - x) - \ln\left(\frac{1 + e^{\lambda(x-1-d_0)}}{2}\right)\right)\Delta t + \sigma\sqrt{dt}W_0$$

The for-loop computed regression parameters at level $l = 0$ providing data to approximate the conditional expectation. With X_1 now given, d_1 is computed by

$$d_1 = \underset{\delta\in[0,d_{max}]}{\arg\max}\left(-p_1\rho(\delta) - c_1(X_1 - \delta)^2\right)\Delta t + \hat{\mathcal{C}}(t_1, X_1; \delta)$$

and is used to calculate

$$X_2 = X_1 + \left(\mu(\Lambda_1 - X_1) - \ln\left(\frac{1 + e^{\lambda(X_1-1-d_1)}}{2}\right)\right)\Delta t + \sigma\sqrt{dt}W_1$$

The comment after the while-loop emphasizes that the Algorithm allows the approximation of the value function on the whole domain, or in practice only in one point (t_n, x_i). For the latter setting, set $p = 1$ and choose the initial point to coincide with the current state of the residual demand. The value function is then evaluated by solving (4.3) with $\delta^* \in [0, \ldots, d_{max}]$. The according conditional value is approximated using the regression parameters from level $L - 1$. The level L isn't reached as the algorithm starts at $l = 0$. Consequently, $L = 3$ corresponds to trajectories that are optimized twice. Throughout the implementation in Chapter 5, the initial values are of the form $(x_1, 0), \ldots, (x_p, 0)$.

Note, that once the while-loop is executed another possibility of reconstructing the value function on the DoI exists. This alternative is motivated by Alasseur et al. [4] and starts by simulating another set of $\tilde{M}$ trajectories $\{X_{t_k}^{d_L} | X_{t_n}^{d_L} = x_i\}_{t_n \le t_k \le T}$ with initial value (t_n, x_i) in the DoI. The cost function is then evaluated and summed up along each trajectory. Their average is an approximated evaluation of the value function in (t_n, x_i). This approach is formalized as follows

Specify DoI, i.e. $t_n = 0, \dots, T$ and $x_i = a, \dots, b$. Initiate $\mathcal{J}_{t_n} = 0$
Sample $\tilde{M}$ times $\{X_{t_k}^{d_L} | X_{t_n}^{d_L} = x_i\}_{t_n \leq t_k \leq T}$
for $t_k \in [t_n, \dots, t_{N-1}]$ **do**
  $\mathcal{J}_{t_{k+1}}^{(m)} = \mathcal{J}_{t_k}^{(m)} - p_{t_k}\rho(d_{t_k}) - c_{t_k}(X_{t_k}^{d_L,(m)} - d_{t_k})^2$
end
output: $v(t_n, x_i) = \frac{1}{\tilde{M}} \sum_{m=1}^{\tilde{M}} \mathcal{J}_{t_N}^{(m)}$

The number of optimization problems increases $\tilde{M} \cdot (N - 1 - n)$-fold due to the resampling d_L and indicates the inefficiency of the approach. Moreover, numerical implementations reveal, that v seems to be more volatile along the DoI for low $\tilde{M}$ and high σ. The equivalence of the two evaluations of v was only tested empirically and no further theoretical explanations are given.

Throughout the algorithm, several methods exist to solve the control optimization problem and to compute δ^*, namely Gradient Descent and the Newton method. In this thesis, a grid optimization method is explored. In the regression part, $\hat{\alpha}$ can be specified in a way that explicitly depends on δ. However, a naïve approach is taken and $\hat{\alpha}_\delta$ is computed for each δ individually. The approach diminishes the computational efficiency and increases the storing complexity for the various δ along the implementation. For the basis functions, polynomials up to degree Q with intercept are considered [4], hence $Q+1$ basis functions in total. Chapter 5 in Devroye et al. [28] suggests including the cost function into the basis functions to enhance approximation accuracy. However, this isn't applicable for $Q > 2$ as f_γ is a quadratic function with respect to the state variable.

The overall complexity of Algorithm 3 approximating the value function along the whole DoI can be summarized as follows. For every level L, $p \cdot M$ trajectories are sampled from time 0 to final time T. Then the for-loop with length N performs $\#|[0, \dots, d_{max}]|$ linear regressions, where the independent and dependent variables are previously corrected by a change of measure. Further in the for-loop, $p \cdot M$ conditional expectations are predicted with the newly obtained regression parameters for all $\delta \in [0, \dots, d_{max}]$. In the last step of the for loop $p \cdot M$ optimization problems are solved to obtain an approximation to the value function along the trajectory points.

This chapter has laid out the details legimitating the implementation of a RMC method when the underlying residual demand process is degenerated. An algorithm has been presented and elaborated and its results are described in the proceeding chapter.

5 Numerical Resolution

Algorithm 3 is now implemented, and the results are presented in this chapter. However, despite the theoretical considerations discussed in Chapter 4, it has become apparent that the implemented Importance Sampling method results in an unexpected distortion of the value function. In order to validate the correctness of the Algorithm, different parts are examined and tested individually.

Algorithm 3 uses the same price processes and discretizations as in Section 3.4. The control space for grid optimization consists of an evenly spaced array of values ranging from 0 to 7, inclusive, with a total of 20 points. The residual demand is modified to

$$\begin{cases} dX_s^d = \left((6\sin(\pi s) - X_s^d) - \ln\left(\frac{1+e^{2(X_s^d - 1 - d_s)}}{2}\right)\right) ds + dW_s \\ X_t^d = x \end{cases}$$

for $t \leq s \leq 2$ and $t \in [0, 2]$. The value function is computed using two methods: first, an explicit FDM as discussed in Section 4.2, and second, the RMC approach presented in Section 4.3. The results are given in Figure 5.1

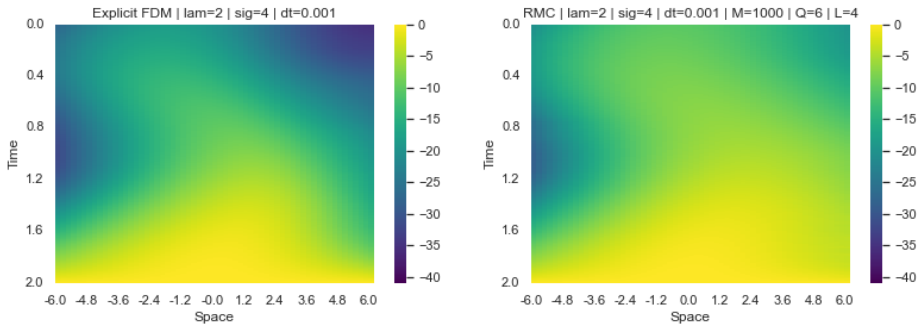

Figure 5.1: Heatmap of the optimal value function for explicit FDM (left) and RMC (right) with discretization size $\Delta t = 0.001$, $h = 0.2$ and $\sigma = 4$. Further parameters for RMC are given in the plot title, coinciding with the notations introduced in this work.

When comparing these results to those in Section 3.4, both plots exhibit a similar overall shape attributed to the mean-reverting behavior of the residual demand. Since the cost function remains unchanged, both approximations to the value function yield negative values. Notably, the penalization term $\lambda = 2$ causes the explicit FDM to exhibit locally more extreme values than those in Section 3.4. Conversely, the value function provided by RMC shows significant smoothing at the beginning of the time interval, resulting in a noticeable distinction between the two approximations. Their difference amounts to 6.4079 in the $||\cdot||_{gl}$ norm. Improving the amount of MC simulations for the different points, the time discretizations Δt and the number of basis function Q for the RMC doesn't reduce the overall divergence. Fitting the value function with the information provided along all trajectories via linear or cubic interpolation instead of evaluating as done at the end of Algorithm 3 returns equally distorted results. This distinction is even more apparent in their according optimal control policies given in Figure 5.2.

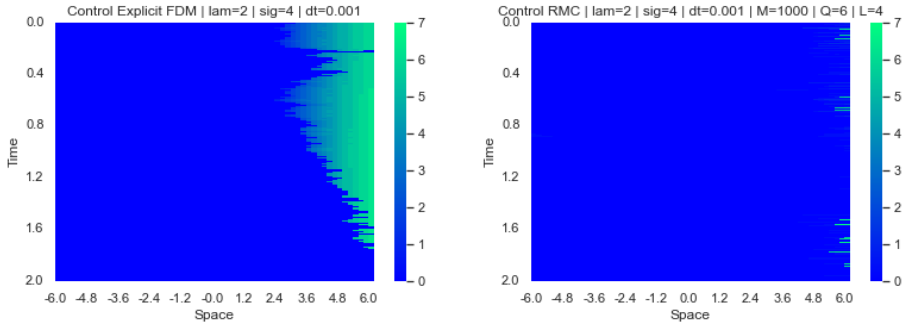

Figure 5.2: Heatmap of the optimal control strategy provided by the explicit FDM (left) and RMC (right) with discretization size $\Delta t = 0.001$, $h = 0.2$ and $\sigma = 4$

The control policy on the left activates the diesel generator only when the residual demand is positive, as discussed in Figure 3.4. This policy appears continuous in space but not in time. In contrast, the control policy on the right is predominantly 0, with occasional deviations, making it noticeably different from what might be considered its 'equivalent' policy on the left. This obvious distinction leads to the question of whether there is a mistake in the Pseudocode or a programming error in the implementation or whether there is a theoretical twist that was overseen when discussing the RMC approach. Assuming the validity of the Pseudocode, the correctness of the implementation is validated in the following by examining and testing different parts individually.

The residual demand Chapter 4 was extended in a way that setting $\lambda = 0$ leads back to Chapter 3. Therefore, both Algorithms should return a value function and control policy that coincides with Section 3.4. Indeed, Figure 5.3 and Figure 5.4 return matching numerical approximations.

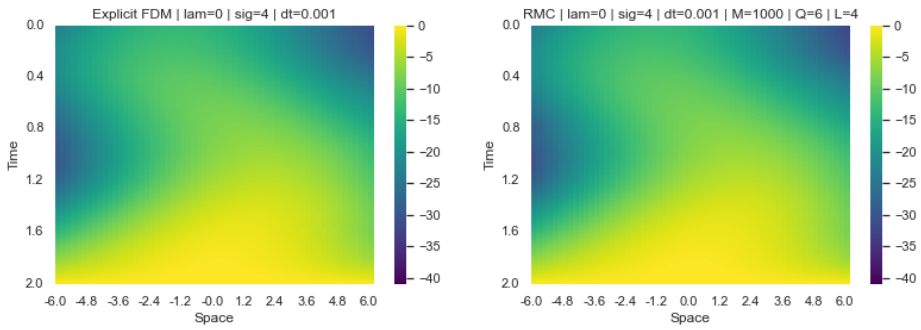

Figure 5.3: Heatmap of the optimal value function provided by the explicit FDM (left) and RMC (right) with vanishing penalization term $\lambda = 0$, $\Delta t = 0.001$, $h = 0.2$ and $\sigma = 4$

The difference compared to the numerical solution proposed by the implicit FDM in Figure 3.3 is 0.0166 for the explicit FDM and 0.158 for RMC in the $||\cdot||_{gl}$ norm. Hence, the error is reasonably low validating the correctness of the approximations. Same results are observed for the corresponding optimal control policies.

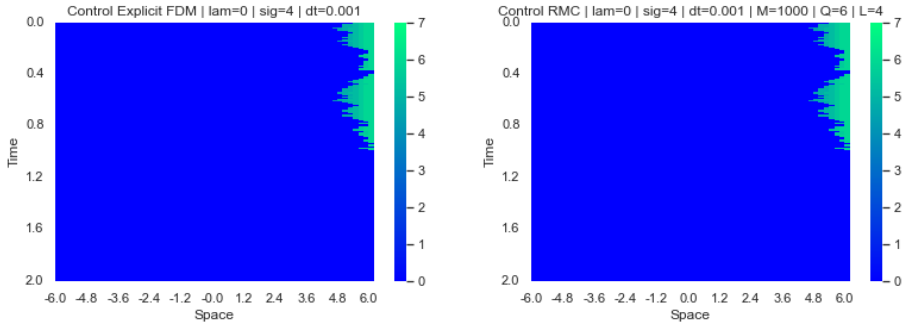

Figure 5.4: Heatmap of optimal control policy provided by the explicit FDM (left) and RMC (right) with vanishing penalization term $\lambda = 0$, $\Delta t = 0.001$, $h = 0.2$ and $\sigma = 4$

Both numerical approximations to the optimal control policy are compared to the analytical one given in Figure 3.4 and lead to the errors 0.0584 for the

FDM and 0.0488 RMC approach in the $||\cdot||_{gl}$ norm. As both errors remain reasonably low, they consequently validate the optimizing part within Algorithm 3.

This test suggests the idea, that the approximations to the value function begin to diverge when λ increases. The argument is examined by modifying Algorithm 3 such that it reconstructs the known verification function $v(s,x) = (T-s)x^2$ for different parameters. This function upholds the terminal condition of Section 4.1 and prevents the potential source of error stemming from an unfavourable pick of basis functions, that is set $Q = 2$. Throughout the modified implementation, the cost function is substituted with an adequate forcing term as done in Section 3.5.1. Table 5.1 captures the error in $||\cdot||_{gl}$ for different levels L of optimized controls, varying discretization sizes Δt and number of simulations M.
Note, that the errors in Table 5.1 have randomness to them and the omission of confidence intervals is justified once again by the global characteristic of the error norm across the entire DoI. Table 5.1 shows, that the overall error across all tables Table 5.1a, Table 5.1b and Table 5.1c increases with higher penalization term λ and supports the suggestion that λ induces an error and distorts the numerical approximation.

Table 5.1a shows, that reducing time discretization Δt leads to a decreasing error. In the case $\Delta t = 0.01$ it continues to decrease when increasing the number of trajectories M. This result doesn't hold for coarser Δt and seems to persist for $L = 1, 3, 5$. The level of optimized controls L, however, is neglectable due to $\lambda = 0$. The discrepancies among the errors for different L shows the variability of the error norm $||\cdot||_{gl}$. With this explanation, the error in $L = 1, M = 100, \Delta t = 0.01$ seems to be particularly high, whereas the error in $L = 3, M = 500, \Delta t = 0.05$ appears to be especially low.

In Table 5.1b and Table 5.1c with $\lambda = 1$ and $\lambda = 2$ respectively, the error tends to decrease once more when refining Δt. An exception to this affirmation is given by the results for $\lambda = 1, 2$ when $L = 1$. Moreover, higher M seem to decrease the error for most Δt and L. The error reduction by increasing M from 100 to 500 tends to be greater than if M is augmented from 500 to 1000. Noticeable exceptions to this assumption exist for $\lambda = 1$ when $L = 1$ and $\Delta t = 0.01, 0.05$ or when $L = 5$ and $\Delta t = 0.1$. For $\lambda = 2$ they arise for $L = 5$ and $\Delta t = 0.1$. Unexpectedly, the error seems to grow in Table 5.1b and Table 5.1c when increasing the level of accuracy L of the trajectories, i.e. the errors tend to worsen if the trajectories are optimized. This leads to the questionability of the discussion held in Section 4.3 about

the local adjustments of the trajectories.

	L=1			L=3			L=5		
$M \setminus \Delta t$	0.01	0.05	0.1	0.01	0.05	0.1	0.01	0.05	0.1
100	0.9213	1.6963	2.8991	0.4007	1.2702	2.9764	0.476	1.2739	3.0736
500	0.5212	1.5372	3.1626	0.3883	1.4356	3.0642	0.5698	1.3007	3.0574
1000	0.2832	1.6594	2.9459	0.3325	1.5703	2.9224	0.3539	1.374	3.1666

(a) $\lambda = 0$

$M \setminus \Delta t$	0.01	0.05	0.1	0.01	0.05	0.1	0.01	0.05	0.1
100	1.9333	1.5623	2.6709	4.0673	4.086	5.126	4.1849	3.9212	5.0409
500	1.9051	1.6906	2.5158	2.9677	3.7496	4.9724	3.0573	3.7284	4.9011
1000	1.9452	1.6385	2.3993	2.8939	3.4478	4.8921	2.8362	3.6679	5.1329

(b) $\lambda = 1$

$M \setminus \Delta t$	0.01	0.05	0.1	0.01	0.05	0.1	0.01	0.05	0.1
100	4.474	3.8274	3.7517	4.5162	5.2201	6.7583	4.4745	4.6127	6.1794
500	4.1922	3.7356	3.6543	3.8335	4.6384	5.918	3.9736	4.5493	5.839
1000	4.085	3.7986	3.661	3.8743	4.6171	5.8839	3.7731	4.5965	6.0175

(c) $\lambda = 2$

Table 5.1: Error comparison for different approximations of $v(s,x) = (T - s)x^2$ in $[0,\dots,2] \times [-6,\dots,6]$ using the $||\cdot||_{gl}$ norm. Changing parameters are the penalization term λ, the number of MC simulations M, the time discretization sizes Δt and the levels of accuracy L for the trajectories

The results obtained and presented throughout this chapter indicate a significant distortion of the value function, that occurs when the penalization term λ exceeds 0. This distortion is particularly evident when the residual demand is transformed into a degenerate form and contradicts the validity of the Importance Sampling approach introduced in Section 4.3.

6 Conclusion

The aim at the outset of this thesis was to get familiar with stochastic optimal control problems and deepen the understanding through practical application. It started by exploring the underlying theory, serving as the basis upon which the hands-on examples were constructed. A simple grid-connected microgrid was set up and subjected to a range of numerical solutions. Subsequently, the microgrid was modified by inserting the control policy into the residual demand process.
This exploration led to the emergence of a change of measure idea, that alters the information provided by the trajectory during the conditional expectation regression. In the beginning, the concept seemed to circumvent the conventional theory of BSDE. However, practical outcomes did not align with the initial expectations. Instead, the numerical algorithms yielded two distinct solutions, with such divergent values that parameter adjustments failed to reconcile them. In response, a detailed examination of Algorithm 3 was performed. The model was restricted to the one presented in Chapter 3 by setting $\lambda = 0$, which led to precise results. Additionally, the algorithm's outcomes were assessed against a known function, shedding light on the error arising from the penalization term.

Future research could delve into potential issues, that arose during this work. A thorough exploration of the theoretical discussion of the measure change concept and its relationship with BSDE theory could serve as a cornerstone for further research. A comprehensive convergence analysis, exploring the limitations of the concept and potential stability criteria, might be a significant step in advancing this line of inquiry.
Furthermore, the thesis hints at the need for a more comprehensive investigation into various elements of Algorithm 3. A complete convergence analysis that encompasses critical parameters, including the time step Δt, spatial step h, the trajectory optimization level L, the penalization term λ, the choice of optimization methods, and the difficulties of the regression technique would lead to a broader comprehension. Also, non-linear dependencies between

regression coefficients and basis functions, possibly employing feature maps (cf. [25]) or Neural Networks hold the potential to significantly enhance the precision of conditional expectations estimation.

In summary, the thesis explored a particular facet of stochastic optimal control problems, offering both theoretical challenges and practical insights.

Acronyms

BE Backward Euler method

BSDE Backwards Stochastic Differential Equation

DoI Domain of Interest

DOM Dominated Convergence Theorem

DPP Dynamic Programming Principle

EM Euler Maruyama

FDM Finite Difference Method

FE Forward Euler method

GBM Geometric Brownian motion

HJB Hamilton-Jacobi-Bellman equation

LRAM Left Rectangle Approximation Method

MC Monte Carlo

MSE Mean squared error

ODE Ordinary Differential Equation

PDE Partial Differential Equation

pdf probability density function

RMC Regression Monte Carlo

SDE Stochastic Differential Equation

Bibliography

[1] Julien Nicolas. *Stochastic optimal control applied to the optimization of a microgrid producing intermittent energy*. 2023, Master thesis at Ecole Polytechnique de Lausanne.

[2] Pierre Haessig et al. "Computing an optimal control policy for an energy storage". In: *Proceedings of the 6th European Conference on Python in Science (EuroSciPy 2013)* (2014). URL: https://doi.org/10.48550/arXiv.1404.6389.

[3] Alexander Boogert and Cyriel De Jong. "Gas storage valuation using a Monte Carlo method". In: *The Journal of Derivatives* 15 (3 2008), pp. 81–98. URL: https://doi.org/10.3905/jod.2008.702507.

[4] Clemence Alasseur et al. "Regression Monte Carlo for Microgrid Management". In: *Esaim: Proceedings and Surveys* 65 (2019), pp. 46–67. URL: https://doi.org/10.48550/arXiv.1802.10352.

[5] Francis A Longstaff and Eduardo S Schwartz. "Valuing American options by simulation: a simple least-squares approach". In: *The Review of Financial Studies* 14 (1 2001), pp. 113–147. URL: https://doi.org/10.1093/rfs/14.1.113.

[6] Fabio Nobile. *Stochastic simulation*. 2022, Lecture Notes in EPFL Moodle, Accessed during fall semester 2022/2023.

[7] Alessandro Balata and Jan Palczewski. "Regress-later Monte Carlo for optimal inventory control with applications in energy". In: *Archive Preprint* (2017). URL: https://doi.org/10.48550/arXiv.1703.06461.

[8] Jared Chessari et al. "Numerical methods for backward stochastic differential equations: A survey". In: *Probability Surveys* 20 (2023), pp. 486–567. URL: https://doi.org/10.1214/23-PS18.

[9] Huyên Pham. *Continuous-time stochastic control and optimization with financial applications*. Vol. 61. Springer Science & Business Media, 2009. URL: https://doi.org/10.1007/978-3-540-89500-8.

[10] Jiongmin Yong and Xun Yu Zhou. *Stochastic Controls: Hamiltonian Systems and HJB Equations*. Vol. 43. Springer Science & Business Media, 1999. URL: https://doi.org/10.1007/978-1-4612-1466-3.

[11] Alexander S Cherny et al. *Singular stochastic differential equations*. Vol. 1. Springer Science & Business Media, 2005. URL: https://doi.org/10.1007/b104187.

[12] Dimitri P Bertsekas et al. "Dynamic programming and optimal control: Volume II, 3rd Edition". In: *Belmont, MA: Athena Scientific* 1 (2011). Accessed 3rd October 2023. URL: https://web.mit.edu/dimitrib/www/dpchapter.pdf.

[13] Nizar Touzi and Agnès Tourin. *Optimal stochastic control, stochastic target problems, and backward SDE*. Vol. 29. Springer Science & Business Media, 2013. URL: https://doi.org/10.1007/978-1-4614-4286-8.

[14] Lawrence Evans. *An Introduction to Mathematical, Optimal Control Theory*. 1983, Lecture Notes, Accessed 26th August 2023.

[15] Idris Kharroubi, Nicolas Langrené, and Huyên Pham. "A numerical algorithm for fully nonlinear HJB equations: An approach by control randomization". In: *Monte Carlo Methods and Applications* 20.2 (2014), pp. 145–165. URL: https://doi.org/10.1515/mcma-2013-0024.

[16] J Michael Steele. *Stochastic Calculus and Financial Applications*. Vol. 45. Springer Science & Business Media, 2001. URL: https://doi.org/10.1007/978-1-4684-9305-4.

[17] Wendell H Fleming and Halil Mete Soner. *Controlled Markov Processes and Viscosity Solutions*. Vol. 2. Springer Science & Business Media, 2006. URL: https://doi.org/10.1007/0-387-31071-1.

[18] Martin Frank. *Optimierungstheorie*. 2022, Lecture Notes in KIT Intranet, Accessed during fall semester 2022/2023.

[19] John Hunter. *Notes on Partial Differential Equations*. 2014, Lecture Notes, Accessed 20th September 2023.

[20] Alfio Quarteroni, Fausto Saleri, Paola Gervasio, et al. *Scientific computing with MATLAB and Octave*. Vol. 4. Springer Science & Business Media, 2006. URL: https://doi.org/10.1007/978-3-642-45367-0.

[21] Alfio Quarteroni and Silvia Quarteroni. *Numerical models for differential problems*. Vol. 2. Springer Science & Business Media, 2009. URL: https://doi.org/10.1007/978-88-470-5522-3.

[22] Adrian Blumenthal. *Numerical integration of stochastic differential equations*. 2023, Lecture Notes in EPFL Moodle, Accessed during summer semester 2023.

[23] Fabio Nobile. *Introduction to Partial Differential Equations*. 2022, Lecture Notes in EPFL Moodle, Accessed during fall semester 2022/2023.

[24] Mathias Trabs. *Wahrscheinlichkeitstheorie*. 2022, Lecture Notes in KIT Intranet, Accessed during summer semester 2022.

[25] Guillaume Obozinski. *Statistical Machine Learning*. 2022, Lecture Notes in EPFL Moodle, Accessed during fall semester 2022/2023.

[26] Rick Durrett. *Probability: Theory and Examples*. Version 5, (2019), https://services.math.duke.edu/∼rtd/PTE/PTE5_011119.pdf, Accessed 26th August 2023.

[27] John Tsitsiklis. *Probabilistic Systems Analysis and Applied Probability*. 2013, Lecture Notes, Accessed 15th September 2023.

[28] Luc Devroye et al. *Recent developments in applied probability and statistics: dedicated to the memory of Jürgen Lehn*. Vol. 1. Springer Science & Business Media, 2010. URL: https://doi.org/10.1007/978-3-7908-2598-5.

Erklärung

Ich versichere wahrheitsgemäß, die Arbeit selbstständig verfasst, alle benutzten Hilfsmittel vollständig und genau angegeben und alles kenntlich gemacht zu haben, was aus Arbeiten anderer unverändert oder mit Abänderungen entnommen wurde, sowie die Satzung des KIT zur Sicherung guter wissenschaftlicher Praxis in der jeweils gültigen Fassung beachtet zu haben.

Karlsruhe, 4ter Oktober 2023