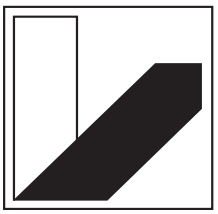


UNIVERSITÄT
BAYREUTH

Model Predictive Control for the Fokker-Planck Equation

Masterarbeit

von

Arthur Fleig

FAKULTÄT FÜR MATHEMATIK, PHYSIK UND INFORMATIK
MATHEMATISCHES INSTITUT

Datum: 31. Januar 2014

Betreuung:
Prof. Dr. L. Grüne
Prof. Dr. H.-J. Pesch
Dipl.-Math. N. Altmüller

Für E.

Danksagung

Zunächst möchte ich mich gerne bei Herrn Prof. Dr. Grüne für die stets ausgezeichnete Betreuung und Beratung während der Entstehung dieser Arbeit bedanken. Weiterhin bedanke ich mich recht herzlich bei Herrn Prof. Dr. Pesch und Nils Altmüller, die immer ein offenes Ohr für meine Fragen hatten und mir mit motivierenden Anregungen bei Problemen zur Seite standen.
Ganz besonderer Dank gebührt meiner Frau für ihre liebe Fürsorge und die viele Geduld mit mir. Außerdem möchte ich mich insbesondere bei meinen Eltern für die Unterstützung während meines Studiums bedanken.

Contents

Chapter 1

Introduction

This work concerns itself with optimal control of one-dimensional stochastic processes by controlling the associated probability density functions using a Receding Horizon Model Predictive Control scheme.

In the 1960s, Hestenes and Pontrjagin laid the foundation for optimal control theory, which deals with maximizing or minimizing functionals, i.e. mappings from a set of functions to real numbers, with side conditions determined by differential equations [23], [24]. Since the 1970s, controlling systems governed by partial differential equations (PDEs) is of interest [21] and it is becoming an increasingly active research field. Obtaining an open-loop optimal control for PDE-constrained problems is treated e.g. in [16], [21], and [30].

When it comes to designing closed-loop controllers, Model Predictive Control (MPC) is very appealing, especially for nonlinear problems. Because of its robustness properties and the ability to easily accommodate control- and state constraints, MPC is utilized in a variety of applications, e.g. chemical process control, electro-mechanical systems, aerospace applications, and many more [11], [25].

Alongside, in the last few decades, researchers have concerned themselves more and more with the modeling and control of uncertainty. This is not surprising, considering the wide range of applications. Much more recently, in 2011, Annunziato and Borzì presented a framework which connects both the optimal control of systems governed by partial differential equations and the control of stochastic processes [3]. To this end, the Fokker-Planck equation, a parabolic PDE, is used to model the evolution of the probability density function (PDF) of certain stochastic processes. Since often the desired outcome can be formulated in terms of the PDF, this approach is viable. This way, the problem is transformed from stochastic to deterministic. A Receding Horizon Model Predictive Control scheme is then used to obtain an optimal closed-loop controller.

The thesis aims to contribute to the above framework in two ways. First, by examining the influence of the horizon length in the MPC scheme on the optimal control. In [3] and [4], only the shortest possible horizon is considered. The question at hand is whether increasing the horizon length yields better results. It is shown numerically that increasing the horizon may substantially improve results under certain circumstances. Yet, it seems more important to cover a certain absolute time frame in the MPC scheme. The second objective

is to investigate the effects of a space-dependent control in contrast to a space-independent one. Numerical experiments show that a space-dependent control can improve the tracking of the desired probability density function tremendously; under certain circumstances, the computed probability density function barely deviates from the desired one. In order to run these numerical simulations, discretization and optimization algorithms were implemented, constituting the practical part of the thesis.
This work is structured as follows. In Chapter 2 the Fokker-Planck (FP) control framework is presented. The connection between optimal control of certain stochastic processes and the Fokker-Planck equation is established. After discussing existence and uniqueness results for the FP equation, the corresponding optimal control problems (OCPs) are introduced. Existence of optimal solutions is addressed, however briefly, and necessary optimality conditions are derived formally. Chapter 3 gives a brief presentation on MPC and shows how to translate the FP optimal control problems into the MPC framework. Chapter 4 describes the discretization and optimization techniques used to obtain numerical results. In Chapter 5 the focus is on the implementation of the presented framework. The numerical results are presented in Chapter 6; Chapter 7 concludes.

Chapter 2

The Fokker-Planck Control Framework

In this chapter, the Fokker-Planck Control Framework for one-dimensional stochastic processes is introduced. The focus is on Markov processes, which are discussed in Section 2.1 along with stochastic differential equations (SDEs). Instead of trying to control these processes directly, an alternative approach is pursued by controlling the respective probability density functions (PDFs). This approach is viable when the collective behavior of the process is of interest.
The state PDF of a Markov process can be obtained by several ways. One way is to employ averaging or interpolation strategies in the stochastic models, see [8], [18]. Another way was discussed by Annunziato and Borzì in [3] and [4]: Starting from a given initial state PDF, the Fokker-Planck (FP) equation is used to model the evolution of the state PDF. This relationship is derived in Section 2.2. Furthermore, results concerning existence and uniqueness of (non-negative) solutions of the FP equation are provided. The Fokker-Planck equation is a partial differential equation (PDE) of parabolic type [28], connecting both the control of stochastic processes and the optimal control of systems governed by PDEs, thereby reformulating the problem from stochastic to deterministic. The corresponding optimal control problem is treated in Section 2.3, along with a discussion of existence and uniqueness of optimal controls. In addition, the necessary optimality conditions are derived using the formal Lagrange technique introduced in [30].

2.1 Stochastic Processes and Differential Equations

A stochastic process is a collection of random variables, used to describe the evolution of these variables (or a subset thereof) over time. Formally, it is defined as follows:

Definition 2.1 (Stochastic process)
Let $(\Omega, \mathcal{F}, P)$ be a probability space, where Ω is the space of outcomes, $\mathcal{F}$ is the σ-algebra of events in Ω and $P\colon \Omega \to [0,1]$ is a probability measure. Further, let $(Z, \mathcal{Z})$ be a measurable

space and I a totally ordered set. A family $X = (X_t)_{t\in I}$ of measurable functions

$$X_t \colon \Omega \to Z, \quad t \in I$$

is called stochastic process *(with state space Z).*

The focus is on continuous-time stochastic processes, i.e. I is an interval with $I \subseteq \mathbb{R}_{\geq 0}$, e.g. $I = \mathbb{R}_{\geq 0}$. A particularly important one for this work is the Wiener Process, also called standard Brownian motion.

Definition 2.2 (Wiener Process, Brownian motion)
A stochastic process $W = (W_t)_{t\geq 0}$ is called Wiener process *(or* Brownian motion*), if the following conditions hold:*

1. *$W_0 = 0$ almost surely.*
2. *$(W_t)_{t\geq 0}$ has independent increments with $W_t - W_s \sim \mathcal{N}(0, t-s), 0 \leq s < t$, i.e. the increments are normally distributed with mean 0 and variance $t - s$.*
3. *$(W_t)_{t\geq 0}$ is a continuous process, i.e. almost all paths (functions $t \mapsto W_t$) of $(W_t)_{t\geq 0}$ are continuous.*

The Wiener Process is omnipresent in scientific literature when it comes to modeling uncertainty. It also appears in the models considered in this work. These models are described by so-called Itô stochastic differential equations, a probabilistic extension of ordinary differential equations.

Definition 2.3 (Stochastic differential equation)
Let $n \in \mathbb{N}$. Given functions $b, \sigma \colon \mathbb{R}^n \times \mathbb{R}_{\geq 0} \to \mathbb{R}$ and a Wiener process $(W_t)_{t\geq 0}$, an (Itô) stochastic differential equation (SDE) *is given by*

$$dX_t = b(X_t, t)\, dt + \sigma(X_t, t)\, dW_t, \tag{2.1}$$

which is a shorter notation for the integral equation

$$X_t = X_0 + \int_0^t b(X_\tau, \tau)\, d\tau + \int_0^t \sigma(X_\tau, \tau)\, dW_\tau. \tag{2.2}$$

The first integral is interpreted as a Lebesgue integral, the second as a so-called Itô integral.

Remark 2.4
The function b is called drift function, σ is the diffusion function. While the drift term shifts the state X_t, the diffusion term indicates the spreading of the state.

In a next step, the control u is added. Only one-dimensional, continuous-time stochastic processes described by the following SDE and initial condition are considered:

$$\begin{aligned} dX_t &= b(X_t, t; u)\, dt + \sigma(X_t, t)\, dW_t \\ X_{t_0} &= X_0 \end{aligned} \tag{2.3}$$

with $X_t, X_{t_0} \in \mathbb{R}$, $b, \sigma \colon \mathbb{R} \times \mathbb{R}_{\geq 0} \to \mathbb{R}$. The control u is either a (given) function of time, i.e. $t \mapsto u(t) \in \mathbb{R}$, or a (given) function of time and space: $(x,t) \mapsto u(x,t) \in \mathbb{R}$. (2.3) is a so-called Markov process. More information regarding Markov processes, stochastic processes in general and Itô integrals can be found e.g. in [9] and [19].
In order to ensure existence and uniqueness of solutions X_t of (2.3), certain requirements need to be satisfied. Therefore, the following assumption on b and σ is made, cf. [9], [19]:

Assumption 2.5
Let $T > 0$. The functions b and σ in (2.3) satisfy the following conditions for a given realization of W_t: $\exists C > 0 \ \forall x, y \in \mathbb{R}, t \in [0,T], u \in \mathbb{R}$:

$$|b(x,t;u) - b(y,t;u)| + |\sigma(x,t) - \sigma(y,t)| \leq C|x-y| \qquad \textit{(global Lipschitz condition for } b, \sigma\textit{)}$$
$$|b(x,t;u)|^2 + |\sigma(x,t)|^2 \leq C^2(1+|x|^2) \qquad \textit{(at most linear growth)}$$

The processes relevant to this work are illustrated below. Notice that if u is bounded, Assumption 2.5 only has to hold for the relevant values, i.e. all u with $|u| \leq M$ for a suitable $M > 0$.

Example 2.6
(a) The Ornstein-Uhlenbeck Process with Additive Control *is defined by*

$$b(x,t;u) := -\mu x + u$$
$$\sigma(x,t) \equiv \tilde{\sigma}$$

with $\mu, \tilde{\sigma} > 0$. It describes the velocity X_t of a massive particle immersed in a viscous fluid subject to random collisions with other molecules of the fluid and the consequential fluctuations. An external force field induces a momentum u and thus influences the velocity of the particle. The values μ and $\tilde{\sigma}$ indicate the impact of the current speed X_t and the randomness W_t on the rate of change dX_t, respectively.
If the control u is a function of time, i.e. $u(t)$, and u is bounded, Assumption 2.5 holds:

$$\begin{aligned} &|b(x,t;u(t)) - b(y,t;u(t))| + |\sigma(x,t) - \sigma(y,t)| \\ =\ &| -\mu x + u(t) + \mu y - u(t)| + |\tilde{\sigma} - \tilde{\sigma}| \\ \leq\ &|\mu||x-y| = \mu|x-y| = C_1|x-y| \end{aligned}$$

with $C_1 := \mu$. Furthermore,

$$\begin{aligned} &|b(x,t;u(t))|^2 + |\sigma(x,t)|^2 = |-\mu x + u(t)|^2 + |\tilde{\sigma}|^2 \\ \leq\ &\left(|\mu x| + |u(t)|\right)^2 + |\tilde{\sigma}|^2 = |\mu x|^2 + \underbrace{2|\mu x||u(t)|}_{\leq |\mu x|^2 + |u(t)|^2} + |u(t)|^2 + |\tilde{\sigma}|^2 \\ \leq\ &2|\mu|^2|x|^2 + 2|u(t)|^2 + |\tilde{\sigma}|^2 = 2\mu^2|x|^2 + 2|u(t)|^2 + \tilde{\sigma}^2 \leq C_2^2(1+|x|^2) \end{aligned}$$

with $C_2 := \max\left\{\sqrt{2}\mu, \sqrt{2\left(\max_t |u(t)|^2\right) + \tilde{\sigma}^2}\right\}$. Define $C := \max\{C_1, C_2\}$.

For $u(x,t)$, this is not as simple, even with a bounded control. In this case, it is sufficient to additionally assume u being (globally) Lipschitz-continuous:

$$\begin{aligned}
&|b(x,t;u(x,t)) - b(y,t;u(y,t))| + |\sigma(x,t) - \sigma(y,t)| \\
&= |-\mu x + u(x,t) + \mu y - u(y,t)| \\
&\leq |\mu||x-y| + \underbrace{|u(x,t) - u(y,t)|}_{\overset{!}{\leq} L|x-y| \text{ for some } L>0 \text{ and all } t\in[0,T]} \\
&\leq (|\mu| + L)|x-y| = (\mu + L)|x-y| = C_1|x-y|
\end{aligned}$$

with $C_1 := \mu + L$ and, analogously to $u(t)$,

$$\begin{aligned}
&|b(x,t;u(x,t))|^2 + |\sigma(x,t)|^2 = |-\mu x + u(x,t)|^2 + |\tilde{\sigma}|^2 \\
&\leq 2\mu^2|x|^2 + 2|u(x,t)|^2 + \tilde{\sigma}^2 \leq C_2^2(1+|x|^2)
\end{aligned}$$

with $C_2 := \max\left\{\sqrt{2}\mu, \sqrt{2\left(\max_{x,t}|u(x,t)|^2\right) + \tilde{\sigma}^2}\right\}$ *and* $C := \max\{C_1, C_2\}$.

(b) The Geometric-Brownian Process with Additive Drift Control *is defined by*

$$\begin{aligned}
b(x,t;u) &:= (\mu + u)x \\
\sigma(x,t) &:= \tilde{\sigma}x
\end{aligned}$$

with $\mu \in \mathbb{R}, \tilde{\sigma} > 0$. It can be used to model a controlled portfolio consisting of risky assets that evolve at a varying, possibly negative interest rate, and bonds, which offer a risk-free return. In this case, X_t represents the wealth, μ is the mean return and $\tilde{\sigma}$ is the volatility parameter. The control u is a portion of the portfolio invested in bonds. For a time-dependent and bounded control $u(t)$, Assumption 2.5 holds:

$$\begin{aligned}
&|b(x,t;u(t)) - b(y,t;u(t))| + |\sigma(x,t) - \sigma(y,t)| \\
&= |(\mu + u(t))x - (\mu + u(t))y| + |\tilde{\sigma}(x-y)| \\
&\leq (|\mu| + |\tilde{\sigma}| + |u(t)|)|x-y| \leq C_1|x-y|
\end{aligned}$$

where $C_1 := \left(|\mu| + \tilde{\sigma} + \max_t |u(t)|\right)$ *and*

$$\begin{aligned}
&|b(x,t;u(t))|^2 + |\sigma(x,t)|^2 = |(\mu + u(t))x|^2 + |\tilde{\sigma}x|^2 \\
&\leq |\mu + u(t)|^2|x|^2 + |\tilde{\sigma}|^2|x|^2 = \left((\mu + u(t))^2 + |\tilde{\sigma}|^2\right)|x|^2 \\
&= \big(\mu^2 + \underbrace{2\mu u(t)}_{\leq \mu^2 + u(t)^2} + u(t)^2 + \tilde{\sigma}^2\big)|x|^2 \leq \left(2\mu^2 + 2u(t)^2 + \tilde{\sigma}^2\right)|x|^2 \leq C_2^2(1+|x|^2),
\end{aligned}$$

where $C_2 := \sqrt{2\mu^2 + 2\left(\max_t u(t)^2\right) + \tilde{\sigma}^2}$ *and* $C := \max\{C_1, C_2\}$.

Unfortunately, this is not true for $u(x,t)$:

$$\begin{aligned} & |b(x,t;u(x,t)) - b(y,t;u(y,t))| + |\sigma(x,t) - \sigma(y,t)| \\ = \; & |(\mu + u(x,t))x - (\mu + u(y,t))y| + |\tilde{\sigma}(x-y)| \\ = \; & |\mu(x-y) + u(x,t)x - u(y,t)y| + |\tilde{\sigma}(x-y)|. \end{aligned}$$

Even for simple, Lipschitz continuous functions like $u(x,t) := x$, the condition is not met:

$$\begin{aligned} & |\mu(x-y) + u(x,t)x - u(y,t)y| + |\tilde{\sigma}(x-y)| \\ = \; & |\mu(x-y) + x^2 - y^2| + |\tilde{\sigma}(x-y)| \\ = \; & |(\mu + x + y)||x-y| + |\tilde{\sigma}||x-y| \\ \geq \; & \underbrace{|\mu + x + y + \tilde{\sigma}|}_{> L \text{ for any } L>0 \text{ and suitable } x,y \in \mathbb{R}} |x-y|. \end{aligned}$$

However, if only localized systems are of interest, i.e. $|x|, |y| \leq M$ for some $M > 0$, then a bounded and Lipschitz continuous function $u(x,t)$ is sufficient to satisfy Assumption 2.5 for the relevant values:

$$\begin{aligned} & |\mu(x-y) + u(x,t)x - u(y,t)y| + |\tilde{\sigma}(x-y)| \\ = \; & |(\mu + x + y)||x-y| + |\tilde{\sigma}||x-y| \\ \leq \; & |\mu||x-y| + |u(x,t)x - u(y,t)y| + \tilde{\sigma}|x-y| \\ = \; & (|\mu| + \tilde{\sigma})\,|x-y| + |u(x,t)x - u(y,t)x + u(y,t)x - u(y,t)y| \\ \leq \; & (|\mu| + \tilde{\sigma})\,|x-y| + \underbrace{|x|}_{\leq M} \underbrace{|u(x,t) - u(y,t)|}_{\leq L|x-y| \text{ for some } L>0 \text{ and all } t \in [0,T]} + |u(y,t)||x-y| \\ \leq \; & (|\mu| + \tilde{\sigma} + ML + |u(y,t)|)\,|x-y| \leq \; C_1|x-y| \end{aligned}$$

with $C_1 := \left(|\mu| + \tilde{\sigma} + ML + \max_{y,t} |u(y,t)|\right)$. *Furthermore,*

$$\begin{aligned} & |b(x,t;u(t))|^2 + |\sigma(x,t)|^2 = |(\mu + u(x,t))x|^2 + |\tilde{\sigma}x|^2 \\ = \; & |\mu + u(x,t)|^2|x|^2 + \tilde{\sigma}^2|x|^2 = \left((\mu + u(x,t))^2 + \tilde{\sigma}^2\right)|x|^2 \\ = \; & \big(\mu^2 + \underbrace{2\mu u(x,t)}_{\leq \mu^2 + u(x,t)^2} + u(x,t)^2 + \tilde{\sigma}^2\big)|x|^2 \leq \left(2\mu^2 + 2u(x,t)^2 + \tilde{\sigma}^2\right)|x|^2 \leq C_2^2(1 + |x|^2), \end{aligned}$$

with $C_2 := \sqrt{2\mu^2 + 2\left(\max_{x,t} u(x,t)^2\right) + \tilde{\sigma}^2}$ *and* $C := \max\{C_1, C_2\}$.

(c) The Shiryaev Process *is defined by*

$$\begin{aligned} b(x,t;u) &:= u + \mu x \\ \sigma(x,t) &:= \tilde{\sigma} x \end{aligned}$$

with $\mu \in \mathbb{R}, \tilde{\sigma} > 0$. This setting is similar to the Geometric-Brownian Process, the difference being that the control u is no longer factored by X_t. In this case, u can be interpreted as the sum of absolute fixed income from bonds and outcome consumption. For bounded $u(t)$, Assumption 2.5 is satisfied:

$$\begin{aligned} & |b(x,t;u(t)) - b(y,t;u(t))| + |\sigma(x,t) - \sigma(y,t)| \\ = \; & |\mu x + u(t) - \mu y - u(t)| + |\tilde{\sigma}(x-y)| \\ \leq \; & (|\mu| + |\tilde{\sigma}|)\,|x-y| = C_1|x-y| \end{aligned}$$

with $C_1 := |\mu| + \tilde{\sigma}$ and

$$\begin{aligned} & |b(x,t;u(t))|^2 + |\sigma(x,t)|^2 = |\mu x + u(t)|^2 + |\tilde{\sigma}x|^2 \\ = \; & (\mu x)^2 + \underbrace{2\mu x u(t)}_{\leq (\mu x)^2 + u(t)^2} + u(t)^2 + \tilde{\sigma}^2|x|^2 \\ \leq \; & \left(2\mu^2 + \tilde{\sigma}^2\right)|x|^2 + 2u(t)^2 \leq C_2^2(1+|x|^2) \end{aligned}$$

with $C_2 := \max\left\{\sqrt{2\mu^2+\tilde{\sigma}^2}, \sqrt{2\max_t u(t)^2}\right\}$ and $C := \max\{C_1, C_2\}$.
As in (a), a Lipschitz-continuous control $u(x,t)$ is assumed in order to satisfy Assumption 2.5 for bounded $u(x,t)$:

$$\begin{aligned} & |b(x,t;u(x,t)) - b(y,t;u(y,t))| + |\sigma(x,t) - \sigma(y,t)| \\ = \; & |\mu x + u(x,t) - \mu y - u(y,t)| + |\tilde{\sigma}(x-y)| \\ \leq \; & (|\mu| + |\tilde{\sigma}|)\,|x-y| + \underbrace{|u(x,t) - u(y,t)|}_{\overset{!}{\leq} L|x-y| \text{ for some } L>0 \text{ and all } t\in[0,T]} \\ \leq \; & (|\mu| + |\tilde{\sigma}| + L)\,|x-y| = C_1|x-y| \end{aligned}$$

with $C_1 := |\mu| + \tilde{\sigma} + L$ and, analogously to $u(t)$,

$$\begin{aligned} & |b(x,t;u(x,t))|^2 + |\sigma(x,t)|^2 = |\mu x + u(x,t)|^2 + |\tilde{\sigma}x|^2 \\ \leq \; & \left(2\mu^2 + \tilde{\sigma}^2\right)|x|^2 + 2u(x,t)^2 \leq C_2^2(1+|x|^2) \end{aligned}$$

with $C_2 := \max\left\{\sqrt{2\mu^2+\tilde{\sigma}^2}, \sqrt{2\max_{x,t} u(x,t)^2}\right\}$ and $C := \max\{C_1, C_2\}$.

2.2 The Fokker-Planck Equation

As mentioned in the beginning of this chapter, the evolution of the state probability density function of a Markov process can be modeled by the Fokker-Planck equation. To this end, let $y(x,t)$ be the probability density to find the process X_t at $x \in \Omega \subset \mathbb{R}$ at time $t \geq 0$. Since the evolution of the PDF is of interest, a function describing the transition density

PDF for the Markov process to move from $z \in \Omega$ at time $s \geq 0$ to $x \in \Omega$ at time $t \geq s$ is employed. Let this function be denoted by $\hat{y}(x,t;z,s)$. This implies

$$\hat{y}(x,s;z,s) = \delta(x-z),$$

where δ is the Dirac delta function. Both y and $\hat{y}$ depict probabilities and therefore have the following properties:

$$\begin{aligned} y(x,t) &\geq 0, \quad (x,t) \in \Omega \times [0,\infty[, \\ \int_\Omega y(x,t)\,dx &= 1, \quad t \in [0,\infty[, \\ \hat{y}(x,t;z,s) &\geq 0, \quad x,z \in \Omega, s,t \in [0,\infty[, t \geq s, \\ \int_\Omega \hat{y}(x,t;z,s)\,dx &= 1, \quad z \in \Omega, s,t \in [0,\infty[, t \geq s. \end{aligned}$$

Note that the focus is on localized systems and Ω is chosen large enough such that the conditions

$$\begin{aligned} y(x,t) &= 0, \quad (x,t) \in \partial\Omega \times [0,\infty[, \\ \hat{y}(x,t;z,s) &= 0, \quad x \in \partial\Omega, z \in \Omega, s,t \in [0,\infty[, t \geq s \end{aligned}$$

can be assumed with a negligible error made by disregarding $\mathbb{R} \setminus \Omega$. The evolution of the transition probability density $\hat{y}$ is modeled by the Fokker-Planck equation:

Definition 2.7 (Fokker-Planck equation for the transition probability density)
Let $x, z \in \Omega \subset \mathbb{R}, s, t \in [0,\infty[$ with $t > s$. The Fokker-Planck equation (for $\hat{y}$) *at x at time t is defined as follows:*

$$\partial_t \hat{y}(x,t;z,s) - \frac{1}{2}\partial_{xx}^2\Big(\sigma(x,t)^2 \hat{y}(x,t;z,s)\Big) + \partial_x\Big(b(x,t;u(x,t))\hat{y}(x,t;z,s)\Big) = 0 \tag{2.4}$$

$$\hat{y}(x,s;z,s) = \delta(x-z) \tag{2.5}$$

where $b, \sigma \colon \mathbb{R} \times \mathbb{R}_{\geq 0} \to \mathbb{R}$ are functions associated with the considered stochastic process in (2.3)*, $u \colon \mathbb{R} \times \mathbb{R}_{\geq 0} \to \mathbb{R}$ is the given control function,* (2.5) *is the initial condition and $\hat{y}$ is the unknown.*

For a given initial PDF of the process at time s, $\rho(z,s)$, the probability density of the process at x at time $t > s$ is given by

$$y(x,t) = \int_\Omega \hat{y}(x,t;z,s)\rho(z,s)\,dz. \tag{2.6}$$

Of course ρ must be a non-negative function and be normalized to one after integration over Ω.

Solving equations (2.4)-(2.5), one obtains the PDF y of the stochastic process for every initial distribution ρ using (2.6). However, this is as complex as to compute a Green function of a PDE. By assuming that the initial condition ρ is given, (2.4) and (2.5) can be reformulated into a Fokker-Planck equation that is much easier to solve, yielding the PDF y only for a certain initial distribution ρ. Additionally, the focus of this work is on localized systems, see above. Ultimately, the problem can therefore be reformulated as the following initial boundary value problem (IBVP).

Definition 2.8 (Fokker-Planck Initial Boundary Value Problem for the PDF)
Let $\Omega \subset \mathbb{R}$ be a bounded and open interval. Further, let $T_E > 0, Q := \Omega\times]0,T_E[$ and $\Sigma := \partial\Omega\times]0,T_E[$. The Fokker-Planck Initial Boundary Value Problem (for the PDF) *is defined as follows:*

$$\partial_t y(x,t) - \frac{1}{2}\partial_{xx}^2\left(\sigma(x,t)^2 y(x,t)\right) + \partial_x\Big(b\big(x,t;u(x,t)\big)y(x,t)\Big) = 0 \qquad \text{in } Q \tag{2.7}$$

$$y(x,0) = \rho(x) \qquad \text{in } \Omega \tag{2.8}$$

$$y(x,t) = 0 \qquad \text{in } \Sigma \tag{2.9}$$

where $b,\sigma\colon Q \to \mathbb{R}$ are given functions associated with the considered stochastic process in (2.3)*, $u\colon Q \to \mathbb{R}$ is the given control function, $\rho\colon \Omega \to \mathbb{R}$ is the initial distribution and $y\colon \bar{Q} \to \mathbb{R}$ is the unknown.*

Remark 2.9
(a) (2.7) *is a linear PDE of parabolic type.*
(b) If the drift term b vanishes, the FP equation (2.7) *becomes the heat equation. If σ vanishes,* (2.7) *becomes the transport equation.*
(c) If $u(x,t) \equiv c \in \mathbb{R}$ then the control is interpreted as a given parameter c.
(d) The problem (2.7)-(2.9) *differs from a classical parabolic problem since the solution y needs to be non-negative and to comply with conservation laws, i.e.*

$$y(x,t) \geq 0,\ (x,t) \in Q, \qquad \int_\Omega y(x,t)\,dx = 1,\ t \geq 0. \tag{2.10}$$

Next, the existence and uniqueness of solutions of the IBVP (2.7)-(2.9) is addressed.

2.2.1 Existence and Uniqueness of Solutions

This section requires some knowledge of PDE theory. In particular, readers should be familiar with weak solutions, Sobolev spaces and the $W(0,T_E)$ space. These preliminaries are established in Appendix A.
The aim is to prove existence and uniqueness of solutions of the Fokker-Planck IBVP (2.7)-(2.9). Since classical solutions may not exist, the focus is on weak solutions in the $W(0,T_E;H,V)$ space with $V := H_0^1(\Omega)$ and $H := L^2(\Omega)$. In the following, $W(0,T_E;H,V)$ will be denoted by $W(0,T_E)$, since neither V nor H will change.

It is helpful to write (2.7) in divergence form, i.e.

$$0 = \partial_t y(x,t) - \partial_x\big(\alpha(x,t)\ \partial_x y(x,t)\big) + \beta(x,t;u(x,t))\ \partial_x y(x,t) + \gamma(x,t;u(x,t))\ y(x,t) \quad (2.11)$$

for suitable coefficient functions α, β, γ. To this end, re-write (2.7) as follows:

$$\begin{aligned}
0 &= \partial_t y(x,t) - \frac{1}{2}\partial_{xx}^2\left(\sigma(x,t)^2 y(x,t)\right) + \partial_x\Big(b\big(x,t;u(x,t)\big)y(x,t)\Big) \\
&= \partial_t y - \frac{1}{2}\partial_x\left(2\sigma\ \partial_x\sigma\ y + \sigma^2\partial_x y\right) + \partial_x b\ y + b\ \partial_x y \\
&= \partial_t y - \partial_x\big(\sigma\ \partial_x\sigma\ y\big) - \partial_x\Big(\underbrace{\frac{1}{2}\sigma^2}_{=:\alpha(x,t)}\ \partial_x y\Big) + \big(D_1(b) + D_3(b)\ \partial_x u\big)y + b\ \partial_x y \\
&= \partial_t y - \left((\partial_x\sigma)^2 + \sigma\ \partial_{xx}^2\sigma\right) y - \sigma\ \partial_x\sigma\ \partial_x y - \partial_x\left(\alpha\ \partial_x y\right) \\
&+ \big(D_1(b) + D_3(b)\ \partial_x u\big)y + b\ \partial_x y \\
&= \partial_t y - \partial_x\left(\alpha\ \partial_x y\right) + \underbrace{\left(b - \sigma\ \partial_x\sigma\right)}_{=:\beta(x,t;u(x,t))}\partial_x y + \underbrace{\left(D_1(b) + D_3(b)\ \partial_x u - (\partial_x\sigma)^2 - \sigma\ \partial_{xx}^2\sigma\right)}_{=:\gamma(x,t;u(x,t))} y.
\end{aligned} \quad (2.12)$$

Example 2.10
For σ, b given by the stochastic processes in Example 2.6, the coefficient functions are:
(a) Ornstein-Uhlenbeck Process with Additive Control:

$$\begin{aligned}
\alpha(x,t) &:= \frac{1}{2}\tilde{\sigma}^2 \\
\beta(x,t;u(x,t)) &:= b(x,t;u(x,t)) = -\mu x + u(x,t) \\
\gamma(x,t;u(x,t)) &:= -\mu + \partial_x u(x,t)
\end{aligned}$$

(b) Geometric-Brownian Process with Additive Drift Control:

$$\begin{aligned}
\alpha(x,t) &:= \frac{1}{2}\tilde{\sigma}^2 x^2 \\
\beta(x,t;u(x,t)) &:= (\mu + u(x,t))x - \tilde{\sigma}^2 x = (\mu + u(x,t) - \tilde{\sigma}^2)x \\
\gamma(x,t;u(x,t)) &:= \mu + u(x,t) + x\ \partial_x u(x,t) - \tilde{\sigma}^2
\end{aligned}$$

(c) Shiryaev Process:

$$\begin{aligned}
\alpha(x,t) &:= \frac{1}{2}\tilde{\sigma}^2 x^2 \\
\beta(x,t;u(x,t)) &:= u(x,t) + \mu x - \tilde{\sigma}^2 x = u(x,t) + (\mu - \tilde{\sigma}^2)x \\
\gamma(x,t;u(x,t)) &:= \mu + \partial_x u(x,t) - \tilde{\sigma}^2
\end{aligned}$$

Note that σ and b are sufficiently regular w.r.t. x in the three stochastic processes discussed here. Therefore, the existence of $\partial_x\sigma, \partial_{xx}^2\sigma$ and $D_1(b)$ is not an issue. However, the existence of $\partial_x u(x,t)$ is a strong requirement. In cases where u does not depend on x, the $\partial_x u$ terms vanish.

Remark 2.11
For $\sigma(x,t) \neq 0$ the Fokker-Planck equation (2.7) is uniformly parabolic[1]*, cf. Definition A.35. In particular, (2.7) is uniformly parabolic in the case of the Ornstein-Uhlenbeck process. Furthermore, if $\Omega \subset]0,\infty[$, (2.7) is uniformly parabolic for the remaining two stochastic processes introduced in Example 2.6.*

The following result guarantees the existence and uniqueness of a weak solution under appropriate assumptions, particularly on the coefficient functions.

Theorem 2.12 (Existence and uniqueness)
Let $\Omega \subset \mathbb{R}$ be open and bounded, $\alpha, \beta, \gamma \in L^\infty(Q)$ as in (2.12), $\rho \in L^2(\Omega)$ and $\sigma(x,t) \neq 0, (x,t) \in Q$. Then the Fokker-Planck IBVP (2.7)-(2.9) has a unique weak solution $y \in W(0,T_E)$ (in the sense of Definition A.37) and satisfies the energy estimate

$$\|y(t)\|^2_{L^2(\Omega)} + \|y\|^2_{L^2(0,T_E;H_0^1(\Omega))} + \|y'\|^2_{L^2(0,T_E;H_0^1(\Omega)^*)} \leq C \|\rho\|^2_{L^2(\Omega)}, \quad t \in]0,T_E[,$$

where $C > 0$ depends only on $C_1 > 0, C_2 \geq 0$ given by

$$a(v,v;t) + C_2 \|v\|^2_{L^2(\Omega)} \geq C_1 \|v\|^2_{H_0^1(\Omega)}, \quad v \in H_0^1(\Omega) \text{ and almost all } t \in]0,T_E[$$

with $a(\cdot,\cdot,t)\colon H_0^1(\Omega) \times H_0^1(\Omega) \to \mathbb{R}$ being a bilinear form for almost all $t \in]0,T_E[$ given by

$$a(v_1,v_2,t) := \int_\Omega \alpha(t)\ \partial_x v_1\ \partial_x v_2 + \beta(t)\ \partial_x v_1\ v_2 + \gamma(t)\ v_1\ v_2\, dx,$$

where the dependence on x has been omitted.

Proof. This Theorem is a special case of [16, p. 46, Corollary 1.1]. □

Remark 2.13
Note that the initial condition (2.8) makes sense, since $W(0,T_E)$ is continuously embedded in $C([0,T_E],L^2(\Omega))$, cf. Theorem A.32.

The regularity of the solution y can be further improved under appropriate conditions:

Theorem 2.14 (Improved regularity)
(a) Let $\rho \in H_0^1(\Omega)$ and suppose $y \in W(0,T_E)$ is a solution of the Fokker-Planck IBVP (2.7)-(2.9). Then

$$y \in L^2(0,T_E;H^2(\Omega)) \cap L^\infty(0,T_E;H_0^1(\Omega)),\ y' \in L^2(0,T_E;L^2(\Omega))$$

with the estimate

$$\operatorname*{ess\,sup}_{t\in[0,T_E]} \|y(t)\|_{H_0^1(\Omega)} + \|y\|_{L^2(0,T_E;H^2(\Omega))} + \|y'\|_{L^2(0,T_E;L^2(\Omega))} \leq C \|\rho\|_{H_0^1(\Omega)}$$

and $C > 0$ depending only on Ω, T_E, and the coefficient functions α, β, and γ.

[1] More precisely, the operator $\frac{\partial}{\partial t} + L$ in Definition A.35 is uniformly parabolic.

(b) If, in addition, $\rho \in H^2(\Omega)$, then

$$\begin{aligned} y &\in L^\infty(0,T_E;H^2(\Omega)),\\ y' &\in L^\infty(0,T_E;L^2(\Omega)) \cap L^2(0,T_E;H_0^1(\Omega)),\\ y'' &\in L^2(0,T_E;H_0^1(\Omega)^*) \end{aligned}$$

with the estimate

$$\operatorname*{ess\,sup}_{t\in[0,T_E]} \Big(\|y(t)\|_{H^2(\Omega)} + \|y'(t)\|_{L^2(\Omega)} \Big) + \|y'\|_{L^2(0,T_E;H_0^1(\Omega))} + \|y''\|_{L^2(0,T_E;H_0^1(\Omega)^*)} \le C\,\|\rho\|_{H^2(\Omega)}\,.$$

Proof. This is a special case of [7, p. 382, Theorem 5]. □

Until now, nothing has been said about the positivity and conservativeness of the weak solution of (2.7)-(2.9). The latter property results from the fact that the FP equation can be written in flux form [4]. The former property requires some assumptions on σ and b, which were introduced in [5] to prove nonnegativity of solutions of linear parabolic equations. They differ from the assumptions in Theorem 2.12, partly because of the slightly different notion of a weak solution and the consideration of linear parabolic PDEs of the form

$$\begin{aligned} 0 = \partial_t y(x,t) &- \partial_x\big(\alpha(x,t)\ \partial_x y(x,t) + \tilde\alpha(x,t)\ y(x,t)\big)\\ &- \tilde\beta(x,t;u(x,t))\ \partial_x y(x,t) - \tilde\gamma(x,t;u(x,t))\ y(x,t) \end{aligned}$$

instead of the divergence form (2.11). In this setting, the coefficient functions are given by

$$\begin{aligned} \alpha(x,t) &:= \frac{1}{2}\sigma(x,t)^2,\\ \tilde\alpha(x,t) &:= \sigma(x,t)\ \partial_x\sigma(x,t),\\ \tilde\beta(x,t;u(x,t)) &:= -b(x,t;u(x,t)),\\ \tilde\gamma(x,t;u(x,t)) &:= -\big(D_1(b)(x,t;u(x,t)) + D_3(b)(x,t;u(x,t))\ \partial_x u(x,t)\big). \end{aligned} \tag{2.13}$$

Remark 2.15
If u does not depend on x, then $\tilde\gamma(x,t;u(x,t)) = \partial_x b(x,t;u(t))$.

To satisfy the assumptions on the coefficient functions $\alpha, \tilde\alpha, \tilde\beta$ and $\tilde\gamma$ made in [5] and adopted in [4], it is sufficient to impose the following requirements.

Assumption 2.16
(i) $\forall \xi \in \mathbb{R}$ *and almost all* $(x,t) \in Q$:
- $\alpha(x,t)\xi^2 \ge \theta\xi^2$ *for some constant* $0 < \theta < \infty$,
- $|\alpha(x,t)| \le M$ *for some constant* $0 < M < \infty$.

(ii) $\tilde\alpha, \tilde\beta \in L^q(0,T_E;L^p(\Omega))$ *with* $2 < p,q \le \infty$ *and* $\frac{1}{2p} + \frac{1}{q} < \frac{1}{2}$.

(iii) $\tilde\gamma \in L^q(0,T_E;L^p(\Omega))$ *with* $1 < p,q \le \infty$ *and* $\frac{1}{2p} + \frac{1}{q} < 1$.

Remark 2.17
For bounded $\alpha(x,t)$ defining a uniformly parabolic operator (cf. Definition A.35) and $\tilde{\alpha}, \tilde{\beta}, \tilde{\gamma} \in L^\infty(Q)$, Assumption 2.16 is immediately satisfied.

With these assumptions, it is possible to prove the following result.

Theorem 2.18 (Existence of non-negative solutions)
Suppose $\alpha, \tilde{\alpha}, \tilde{\beta}$ and $\tilde{\gamma}$ in (2.13) satisfy Assumption 2.16. Let $\rho \in L^2(\Omega)$.
$\Rightarrow \exists_1 y \in L^2(0,T_E;H_0^1(\Omega)) \cap L^\infty(0,T_E;L^2(\Omega))$:

$$\int_0^{T_E}\int_\Omega -y\ \partial_t\varphi + \alpha\ \partial_x y\ \partial_x\varphi + \tilde{\alpha}\ y\ \partial_x\varphi - \tilde{\beta}\ \partial_x y\ \varphi - \tilde{\gamma}\ y\ \varphi\,dx\,dt = 0, \quad \varphi \in C_0^1(Q), \tag{2.14}$$

$$\lim_{t\to 0}\int_\Omega y(\cdot,t)\ \varphi\,dx = \int_\Omega \rho\ \varphi\,dx, \quad \varphi \in C_0^1(\Omega). \tag{2.15}$$

Moreover, if $0 \le \rho \le m$ almost everywhere in Ω, then

$$0 \le y(x,t) \le m(1+Ck) \text{ in } Q,$$

where

$$k := \left(\int_0^{T_E}\left(\int_\Omega |\tilde{\alpha}(x,t)|^p\,dx\right)^{q/p} dt\right)^{1/q} + \left(\int_0^{T_E}\left(\int_\Omega |\tilde{\gamma}(x,t)|^p\,dx\right)^{q/p} dt\right)^{1/q},$$

and the constant $C > 0$ depends only on T_E, Ω, and the structure of the Fokker-Planck operator, i.e. on the coefficient functions $\alpha, \tilde{\alpha}, \tilde{\beta}$, and $\tilde{\gamma}$.

Proof. This is a special case of [5, p. 634, Theorem 1]. □

Since the setting in Theorem 2.18 differs from the one in Theorems 2.12 and 2.14, the question remains whether the unique weak solution $y \in W(0,T_E)$ – which exists according to Theorem 2.12 – is indeed non-negative. To this end, the three above theorems are brought together.

Corollary 2.19
Let $\Omega \subset \mathbb{R}$ be open and bounded. Furthermore, let

$$\sigma^2, b, \sigma\ \partial_x\sigma, D_1(b) + D_3(b)\ \partial_x u, (\partial_x\sigma)^2 + \sigma\ \partial_{xx}^2\sigma \in L^\infty(Q) \tag{2.16}$$

and

$$\sigma(x,t) \neq 0, \quad (x,t) \in Q. \tag{2.17}$$

Additionally, let $\rho \in H_0^1(\Omega)$ be a probability density function, i.e.

$$\rho(x) \ge 0 \text{ in } \Omega \quad \text{and} \quad \int_\Omega \rho(x)\,dx = 1.$$

Then the Fokker-Planck IBVP (2.7)-(2.9) *has a unique non-negative weak solution* y *in the sense of Definition A.37 with*

$$\begin{aligned} &y \in L^2(0,T_E;H^2(\Omega)) \cap L^\infty(0,T_E;H_0^1(\Omega)),\\ &y' \in L^2(0,T_E;L^2(\Omega)). \end{aligned} \tag{2.18}$$

In particular, $y \in C([0,T_E];L^2(\Omega))$*. Moreover, the following estimate holds:*

$$\operatorname*{ess\,sup}_{t\in[0,T_E]} \|y(t)\|_{H_0^1(\Omega)} + \|y\|_{L^2(0,T_E;H^2(\Omega))} + \|y'\|_{L^2(0,T_E;L^2(\Omega))} \le C\,\|\rho\|_{H_0^1(\Omega)} \tag{2.19}$$

with $C>0$ *depending only on* Ω*,* T_E*, and the coefficient functions* α*,* β*, and* γ *defined as in* (2.12)*.*

Proof. From (2.16) it follows that the coefficient functions α,β, and γ defined in (2.12) are $L^\infty(Q)$ functions. With (2.17) and $\rho \in H_0^1(\Omega)$ the remaining conditions of Theorem 2.12 are satisfied as well. Theorem 2.12 now ensures the existence and uniqueness of a weak solution $y \in W(0,T_E)$ (in the sense of Definition A.37). Applying Theorem A.32 results in $y \in C([0,T_E];L^2(\Omega))$. Moreover, since $\rho \in H_0^1(\Omega)$, Theorem 2.14 guarantees (2.18) and the estimate (2.19).
Additionally, from (2.16) it follows that the coefficient functions α, $\tilde{\alpha}$, $\tilde{\beta}$, and $\tilde{\gamma}$ defined in (2.13) are $L^\infty(Q)$ functions and thus satisfy Assumption 2.16. Note that because of (2.17), Assumption 2.16(i) holds in particular.
Since all conditions of Theorem 2.18 hold, it remains to show that the unique weak solution y obtained from Theorem 2.12 satisfies (2.14)-(2.15). Then from the uniqueness property in Theorem 2.18 and (2.18), the non-negativity of the unique weak solution in the sense of Definition A.37 follows. To this end, the equivalent formulation of a weak solution of a parabolic PDE is used, cf. Definition A.38. According to Definition A.38, the following holds:

$$\begin{aligned} 0 &= \int_0^{T_E}\int_\Omega \partial_t y\; v\,dx\,dt + \int_0^{T_E}\int_\Omega \alpha\;\partial_x y\;\partial_x v + \beta\;\partial_x y\;v + \gamma\;y\;v\,dx\,dt\\ &= \int_0^{T_E}\int_\Omega \partial_t y\; v\,dx\,dt + \int_0^{T_E}\int_\Omega \alpha\;\partial_x y\;\partial_x v\,dx\,dt + \int_0^{T_E}\int_\Omega b\;\partial_x y\;v\,dx\,dt - \int_0^{T_E}\int_\Omega \sigma\;\partial_x\sigma\;\partial_x y\;v\,dx\,dt\\ &\quad + \int_0^{T_E}\int_\Omega \left(D_1(b) + D_3(b)\;\partial_x u\right) y\;v\,dx\,dt - \int_0^{T_E}\int_\Omega \left((\partial_x\sigma)^2 + \sigma\;\partial_{xx}^2\sigma\right) y\;v\,dx\,dt \end{aligned}$$

for all $v \in L^2(0,T_E;H_0^1(\Omega))$ and therefore, for all $v \in C_0^1(Q)$ in particular. For $v \in C_0^1(Q)$, the formula for integration by parts is used for the first (w.r.t. t) and the fourth integral

(w.r.t. x), resulting in

$$
\begin{aligned}
0 &= \int_0^{T_E}\int_\Omega \partial_t y\ v\,dx\,dt + \int_0^{T_E}\int_\Omega \alpha\ \partial_x y\ \partial_x v\,dx\,dt + \int_0^{T_E}\int_\Omega \underbrace{b}_{=-\tilde{\beta}}\ \partial_x y\ v\,dx\,dt \\
&- \int_0^{T_E}\int_\Omega \sigma\ \partial_x\sigma\ v\ \partial_x y\,dx\,dt + \int_0^{T_E}\int_\Omega \underbrace{(D_1(b) + D_3(b)\ \partial_x u)}_{=-\tilde{\gamma}}\ y\ v\,dx\,dt \\
&- \int_0^{T_E}\int_\Omega \left((\partial_x\sigma)^2 + \sigma\ \partial^2_{xx}\sigma\right) y\ v\,dx\,dt \\
&= \int_0^{T_E}\int_\Omega -y\ \partial_t v\,dx\,dt + \int_0^{T_E}\int_\Omega \alpha\ \partial_x y\ \partial_x v\,dx\,dt - \int_0^{T_E}\int_\Omega \tilde{\beta}\ \partial_x y\ v\,dx\,dt \\
&+ \int_0^{T_E}\int_\Omega \underbrace{\partial_x\left(\sigma\ \partial_x\sigma\ v\right)}_{=\left((\partial_x\sigma)^2+\sigma\ \partial^2_{xx}\sigma\right)v+\sigma\ \partial_x\sigma\ \partial_x v}\ y\,dx\,dt - \int_0^{T_E}\int_\Omega \tilde{\gamma}\ y\ v\,dx\,dt \\
&- \int_0^{T_E}\int_\Omega \left((\partial_x\sigma)^2 + \sigma\ \partial^2_{xx}\sigma\right) y\ v\,dx\,dt \\
&= \int_0^{T_E}\int_\Omega -y\ \partial_t v + \alpha\ \partial_x y\ \partial_x v - \tilde{\beta}\ \partial_x y\ v + \underbrace{\sigma\ \partial_x\sigma}_{=\tilde{\alpha}}\ \partial_x v\ y - \tilde{\gamma}\ y\ v\,dx\,dt,
\end{aligned}
$$

which is exactly (2.14). Notice that no additional terms appear since $v \in C_0^1(Q)$. Obviously, y satisfies condition (2.15), concluding the proof. □

Remark 2.20
If the control u does not depend on x, no $\partial_x u$ terms appear. This makes the requirements in Corollary 2.19 much easier to cope with, especially in light of optimal control. Often the control is assumed to be less regular, e.g. a piecewise constant function w.r.t. time t or $u \in L^2(Q)$.

For every fixed control function u, Corollary 2.19 guarantees a unique non-negative weak solution under certain regularity assumptions. It is therefore meaningful to formulate optimal control problems. This is done in the next section.

2.3 The Fokker-Planck Optimal Control Problem

Instead of fixing the control, the task now is to find a control u such that, starting with a certain initial probability density function ρ, the stochastic process evolves such that a

sequence of desired target probability densities is matched as good as possible at given terminal times within the time interval $]0, T_E[$, cf. Definition 2.8. Consider terminal times of the form nT_s with $T_s > 0$ and $n \in \mathbb{N}$ and partition $]0, T_E[$ into M pieces of equal length T_s, i.e. $]t_n, t_n + T_s[$ with $t_n := nT_s, n = 0, ..., M-1$.
At first, only one such interval is looked at at a time. To formulate the optimal control problem, define $Q := \Omega \times]t_n, t_n + T_s[$. Analogously, using the same notation as in Definition 2.8, $\Sigma := \partial\Omega \times]t_n, t_n + T_s[$. For $y_d \in L^2(Q)$ denoting the desired probability density function the first optimal control problem is given by:

$$\min_{u \in U} J(y,u) := \frac{1}{2} \left\| y(\cdot, t_n + T_s) - y_d(\cdot, t_n + T_s) \right\|^2_{L^2(\Omega)} + \frac{\lambda}{2}|u|^2, \tag{2.20}$$

with a so-called regularization parameter $\lambda > 0$, subject to

$$\begin{aligned} \partial_t y(x,t) - \frac{1}{2}\partial^2_{xx}\left(\sigma(x,t)^2 y(x,t)\right) + \partial_x\Big(b\big(x,t;u\big)y(x,t)\Big) &= 0 \qquad \text{in } Q \\ y(x,t_n) &= \rho(x) \quad \text{in } \Omega \\ y(x,t) &= 0 \qquad \text{in } \Sigma \end{aligned} \tag{2.21}$$

and

$$u \in U_{ad} := \{ u \in U \mid u_a \leq u \leq u_b \} \subset U := \mathbb{R}, \tag{2.22}$$

where

$$u_a, u_b \in \mathbb{R}. \tag{2.23}$$

In this case, the control u is simply a real value. The optimal control problem (2.20)-(2.23) is solved on $]0, T_s[$, i.e. for $n = 0$. Then the next interval is looked at, i.e. n is increased and the OCP is solved on $]t_1, t_1 + T_s[=]T_s, 2T_s[$, and so on. This is exactly the problem considered in [3]. The question is whether the numerical results in [3] can be improved by considering $N \in \{2, ..., M\}$ consecutive intervals at a time instead of one.[2] To this end, the time interval is extended to $]t_n, t_n + (N-1)T_s[$. Hence, $Q := \Omega \times]t_n, t_n + (N-1)T_s[$ and $\Sigma := \partial\Omega \times]t_n, t_n + (N-1)T_s[$. The idea is still to track a sequence of desired target probability densities. For sufficiently high M or, equivalently, for sufficiently low T_s, this task is (approximately) captured by the following optimal control problem:

$$\min_{u \in U} J(y,u) := \int_{t_n}^{t_n + (N-1)T_s} \frac{1}{2} \left\| y(\cdot,t) - y_d(\cdot,t) \right\|^2_{L^2(\Omega)} + \frac{\lambda}{2} u(t)^2 \, dt, \tag{2.24}$$

with $\lambda > 0$, subject to

$$\begin{aligned} \partial_t y(x,t) - \frac{1}{2}\partial^2_{xx}\left(\sigma(x,t)^2 y(x,t)\right) + \partial_x\Big(b\big(x,t;u(t)\big)y(x,t)\Big) &= 0 \qquad \text{in } Q \\ y(x,t_n) &= \rho(x) \quad \text{in } \Omega \\ y(x,t) &= 0 \qquad \text{in } \Sigma \end{aligned} \tag{2.25}$$

[2] Why this might lead to better results is explained in Chapter 3.

and

$$u \in U_{ad} \subset U \tag{2.26}$$

with

$$\begin{aligned} U &:= L^2(]t_n, t_n + (N-1)T_s[), \\ U_{ad} &:= \{u \in U \mid u_a \leq u \leq u_b \text{ a.e. in }]t_n, t_n + (N-1)T_s[\}, \end{aligned} \tag{2.27}$$

where

$$u_a, u_b \in L^\infty(]t_n, t_n + (N-1)T_s[). \tag{2.28}$$

The control u is now depending on time, but not on the space x. For a given time t, the same control value is applied on the whole space domain Ω. Intuitively, one expects better results if u is depending on x. Therefore, the next step is to introduce a space-dependent control. For Q and Σ as in (2.20)-(2.23), this leads to

$$\min_{u \in U} J(y,u) := \frac{\alpha}{2} \left\| y(\cdot, t_n + T_s) - y_d(\cdot, t_n + T_s) \right\|^2_{L^2(\Omega)} + \frac{\lambda}{2} \left\| u \right\|^2_{L^2(\Omega)}, \tag{2.29}$$

with $\alpha, \lambda > 0$, subject to

$$\begin{aligned} \partial_t y(x,t) - \frac{1}{2} \partial^2_{xx} \left(\sigma(x,t)^2 y(x,t) \right) + \partial_x \Big(b\big(x,t;u(x)\big) y(x,t) \Big) &= 0 \qquad \text{in } Q \\ y(x,t_n) &= \rho(x) \quad \text{in } \Omega \\ y(x,t) &= 0 \qquad \text{in } \Sigma \end{aligned} \tag{2.30}$$

and

$$u \in U_{ad} \subset U \tag{2.31}$$

with

$$\begin{aligned} U &:= L^2(\Omega), \\ U_{ad} &:= \{u \in U \mid u_a \leq u \leq u_b \text{ almost everywhere in } \Omega\}, \end{aligned} \tag{2.32}$$

where

$$u_a, u_b \in L^\infty(\Omega). \tag{2.33}$$

The additional parameter α is introduced to vary the weight of the state penalty term. The ratio of α and λ is more important than the value of α itself. However, for numerical reasons, it is beneficial not to lower λ too much, cf. Chapter 6. The solution is then to increase α instead.
Eventually, for Q and Σ as in (2.24)-(2.28), the optimal control problem is given by

$$\begin{aligned} \min_{u \in U} J(y,u) &:= \int_{t_n}^{t_n + (N-1)T_s} \frac{\alpha}{2} \left\| y(\cdot,t) - y_d(\cdot,t) \right\|^2_{L^2(\Omega)} + \frac{\lambda}{2} \left\| u(\cdot,t) \right\|^2_{L^2(\Omega)} \, dt \\ &= \frac{\alpha}{2} \left\| y(\cdot,t) - y_d(\cdot,t) \right\|^2_{L^2(Q)} + \frac{\lambda}{2} \left\| u(\cdot,t) \right\|^2_{L^2(Q)}, \end{aligned} \tag{2.34}$$

with $\alpha, \lambda > 0$, subject to

$$\begin{aligned} \partial_t y(x,t) - \frac{1}{2}\partial_{xx}^2\left(\sigma(x,t)^2 y(x,t)\right) + \partial_x\Big(b\big(x,t;u(x,t)\big)y(x,t)\Big) &= 0 \qquad \text{in } Q \\ y(x,t_n) &= \rho(x) \quad \text{in } \Omega \\ y(x,t) &= 0 \qquad \text{in } \Sigma \end{aligned} \tag{2.35}$$

and

$$u \in U_{ad} \subset U \tag{2.36}$$

with

$$\begin{aligned} U &:= L^2(Q), \\ U_{ad} &:= \{u \in U \mid u_a \leq u \leq u_b \text{ almost everywhere in } Q\}, \end{aligned} \tag{2.37}$$

where

$$u_a, u_b \in L^\infty(Q). \tag{2.38}$$

Remark 2.21
In (2.29)-(2.33) *and* (2.34)-(2.38)*, the set of admissible controls* U_{ad} *consists of* L^2 *functions. However, in order to prove that a solution of the Fokker-Planck equation exists, additional regularity of* u *was needed, cf. Subsection 2.2.1. In particular, the existence of* $\partial_x u$ *was required. Ultimately,* u *has to meet the most restrictive demand.*

It is questionable whether each of the optimal control problems (2.20)-(2.23), (2.24)-(2.28), (2.29)-(2.33), and (2.34)-(2.38) possesses a (unique) optimal control u^*. Even if u^* exists, the next question is how to obtain it. These two problems are addressed in the two following subsections.

2.3.1 Existence and Uniqueness of Optimal Solutions

One possible way to prove existence and uniqueness of optimal solutions is illustrated in [30]: After the existence and uniqueness of solutions of the PDE have been established, the state y in the cost functional can be eliminated formally by replacing it with a solution operator S, i.e. $y = S(u)$. This "new" cost functional is called the reduced cost functional. Then, for quadratic cost functionals like (2.34) and linear and continuous (i.e. linear and bounded) S, one can prove that – under certain technical assumptions[3] – an optimal solution exists. Moreover, it is unique if $\lambda > 0$. The linearity and continuity of S and the convexity of the reduced cost functional[4] are crucial.
However, for the optimal control problems (2.20)-(2.23), (2.24)-(2.28), (2.29)-(2.33), and (2.34)-(2.38), clearly the mapping $u \mapsto S(u)$ is not linear and the (reduced) cost functional is non-convex. Therefore, the above statement does not necessarily hold. A remedy is

[3]These assumptions are not relevant here as they are not critical; they are satisfied for the Fokker-Planck optimal control problems introduced above.

[4]The convexity property holds for quadratic cost functionals like (2.34) and linear and continuous S.

provided in [16, Section 1.5], where the theorem about the existence of optimal solutions is extended to (general) nonlinear problems

$$\min_{(y,u)\in Y\times U} J(y,u) \tag{2.39}$$

subject to

$$\begin{aligned} e(y,u) &= 0, \\ u &\in U_{ad}, \\ y &\in Y_{ad}, \end{aligned} \tag{2.40}$$

where $J\colon Y \times U \to \mathbb{R}$ and $e\colon Y \times U \to Z$ are continuous, Z is a Banach space and U, Y are reflexive Banach spaces (e.g. real Hilbert spaces). Then, under the following assumptions, an optimal solution of (2.39)-(2.40) exists, cf. [16, p. 55, Theorem 1.45]:

Assumption 2.22

(i) $\emptyset \neq U_{ad} \subset U$ *is convex, bounded and closed.*
(ii) $Y_{ad} \subset Y$ *is convex and closed, such that* (2.39)-(2.40) *has a feasible point.*
(iii) The state equation (i.e. the PDE) $e(y,u) = 0$ *has a bounded solution operator* $U_{ad} \ni u \mapsto S(u) \in Y$.
(iv) $Y \times U \ni (y,u) \mapsto e(y,u) \in Z$ *is continuous under weak convergence.*
(v) J *is sequentially weakly lower continuous.*

Remark 2.23

(a) Since U_{ad} *is not necessarily compact ("bounded" and "closed" does not imply "compact" in infinite dimensional spaces like* $L^2(Q)$*), a detour involving the notion of weak convergence must be taken. For more details, see e.g. [16] or [30].*
(b) Out of the five assumptions in Assumption 2.22, (iv) and (v) are of particular importance in order to incorporate the more general setting.
(c) To facilitate the verification of Assumption 2.22(iv), often compact embeddings $Y \hookrightarrow\hookrightarrow \tilde{Y}$ *are utilized to convert weak convergence in* Y *to strong convergence in* $\tilde{Y}$ *[16].*

In [4, Section 3] existence and uniqueness of optimal solutions of a Fokker-Planck OCP similar to (2.20)-(2.23) have been discussed. In the case of the Ornstein-Uhlenbeck process, cf. Example 2.6, (2.20)-(2.23) is included in the considered problem. In all other cases considered in this thesis, an existence result regarding optimal control is yet to be provided. To this end, one could try to verify Assumption 2.22. However, this task is not tackled in this work and remains an open question. In the following, the existence of a unique optimal solution is assumed whenever needed.

2.3.2 Necessary Optimality Conditions

As in finite-dimensional optimization, an optimal control is obtained by deriving and then verifying necessary (and sufficient) optimality conditions. To this end, in the case of optimal control problems in infinite-dimensional spaces, differentiability in Banach spaces is needed.

This leads to the notion of Gâteaux- and Fréchet differentiability, cf. [16], [30]. The reduced cost functional is then differentiated w.r.t. u, yielding the so-called variational inequality. This necessary condition is subsequently simplified by introducing the adjoint state p, which is characterized by a terminal boundary value problem that is connected to the IBVP in the side conditions of the optimal control problem. The adjoint state p can be interpreted as the Lagrangian multiplier of the original IBVP. If the reduced cost functional is convex, the necessary condition is sufficient and characterizes a global minimum.

This course of action has been taken in [4] to derive an optimality system that, after solving, yields a local minimum. The (non-convex) optimal control problem considered in the theoretical part of that paper is similar to (2.20)-(2.23). It includes (2.20)-(2.23) in case of the Ornstein-Uhlenbeck process, cf. Example 2.6.

Another possible way to determine the necessary conditions rigorously is to employ the Karush-Kuhn-Tucker theory for optimization in Banach spaces. The problem here is to coordinate operators, functionals and suitable function spaces: Some differentiability of the given variables is needed in certain function spaces, adjoint operators need to be determined and the Lagrangian multipliers must exist in some specified spaces. For more details, see e.g. [30, Chapter 6] and the references therein.

Keeping this in mind, if one's intention is not to prove the necessary conditions but to quickly get an impression of how they and the adjoint equation in particular look like, the formal Lagrange technique introduced in [30] is well-suited. It is based on the Karush-Kuhn-Tucker theory, but ignoring any technical difficulties like heeding the different function spaces. Instead, it is assumed that the state y, the Lagrangian multipliers, and all occurring derivatives are L^2 functions. The idea is to take the cost functional and incorporate the "complicated" side conditions, i.e. the PDE and possibly the boundary conditions, into the Lagrange function and use the Lagrange principle to determine the necessary conditions. Note that neither does this technique prove the existence of an optimal solution, nor does it prove that the conditions derived with this technique are indeed necessary. Still, it provides a starting point for a rigorous derivation.

In the following, this technique is employed for the four optimal control problems introduced above. As already mentioned in Subsection 2.3.1, these problems are non-convex. Because of this, one is content with a local minimum. Note further that, even if the conditions derived using this technique are correct, they are only necessary. Yet, numerical results look promising and indicate that these conditions are indeed appropriate to find a local minimum, cf. Chapter 6.

Example 2.24
Consider the optimal control problem (2.20)-(2.23) *with* $\Omega :=]x_L, x_R[$. *Define the Lagrange*

function by

$$\begin{aligned}\mathcal{L}(y,u,p) &= J(y,u) - \iint\limits_Q \left(\partial_t y - \frac{1}{2}\partial_{xx}^2\left(\sigma^2 y\right) + \partial_x\big(b\ y\big)\right) p\,dx\,dt \\ &= \frac{1}{2}\int\limits_\Omega \big(y(\cdot,t_n+T_s) - y_d(\cdot,t_n+T_s)\big)^2\,dx + \frac{\lambda}{2}\underbrace{|u|^2}_{=u^2} \\ &\quad - \iint\limits_Q \left(\partial_t y - \frac{1}{2}\partial_{xx}^2\left(\sigma^2 y\right) + \partial_x\big(b\ y\big)\right) p\,dx\,dt\end{aligned}$$

Since the condition $u \in U_{ad}$ is not included in the Lagrange function, the variational inequality

$$D_u\mathcal{L}(\bar{y},\bar{u},p)(u-\bar{u}) \geq 0, \quad u \in U_{ad} \tag{2.41}$$

follows from the Lagrange principle, where $\bar{u}$ and $\bar{y}$ denote the optimal control and the corresponding state, respectively. Due to the initial and boundary conditions $y(\cdot,t_n) = \rho$ and $y_{|\Sigma} = 0$, which too have not been incorporated into the Lagrange function, it follows analogously that

$$D_y\mathcal{L}(\bar{y},\bar{u},p)(y-\bar{y}) \geq 0$$

for all y with $y(\cdot,t_n) = \rho$ and $y_{|\Sigma} = 0$. Substituting $y := y - \bar{y}$ yields $D_y\mathcal{L}(\bar{y},\bar{u},p)y \geq 0$ for all y with $y(\cdot,t_n) = 0$ and $y_{|\Sigma} = 0$. Another substitution $y := -y$ then finally results in

$$D_y\mathcal{L}(\bar{y},\bar{u},p)y = 0, \quad y \text{ with } y(\cdot,t_n) = 0 \text{ and } y_{|\Sigma} = 0. \tag{2.42}$$

In a next step, $D_u\mathcal{L}(\bar{y},\bar{u},p)(u-\bar{u})$ and $D_y\mathcal{L}(\bar{y},\bar{u},p)y$ are computed.

$$D_u\mathcal{L}(\bar{y},\bar{u},p)(u-\bar{u}) = \left(\lambda\bar{u} - \iint\limits_Q \partial_x\left(\frac{\partial b}{\partial u}\ \bar{y}\right) p\,dx\,dt\right)(u-\bar{u}) \geq 0, \quad u \in U_{ad}.$$

The Lagrangian multiplier p has not been characterized yet. (2.42) helps in that matter:

$$\begin{aligned}D_y\mathcal{L}(\bar{y},\bar{u},p)\ y &= \int\limits_\Omega \big(\bar{y}(\cdot,t_n+T_s) - y_d(\cdot,t_n+T_s)\big)\ y(t_n+T_s)\,dx \\ &\quad - \iint\limits_Q \left(\partial_t y - \frac{1}{2}\partial_{xx}^2\left(\sigma^2 y\right) + \partial_x\big(b\ y\big)\right) p\,dx\,dt.\end{aligned} \tag{2.43}$$

Note that the derivative of a linear continuous mapping $y \mapsto y(\cdot,t_n+T_s)$ is the mapping itself. Plugging in (2.43) in (2.42) and integrating by parts w.r.t. t in $\partial_t y$ and w.r.t. x in

$\partial_{xx}^2(\sigma^2 y)$ (twice) and $\partial_x(b\ y)$ yields

$$
\begin{aligned}
0 &= \int_\Omega \big(\bar{y}(\cdot, t_n+T_s) - y_d(\cdot, t_n+T_s)\big)\ y(\cdot, t_n+T_s)\, dx \\
&+ \iint_Q \partial_t p\ y\, dx\, dt - \int_\Omega (y\ p)\big|_{t_n}^{t_n+T_s}\ dx \\
&+ \iint_Q \frac{1}{2}\sigma^2 y\ \partial_{xx}^2 p\, dx\, dt + \int_{t_n}^{t_n+T_s} \left(\frac{1}{2} p\ \partial_x\left(\sigma^2 y\right)\right)\Big|_{x_L}^{x_R} dt - \underbrace{\int_{t_n}^{t_n+T_s} \left(\frac{1}{2}\sigma^2 y\ \partial_x p\right)\Big|_{x_L}^{x_R} dt}_{=0} \\
&+ \iint_Q b\ y\ \partial_x p\, dx\, dt - \int_{t_n}^{t_n+T_s} \underbrace{(p\ b\ y)\big|_{x_L}^{x_R}}_{=0} dt \\
&= \int_\Omega \big(\bar{y}(\cdot, t_n+T_s) - y_d(\cdot, t_n+T_s)\big)\ y(\cdot, t_n+T_s) - y(\cdot, t_n+T_s)\ p(\cdot, t_n+T_s)\, dx \\
&- \iint_Q -\partial_t p\ y - \frac{1}{2}\sigma^2 y\ \partial_{xx}^2 p - b\ y\ \partial_x p\, dx\, dt + \int_{t_n}^{t_n+T_s} \left(\frac{1}{2} p\ \partial_x\left(\sigma^2 y\right)\right)\Big|_{x_L}^{x_R} dt \\
&= \int_\Omega \big(\bar{y}(\cdot, t_n+T_s) - y_d(\cdot, t_n+T_s) - p(\cdot, t_n+T_s)\big)\ y(\cdot, t_n+T_s)\, dx \\
&- \iint_Q \left(-\partial_t p - \frac{1}{2}\sigma^2 \partial_{xx}^2 p - b\ \partial_x p\right) y\, dx\, dt + \int_{t_n}^{t_n+T_s} \left(\frac{1}{2} p\ \partial_x\left(\sigma^2 y\right)\right)\Big|_{x_L}^{x_R} dt
\end{aligned}
\tag{2.44}
$$

for all sufficiently smooth y with $y(\cdot, t_n) = 0$ and $y_{|\Sigma} = 0$. Choosing $y \in C_0^\infty(Q)$ results in

$$
0 = -\iint_Q \left(-\partial_t p - \frac{1}{2}\sigma^2 \partial_{xx}^2 p - b\ \partial_x p\right) y\, dx\, dt, \quad y \in C_0^\infty(Q).
$$

Since $C_0^\infty(Q)$ is dense in $L^2(Q)$, it follows

$$
-\partial_t p - \frac{1}{2}\sigma^2 \partial_{xx}^2 p - b\ \partial_x p = 0 \quad \text{in } Q, \tag{2.45}
$$

i.e. the integral over Q in (2.44) vanishes. Next, the requirement $y(\cdot, t_n+T_s) = 0$ is dropped, resulting in

$$
0 = \int_\Omega \big(\bar{y}(\cdot, t_n+T_s) - y_d(\cdot, t_n+T_s) - p(\cdot, t_n+T_s)\big)\ y(\cdot, t_n+T_s)\, dx.
$$

Analogously[5]*, this yields*

$$p(\cdot, t_n + T_s) = \bar{y}(\cdot, t_n + T_s) - y_d(\cdot, t_n + T_s) \quad \text{in } \Omega. \tag{2.46}$$

Finally, $(\partial_x y)_{|\Sigma} = 0$ *is dropped. Therefore,*

$$0 = \int_{t_n}^{t_n+T_s} \left(\frac{1}{2} p \; \partial_x \left(\sigma^2 y \right) \right) \Big|_{x_L}^{x_R} dt,$$

and, consequently,

$$p = 0 \quad \text{in } \Sigma, \tag{2.47}$$

hold for all sufficiently smooth y *with* $y(\cdot, t_n) = 0$ *and* $y_{|\Sigma} = 0$*. Thus,* p *is characterized by* (2.45)*,* (2.46)*, and* (2.47)*.*
In summary, the optimality system consists of the original IBVP

$$\begin{aligned} \partial_t y - \frac{1}{2} \partial_{xx}^2 \left(\sigma^2 y \right) + \partial_x \big(b \; y \big) &= 0 \quad \text{in } Q, \\ y(\cdot, t_n) &= \rho \quad \text{in } \Omega, \\ y &= 0 \quad \text{in } \Sigma, \end{aligned} \tag{2.48}$$

the corresponding adjoint problem

$$\begin{aligned} -\partial_t p - \frac{1}{2} \sigma^2 \partial_{xx}^2 p - b \; \partial_x p &= 0 && \text{in } Q, \\ p(\cdot, t_n + T_s) &= y(\cdot, t_n + T_s) - y_d(\cdot, t_n + T_s) && \text{in } \Omega, \\ p &= 0 && \text{in } \Sigma, \end{aligned} \tag{2.49}$$

and the variational inequality

$$\left(\lambda \bar{u} - \iint_Q \partial_x \left(\frac{\partial b}{\partial u} \; y \right) p \, dx \, dt \right) (u - \bar{u}) \geq 0, \quad u \in U_{ad}. \tag{2.50}$$

Solving the optimality system (2.48)-(2.50) yields a candidate for an optimal control $\bar{u}$ with corresponding state y and adjoint state p.[6] This process is applied to the other control problems as well.

Example 2.25
Consider the optimal control problem (2.24)-(2.28) *with* $\Omega :=]x_L, x_R[$*. The Lagrange function*

[5]The fact that the possible values for $y(\cdot, t_n + T_s)$ and $\partial_\nu y$ (cf. the last integral term in (2.44)) are dense in $L^2(\Omega)$ and $L^2(\Sigma)$, respectively, is used without proof in this formal derivation.

[6]The bar above y is omitted intentionally.

is given by

$$\mathcal{L}(y,u,p) = \frac{1}{2}\iint\limits_Q (y-y_d)^2\,dx\,dt + \frac{\lambda}{2}\int\limits_{t_n}^{t_n+(N-1)T_s} u^2\,dt - \iint\limits_Q \left(\partial_t y - \frac{1}{2}\partial_{xx}^2\left(\sigma^2 y\right) + \partial_x\big(b\ y\big)\right) p\,dx\,dt.$$

Then

$$\begin{aligned} D_u\mathcal{L}(\bar y,\bar u,p)(u-\bar u) &= \int\limits_{t_n}^{t_n+(N-1)T_s} \lambda\bar u\ (u-\bar u)\,dt - \iint\limits_Q \partial_x\big(D_3(b)\ \bar y\big)p\ (u-\bar u)\,dx\,dt \\ &= \int\limits_{t_n}^{t_n+(N-1)T_s} \left(\lambda\bar u - \int\limits_\Omega \partial_x\big(D_3(b)\ \bar y\big)p\,dx\right)(u-\bar u)\ dt \end{aligned}$$

and

$$D_y\mathcal{L}(\bar y,\bar u,p)\ y = \iint\limits_Q (\bar y - y_d)\ y\,dx\,dt - \iint\limits_Q \left(\partial_t y - \frac{1}{2}\partial_{xx}^2\left(\sigma^2 y\right) + \partial_x\big(b\ y\big)\right) p\,dx\,dt.$$

Therefore, analogous to (2.44)*,*

$$\begin{aligned} 0 = D_y\mathcal{L}(\bar y,\bar u,p)\ y &= -\int\limits_\Omega p(\cdot,t_n+(N-1)T_s)\ y(\cdot,t_n+(N-1)T_s)\,dx \\ &\quad - \iint\limits_Q \left(-(\bar y - y_d) - \partial_t p - \frac{1}{2}\sigma^2\partial_{xx}^2 p - b\ \partial_x p\right) y\,dx\,dt \\ &\quad + \int\limits_{t_n}^{t_n+(N-1)T_s} \left(\frac{1}{2}p\ \partial_x\left(\sigma^2 y\right)\right)\Big|_{x_L}^{x_R} dt \end{aligned} \tag{2.51}$$

for all sufficiently smooth y with $y(\cdot,t_n)=0$ and $y_{|\Sigma}=0$. Using the same arguments as in Example 2.24, the adjoint problem is determined. Eventually, the optimality system is given by

$$(2.48)\ \textit{with adjusted } Q \textit{ and } \Sigma \textit{ and } \bar u = \bar u(t), \tag{2.52}$$

the corresponding adjoint problem

$$\begin{aligned} -\partial_t p - \frac{1}{2}\sigma^2\partial_{xx}^2 p - b\ \partial_x p = y - y_d \quad & \textit{in } Q, \\ p(\cdot,t_n+(N-1)T_s) = 0 \quad & \textit{in } \Omega, \\ p = 0 \quad & \textit{in } \Sigma, \end{aligned} \tag{2.53}$$

and the variational inequality

$$\int_{t_n}^{t_n+(N-1)T_s} \left(\lambda\bar{u} - \int_\Omega \partial_x\big(D_3(b)\ y\big)p\,dx \right) (u-\bar{u})\ dt \geq 0, \quad u \in U_{ad}. \tag{2.54}$$

Example 2.26
Consider the optimal control problem (2.29)-(2.33) *with* $\Omega :=]x_L, x_R[$. *The Lagrange function is defined by*

$$\begin{aligned}\mathcal{L}(y,u,p) = \frac{\alpha}{2}\int_\Omega \big(y(\cdot,t_n+T_s) - y_d(\cdot,t_n+T_s)\big)^2\,dx + \frac{\lambda}{2}\int_\Omega u^2\,dx \\ - \iint_Q \left(\partial_t y - \frac{1}{2}\partial_{xx}^2\left(\sigma^2 y\right) + \partial_x\big(b\ y\big)\right) p\,dx\,dt\end{aligned}$$

Compared to the other two examples, it is harder to determine $D_u\mathcal{L}(\bar{y},\bar{u},p)(u-\bar{u})$ *since the control* $u = u(x)$ *is now space-dependent. The problem is the* $\partial_x\big(b\ y\big)$ *or, more precisely, the resulting* $\partial_x u$ *term: Since*

$$\partial_x\big(b\ y\big) = (D_1(b) + D_3(b)\ \partial_x u)\ y + b\ \partial_x y,$$

it follows

$$\begin{aligned}&D_u\mathcal{L}(\bar{y},\bar{u},p)(u-\bar{u}) \\ &= \int_\Omega \lambda\bar{u}\ (u-\bar{u})\,dx - D_u\left(\iint_Q \left[(D_1(b) + D_3(b)\ \partial_x u)\ y + b\ \partial_x y\right] p\,dx\,dt\right)\Big|_{(y,u)=(\bar{y},\bar{u})} (u-\bar{u}).\end{aligned}$$

Because of

$$D_u\left(\iint_Q \partial_x u\,dx\,dt\right)(u-\bar{u}) = \iint_Q \partial_x\,(u-\bar{u})\ dx\,dt,$$

it is not possible to factor out $u-\bar{u}$. *However, this will be essential later, e.g. when the gradient of the reduced cost functional is sought, cf.* (2.69). *This difficulty can be circumvented*

by integrating by parts:

$$\begin{aligned}
&D_u\mathcal{L}(\bar{y},\bar{u},p)(u-\bar{u})\\
&= \int_\Omega \lambda\bar{u}\ (u-\bar{u})\,dx - D_u\left(\iint_Q \partial_x\big(b\ y\big)\ p\,dx\,dt\right)\Big|_{(y,u)=(\bar{y},\bar{u})} (u-\bar{u})\\
&= \int_\Omega \lambda\bar{u}\ (u-\bar{u})\,dx\\
&\quad - D_u\left(-\iint_Q b\ y\ \partial_x p\,dx\,dt + \int_{t_n}^{t_n+T_s} \underbrace{(b\ y\ p)|_{x_L}^{x_R}}_{=0}\,dt\right)\Big|_{(y,u)=(\bar{y},\bar{u})} (u-\bar{u})\\
&= \int_\Omega \left(\lambda\bar{u} + \int_{t_n}^{t_n+T_s} D_3(b)\ \bar{y}\ \partial_x p\,dt\right)(u-\bar{u})\,dx.
\end{aligned} \tag{2.55}$$

Apart from the factor α, $\mathcal{L}(y,u,p)$ is identical to the Lagrange function in Example 2.24. Hence,

$$\begin{aligned}
0 = D_y\mathcal{L}(\bar{y},\bar{u},p)\ y &= \int_\Omega \Big(\alpha\big(\bar{y}(\cdot,t_n+T_s) - y_d(\cdot,t_n+T_s)\big) - p(\cdot,t_n+T_s)\Big)\ y(\cdot,t_n+T_s)\,dx\\
&\quad - \iint_Q \left(-\partial_t p - \frac{1}{2}\sigma^2\partial_{xx}^2 p - b\ \partial_x p\right) y\,dx\,dt + \int_{t_n}^{t_n+T_s}\left(\frac{1}{2}p\ \partial_x\left(\sigma^2 y\right)\right)\Big|_{x_L}^{x_R} dt
\end{aligned}$$

for all sufficiently smooth y with $y(\cdot,t_n)=0$ and $y_{|\Sigma}=0$. Analogously to Example 2.24, the adjoint problem is determined. Eventually, the result is the following optimality system, which consists of

$$(2.48)\ \textit{with}\ \bar{u}=\bar{u}(x), \tag{2.56}$$

the corresponding adjoint problem

$$\begin{aligned}
-\partial_t p - \frac{1}{2}\sigma^2\partial_{xx}^2 p - b\ \partial_x p = 0 &\qquad \textit{in } Q,\\
p(\cdot,t_n+T_s) = \big(y(\cdot,t_n+T_s) - y_d(\cdot,t_n+T_s)\big)\,\alpha &\qquad \textit{in } \Omega,\\
p = 0 &\qquad \textit{in } \Sigma,
\end{aligned} \tag{2.57}$$

and the variational inequality

$$\int_\Omega \left(\lambda\bar{u} + \int_{t_n}^{t_n+T_s} D_3(b)\ y\ \partial_x p\,dt\right)(u-\bar{u})\,dx \geq 0, \quad u\in U_{ad}. \tag{2.58}$$

Example 2.27
Consider the optimal control problem (2.34)-(2.38) *with* $\Omega :=]x_L, x_R[$. *The Lagrange function is given by*

$$\mathcal{L}(y,u,p) = \frac{\alpha}{2}\iint\limits_Q (y-y_d)^2\,dx\,dt + \frac{\lambda}{2}\iint\limits_Q u^2\,dx\,dt - \iint\limits_Q \Big(\partial_t y - \frac{1}{2}\partial_{xx}^2\left(\sigma^2 y\right) + \partial_x\big(b\ y\big)\Big)\,p\,dx\,dt.$$

Determining $D_u\mathcal{L}(\bar y,\bar u,p)(u-\bar u)$ *is analogous to Example 2.26, i.e.*

$$D_u\mathcal{L}(\bar y,\bar u,p)(u-\bar u) = \iint\limits_Q \Big(\lambda\bar u + D_3(b)\ \bar y\ \partial_x p\Big)(u-\bar u)\,dx\,dt. \tag{2.59}$$

$D_y\mathcal{L}(\bar y,\bar u,p)\ y$ *is almost the same as in Example 2.25:*

$$D_y\mathcal{L}(\bar y,\bar u,p)\ y = \alpha\iint\limits_Q (\bar y - y_d)\ y\,dx\,dt - \iint\limits_Q \Big(\partial_t y - \frac{1}{2}\partial_{xx}^2\left(\sigma^2 y\right) + \partial_x\big(b\ y\big)\Big)\,p\,dx\,dt.$$

Therefore,

$$0 = -\int\limits_\Omega p(\cdot,t_n+(N-1)T_s)\ y(\cdot,t_n+(N-1)T_s)\,dx$$
$$-\iint\limits_Q \Big(-\alpha(\bar y - y_d) - \partial_t p - \frac{1}{2}\sigma^2\partial_{xx}^2 p - b\ \partial_x p\Big)\,y\,dx\,dt + \int\limits_{t_n}^{t_n+(N-1)T_s}\Big(\frac{1}{2}p\ \partial_x\left(\sigma^2 y\right)\Big)\Big|_{x_L}^{x_R} dt.$$

As before, the adjoint problem is determined, resulting in the following optimality system:

$$(2.48)\ \textit{with adjusted}\ Q\ \textit{and}\ \Sigma\ \textit{and}\ \bar u = \bar u(x,t), \tag{2.60}$$

the corresponding adjoint problem

$$\begin{aligned} -\partial_t p - \frac{1}{2}\sigma^2\partial_{xx}^2 p - b\ \partial_x p &= (y-y_d)\ \alpha \quad \textit{in } Q,\\ p(\cdot,t_n+(N-1)T_s) &= 0 \qquad \textit{in } \Omega,\\ p &= 0 \qquad \textit{in } \Sigma, \end{aligned} \tag{2.61}$$

and the variational inequality

$$\iint\limits_Q \Big(\lambda\bar u + D_3(b)\ y\ \partial_x p\Big)(u-\bar u)\,dx\,dt \geq 0, \quad u\in U_{ad}. \tag{2.62}$$

Remark 2.28
The adjoint problems (2.49), (2.53), (2.57), *and* (2.61) *evolve backwards in time. However, they can be rewritten such that the time ordering is reversed. Exemplarily, this is done below for* (2.61).
Let $T := t_n + (N-1)T_s$. *Then by the time transformation*

$$\tau := T - t$$

and by defining

$$\begin{aligned}
\tilde{p}(x,\tau) &:= p(x,\underbrace{T-\tau}_{=t}),\\
\tilde{\sigma}(x,\tau) &:= \sigma(x, T-\tau),\\
\tilde{b}(x,\tau;\tilde{u}(x,\tau)) &:= b(x, T-\tau; u(x, T-\tau)),\\
\tilde{y}(x,\tau) &:= y(x, T-\tau),\\
\tilde{y}_d(x,\tau) &:= y_d(x, T-\tau),
\end{aligned}$$

the time ordering is reversed, resulting in the following IBVP:

$$\begin{aligned}
\partial_\tau \tilde{p}(x,\tau) - \frac{1}{2}\tilde{\sigma}^2(x,\tau)\partial^2_{xx}\tilde{p}(x,\tau) - \tilde{b}(x,\tau;\tilde{u}(x,\tau))\ \partial_x\tilde{p}(x,\tau) &= (\tilde{y}(x,\tau) - \tilde{y}_d(x,\tau))\ \alpha \quad \text{in } Q\\
\tilde{p}(\cdot,0) &= 0 \quad \text{in } \Omega\\
\tilde{p} &= 0 \quad \text{in } \Sigma
\end{aligned}$$

Note that

$$D_\tau\tilde{p}(x,\tau) = D_\tau p(x, T-\tau) = -D_t p(x,t).$$

Similar to Theorem 2.12, [16, p. 46, Corollary 1.1] ensures a unique solution $\tilde{p}$ *of the transformed adjoint problem. Backward substitution by* $p(x,t) = \tilde{p}(x,\tau)$ *yields the adjoint state* p.

The optimality systems (2.48)-(2.50), (2.52)-(2.54), (2.56)-(2.58), and (2.60)-(2.62) derived with the help of the formal Lagrange technique are used to solve the optimal control problems (2.20)-(2.23), (2.24)-(2.28), (2.29)-(2.33), and (2.34)-(2.38), respectively. However, it is not clear how to algorithmically process the variational inequalities (2.54), (2.58), and (2.62), i.e. if u is a function of time (and space). The solution is to formulate a point-wise variational inequality in $\mathbb{R}$ instead. Exemplarily, this is done below for (2.62). However, the same arguments apply to (2.54) and (2.58).
The point-wise variational inequality in the case of (2.62) is given by

$$\Big(\lambda\bar{u}(x,t) + D_3(b)(x,t;\bar{u}(x,t))\ y(x,t)\ \partial_x p(x,t)\Big)(v - \bar{u}(x,t)) \geq 0 \tag{2.63}$$

for all $v \in [u_a(x,t), u_b(x,t)]$ and almost all $(x,t) \in Q$. Moreover, the following Lemma[7] holds.

[7]This Lemma and its proof are based on [30, Lemma 2.26] and the proof thereof.

Lemma 2.29
The following statements are equivalent:
(i) (2.62) *holds*
(ii) $\bar{u}(x,t) \begin{cases} = u_a(x,t), & \text{if } \lambda\bar{u}(x,t) + D_3(b)(x,t;\bar{u}(x,t))\ y(x,t)\ \partial_x p(x,t) > 0 \\ \in [u_a(x,t), u_b(x,t)], & \text{if } \lambda\bar{u}(x,t) + D_3(b)(x,t;\bar{u}(x,t))\ y(x,t)\ \partial_x p(x,t) = 0 \\ = u_b(x,t), & \text{if } \lambda\bar{u}(x,t) + D_3(b)(x,t;\bar{u}(x,t))\ y(x,t)\ \partial_x p(x,t) < 0 \end{cases}$
(iii) (2.63) *holds for all* $v \in [u_a(x,t), u_b(x,t)]$ *and almost all* $(x,t) \in Q$.

Proof. (i) $\Rightarrow$ (ii): (Proof by contradiction.) Assume (ii) does not hold. Define the measurable sets

$$A_+(\bar{u}) := \{(x,t) \in Q \mid \lambda\bar{u}(x,t) + D_3(b)(x,t;\bar{u}(x,t))\ y(x,t)\ \partial_x p(x,t) > 0\},$$
$$A_-(\bar{u}) := \{(x,t) \in Q \mid \lambda\bar{u}(x,t) + D_3(b)(x,t;\bar{u}(x,t))\ y(x,t)\ \partial_x p(x,t) < 0\},$$

where $\bar{u}, u_a$, and u_b are each arbitrary but fixed representatives of their respective equivalence class. Since (ii) does not hold, there exists a set $E_+ \subset A_+(\bar{u})$ with $|E_+| > 0$ and $\bar{u}(x,t) > u_a(x,t)$ for all $(x,t) \in E_+$ or there exists a set $E_- \subset A_-(\bar{u})$ with $|E_-| > 0$ and $\bar{u}(x,t) < u_b(x,t)$ for all $(x,t) \in E_-$. Define

$$u(x,t) := \begin{cases} u_a(x,t), & (x,t) \in E_+, \\ u_b(x,t), & (x,t) \in E_-, \\ \bar{u}(x,t), & (x,t) \in Q \setminus (E_+ \cup E_-). \end{cases}$$

Then from (i) follows

$$\begin{aligned} &\iint_Q \Big(\lambda\bar{u}(x,t) + D_3(b)(x,t;\bar{u}(x,t))\ y(x,t)\ \partial_x p(x,t)\Big)(u(x,t) - \bar{u}(x,t))\,dx\,dt \\ &= \iint_{E_+} \underbrace{\Big(\lambda\bar{u}(x,t) + D_3(b)(x,t;\bar{u}(x,t))\ y(x,t)\ \partial_x p(x,t)\Big)}_{>0} \underbrace{(u_a(x,t) - \bar{u}(x,t))}_{<0}\,dx\,dt \\ &+ \iint_{E_-} \underbrace{\Big(\lambda\bar{u}(x,t) + D_3(b)(x,t;\bar{u}(x,t))\ y(x,t)\ \partial_x p(x,t)\Big)}_{<0} \underbrace{(u_b(x,t) - \bar{u}(x,t))}_{>0}\,dx\,dt < 0, \end{aligned}$$

which contradicts (i).
(ii) $\Rightarrow$ (iii): From (ii) follows $\bar{u}(x,t) = u_a(x,t)$ almost everywhere in $A_+(\bar{u})$ and therefore, $v - \bar{u}(x,t) \geq 0$ for all $v \in [u_a(x,t), u_b(x,t)]$. Hence,

$$\underbrace{\Big(\lambda\bar{u}(x,t) + D_3(b)(x,t;\bar{u}(x,t))\ y(x,t)\ \partial_x p(x,t)\Big)}_{>0} \underbrace{(v - \bar{u}(x,t))}_{\geq 0} \geq 0$$

almost everywhere in $A_+(\bar{u})$. Analogously, the same inequality holds almost everywhere in $A_-(\bar{u})$. Almost everywhere in $Q \setminus (A_+(\bar{u}) \cup A_-(\bar{u}))$ this inequality obviously holds as well: there,

$$\Big(\lambda\bar{u}(x,t) + D_3(b)(x,t;\bar{u}(x,t))\ y(x,t)\ \partial_x p(x,t)\Big) = 0$$

holds. Therefore, (2.63) holds for all $v \in [u_a(x,t), u_b(x,t)]$ and almost all $(x,t) \in Q$.
(iii) $\Rightarrow$ (i): Let $u \in U_{ad}$. Since $\bar{u}(x,t) \in [u_a(x,t), u_b(x,t)]$ almost everywhere in Q, with $v := u(x,t)$ it follows from (iii) that

$$\Big(\lambda\bar{u}(x,t) + D_3(b)(x,t;\bar{u}(x,t))\; y(x,t)\; \partial_x p(x,t)\Big)(u(x,t) - \bar{u}(x,t)) \geq 0, \quad \text{almost all } (x,t) \in Q.$$

Integrating this inequality yields (i). □

In the case of

$$\lambda\bar{u}(x,t) + D_3(b)(x,t;\bar{u}(x,t))\; y(x,t)\; \partial_x p(x,t) \neq 0$$

Lemma 2.29(ii) determines $\bar{u}(x,t)$. If

$$\lambda\bar{u}(x,t) + D_3(b)(x,t;\bar{u}(x,t))\; y(x,t)\; \partial_x p(x,t) = 0,$$

then

$$\bar{u}(x,t) = -\frac{1}{\lambda} D_3(b)(x,t;\bar{u}(x,t))\; y(x,t)\; \partial_x p(x,t) \tag{2.64}$$

at least indicates how the control $\bar{u}(x,t)$ should be chosen. If $D_3(D_3(b)) \equiv 0$ then (2.64) is an explicit formula for $\bar{u}(x,t)$. For $D_3(D_3(b)) \equiv 0$, the different cases can be subsumed by the explicit projection formula

$$\bar{u}(x,t) = \mathbb{P}_{[u_a(x,t),u_b(x,t)]}\left\{-\frac{1}{\lambda} D_3(b)(x,t;\bar{u}(x,t))\; y(x,t)\; \partial_x p(x,t)\right\} \tag{2.65}$$

for almost all $(x,t) \in Q$, where $\mathbb{P}_{[a,b]}$ with $a \leq b \in \mathbb{R}$ is the projection of $\mathbb{R}$ onto $[a,b]$, i.e.

$$\mathbb{P}_{[a,b]}(u) := \min\{b, \max\{a,u\}\},$$

cf. Figure 2.1.

Remark 2.30
$D_3(D_3(b)) \equiv 0$ holds for all stochastic processes considered in this work, cf. Example 2.6.

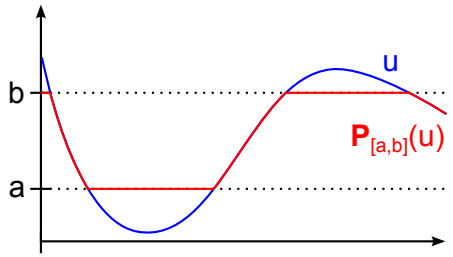


Figure 2.1: Exemplary projection

For completeness, the projection formulas for Examples 2.24, 2.25, and 2.26, respectively, are given below:

$$\bar{u} = \mathbb{P}_{[u_a,u_b]}\left\{\frac{1}{\lambda}\iint\limits_Q \partial_x\left(\frac{\partial b}{\partial u}(x,t;\bar{u})\ y(x,t)\right)p(x,t)\,dx\,dt\right\}, \tag{2.66}$$

$$\bar{u}(t) = \mathbb{P}_{[u_a(t),u_b(t)]}\left\{\frac{1}{\lambda}\int\limits_\Omega \partial_x\big(D_3(b)(x,t;\bar{u}(t))\ y(x,t)\big)\ p(x,t)\,dx\right\}, \tag{2.67}$$

$$\bar{u}(x) = \mathbb{P}_{[u_a(x),u_b(x)]}\left\{-\frac{1}{\lambda}\int\limits_{t_n}^{t_n+T_s} D_3(b)(x,t;\bar{u}(x))\ y(x,t)\ \partial_x p(x,t)\,dt\right\}. \tag{2.68}$$

Remark 2.31
It is also possible to use integration by parts to "transfer" the derivative w.r.t. x to p in (2.66) *and* (2.67)*, resulting in formulas that highlight the similarity to* (2.68) *and* (2.65)*.*

As mentioned in the beginning of this subsection, the variational inequalities can also be obtained by replacing the state y in the cost functional $J(y,u)$ with a solution operator S, i.e. $y = S(u)$, which leads to the reduced cost functional $\hat{J}(u) = J(S(u),u)$, cf. Subsection 2.3.1. Differentiating it w.r.t. u yields the variational inequality

$$\langle \hat{J}'(\bar{u}), u-\bar{u}\rangle_{U^*,U} \geq 0, \quad u \in U_{ad} \subset U,$$

where $\langle\cdot,\cdot\rangle$ is the duality pairing, cf. Definition A.2. By the Riesz Representation Theorem A.4 the derivative can be identified with the gradient $\nabla\hat{J}(u)$ by means of

$$\langle \hat{J}'(\bar{u}), u-\bar{u}\rangle_{U^*,U} = \Big(\nabla\hat{J}(\bar{u}), u-\bar{u}\Big)_U, \quad u \in U_{ad} \subset U. \tag{2.69}$$

Remark 2.32
In the literature, sometimes $\hat{J}'(u)$ is written for the gradient $\nabla\hat{J}(u)$. The difference between $\hat{J}'(u)$ and $\nabla\hat{J}(u)$ is that $\hat{J}'(u)$ is an element in U^*, the dual space of U, while the Riesz representative $\nabla\hat{J}(u)$ is an element in U.

Notice that the inner product in U is used on the right-hand side of (2.69). In the optimal control problems (2.20)-(2.23), (2.24)-(2.28), (2.29)-(2.33), and (2.34)-(2.38), U is either some L^2 space or the space of real values, $\mathbb{R}$. In either case, assuming the corresponding optimality systems derived in Examples 2.24-2.27 are correct, $\nabla\hat{J}(\bar{u})$ can be written as

$$\nabla\hat{J}(\bar{u}) = \lambda\bar{u} - \iint\limits_Q \partial_x\left(\frac{\partial b}{\partial u}\ y\right)p\,dx\,dt \tag{2.70}$$

in the case of (2.20)-(2.23). Analogously,

$$\nabla\hat{J}(\bar{u}) = \lambda\bar{u} - \int\limits_\Omega \partial_x\big(D_3(b)\ y\big)p\,dx \tag{2.71}$$

in the case of (2.24)-(2.28),

$$\nabla \hat{J}(\bar{u}) = \lambda \bar{u} + \int_{t_n}^{t_n+T_s} D_3(b)\ y\ \partial_x p\, dt \tag{2.72}$$

in the case of (2.29)-(2.33), and

$$\nabla \hat{J}(\bar{u}) = \lambda \bar{u} + D_3(b)\ y\ \partial_x p \tag{2.73}$$

in the case of (2.34)-(2.38). Notice that for $\nabla \hat{J}(\bar{u}) = 0$ and $D_3(D_3(b)) \equiv 0$, one can solve (2.70)-(2.73) for $\bar{u}$. Doing so yields the arguments of the above projection formulas (2.66)-(2.68) and (2.65).

Chapter 3

Model Predictive Control

Model predictive control is an optimization based method for the feedback control of (nonlinear) systems [11]. In contrast to open-loop optimal control problems, such as, for example (2.34)-(2.38), where the control u is optimized once on the whole time interval (i.e. the whole optimization horizon), MPC computes a closed-loop control. One advantage of a closed-loop controller is the ability to incorporate disturbances that may shift the state. In addition, the main benefits of MPC are 1) the ability to deal with infinite optimization horizon, 2) the anticipation of future events in the current optimization step and 3) the possibility to easily implement control and state constraints.
In this thesis, MPC is used to compute an optimal control for the Fokker-Planck equation with the objective to track a sequence of target probability density functions, trying to match the respective densities as good as possible at given times. This has already been done in [3] and [4] with a space-independent control. However, only the shortest possible horizon was considered. The horizon length is a key parameter to tune in the MPC algorithm, cf. Section 3.1. In many cases, satisfying results can only be achieved by a high enough horizon, which entails more computing time. Therefore, an analysis whether a longer horizon yields better results is conducted in this work. To this end, a short introduction to MPC is given in Section 3.1. Stability of MPC is briefly discussed in Section 3.2. Details and many results, especially concerning stability of MPC, can be found in [11] and [26]. Afterwards, in Section 3.3, the MPC scheme is applied to solve the above tracking problem.

3.1 Introduction to Model Predictive Control

In the MPC setting, the considered systems are the following.

Definition 3.1 (Nonlinear discrete time control system)
Let Z, U be metric spaces and $f\colon Z \times U \to Z$. The nonlinear discrete time control system *is defined as follows:*

$$z(n+1) = f(z(n), u(n)) \tag{3.1}$$

Notation

- $\mathcal{U} := \{u \mid u\colon \mathbb{N}_0 \to U\}$ *is the space of control sequences.*
- *The solution trajectory for given initial state* z_0 *and control sequence* $(u(n))_{n\in\mathbb{N}_0} \subseteq U$ *is denoted by* $z_u(\cdot) = z_u(\cdot; z_0)$.

The main idea behind MPC is to truncate the original optimization horizon, solve the optimal control problem only on a smaller, finite horizon and thereby obtain an optimal control sequence. The horizon length is defined by a parameter $N \in \mathbb{N}_{\geq 2}$. Then, instead of simply implementing the sequence, only the first value of the sequence is implemented. Afterwards, the optimization (or prediction) horizon is shifted in time so that the new (measured) state becomes the present state. Based on this state, another optimal control problem is solved on the new horizon, and so on, until a terminal time T_E is reached. This process is depicted in Figure 3.1. In the literature, MPC is also denoted by Receding Horizon Control (RHC) for its moving horizon characteristic. By only implementing a part of the optimal control sequence and measuring the current state of the system, disturbances in the state and new, possibly unexpected information regarding the reference trajectory are accounted for when optimizing.
The optimization problem that needs to be solved in every time step is given by

$$\begin{aligned} \min_{u(\cdot)\in\mathcal{U}} J_N(z_0, u(\cdot)) := \sum_{k=0}^{N-1} l(z_u(k; z_0), u(k)) \quad \text{s.t.} \\ z_u(0; z_0) = z_0, \\ z_u(k+1; z_0) = f(z_u(k; z_0), u(k)), \end{aligned} \tag{OCP_N}$$

where $l\colon Z \times U \to \mathbb{R}_{\geq 0}$ is a continuous function defining the so-called stage costs and f is given by the nonlinear discrete time control system, cf. Definition 3.1.
By following the above procedure, an MPC-Feedback $\mu_N\colon Z \to U$ and the corresponding solution trajectory z_{μ_N} of the closed-loop system

$$z_{\mu_N}(n+1; z_0) = f(z_{\mu_N}(n; z_0), \mu_N(z_{\mu_N}(n; z_0))) \tag{3.2}$$

is obtained. This iterative process is specified in the following algorithm.

Algorithm 3.2 (MPC Algorithm)
Fix some $N \in \mathbb{N}_{\geq 2}$. *At each time* $t_n, n = 0, 1, 2, \ldots$

1. *Measure the state* $z(n) \in Z$ *of the system.*
2. *Set* $z_0 := z(n)$ *and solve* (OCP_N) *in order to obtain the optimal control sequence denoted by* $u^*(0), \ldots, u^*(N-1)$.
3. *Set* $\mu_N(z_{\mu_N}(n; z_0)) := u^*(0)$, *compute* $z_{\mu_N}(n+1; z_0)$ *according to* (3.2), *set* $n := n+1$ *and go to 1.*

Remark 3.3
Only the first value of the optimal control sequence $u^*(0), ..., u^*(N-1)$ *is used. However, the other values may be used as initial guesses for the new optimal control in the next optimization step.*

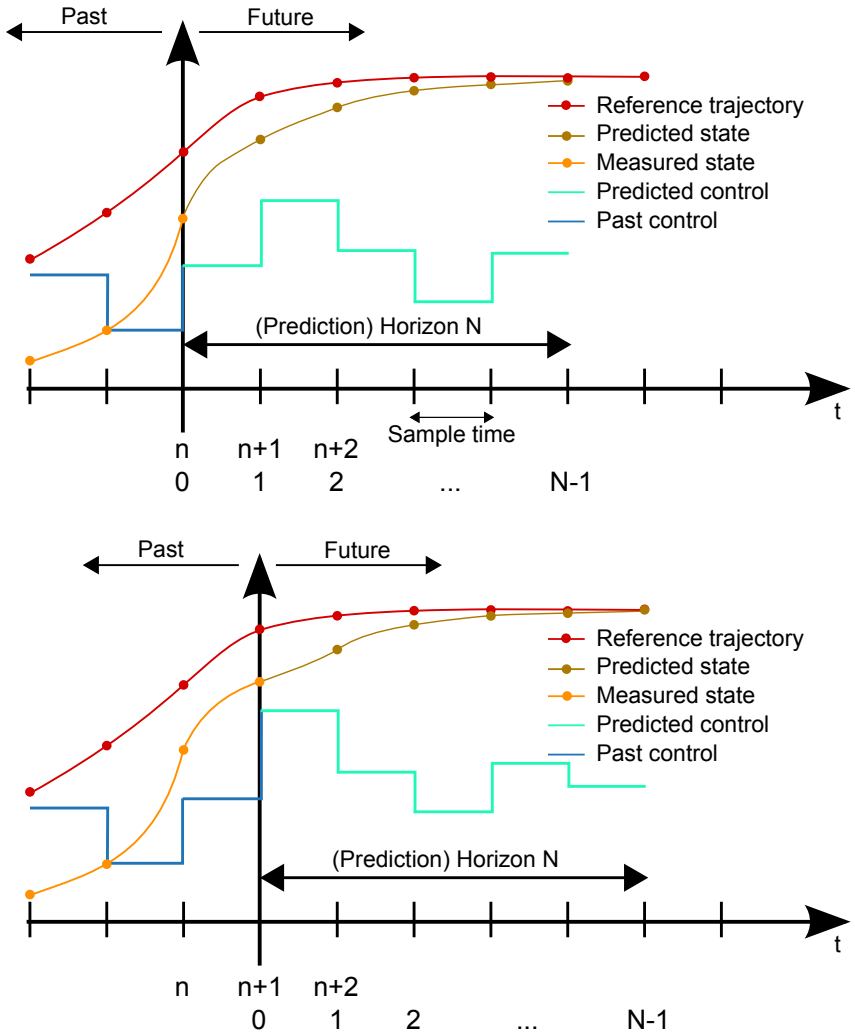


Figure 3.1: MPC scheme (discrete times n (top) and $n+1$ (bottom)) (Template source: `http://commons.wikimedia.org/wiki/File:MPC_scheme_basic.svg`, retrieved on October 24, 2013)

3.2 Stability

As already mentioned in the introduction of this chapter, MPC is used to compute an optimal control for the Fokker-Planck equation with the objective to track a sequence of target probability density functions. More generally, the goal of MPC is to control the state $z(n)$ of a (nonlinear) system toward a reference trajectory and keep it close to this reference [11], cf. Figure 3.1. Under certain assumptions theoretical results guarantee that the MPC feedback law μ_N does exactly that, i.e. the closed-loop system (3.2) is "asymptotically stable". Hence, in the following, stability of nonlinear discrete systems

$$z(n+1) = f(z(n)) \tag{3.3}$$

with $f\colon Z \to Z$ and Z being a metric space is scrutinized. In the discussion to follow, only constant reference trajectories are considered. For time-varying reference trajectories, see [11].
If the reference trajectory is constant, i.e. if convergence to a certain state $z^* \in Z$ is sought, it is desirable for z^* to be an equilibrium point.

Definition 3.4 (Equilibrium point)
A point $z^* \in Z$ is an equilibrium point *of* (3.3) $:\Leftrightarrow \big(z(0) = z^* \Rightarrow z(n) = z^*$ *for all* $n \in \mathbb{N}_0\big)$, *i.e.* $z^* = f(z^*)$.

If convergence to a state z^* is impossible, one might be content with convergence to a set $\mathcal{A} \subset Z$. In this case, it is desirable that once the set is reached, the system stays in that set. This property is called positive invariance.

Definition 3.5 (Positive invariant set)
A set $\mathcal{A} \subset Z$ is positive invariant *for* (3.3) $:\Leftrightarrow \big(z \in \mathcal{A} \Rightarrow f(z) \in \mathcal{A}\big)$.

So-called comparison functions provide a convenient way to define stability properties of nonlinear systems.

Definition 3.6 ($\mathcal{K}, \mathcal{K}_\infty, \mathcal{L}$ and $\mathcal{KL}$ functions)
The following sets each define a class of comparison functions.

$$\mathcal{K} := \{\alpha\colon \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0} \mid \alpha \text{ continuous and strictly monotonically increasing with } \alpha(0) = 0\}$$
$$\mathcal{K}_\infty := \{\alpha\colon \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0} \mid \alpha \in \mathcal{K}, \alpha \text{ unbounded}\}$$
$$\mathcal{L} := \left\{\delta\colon \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0} \,\middle|\, \delta \text{ continuous, strictly monotonically decreasing and } \lim_{t\to\infty} \delta(t) = 0\right\}$$
$$\mathcal{KL} := \{\beta\colon \mathbb{R}_{\geq 0} \times \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0} \mid \beta \text{ continuous}, \beta(\cdot, t) \in \mathcal{K}, \beta(r, \cdot) \in \mathcal{L}\}$$

Furthermore, the following notation is used in the definition of stability.

Notation
Let Z be a metric space.

$$|z_1|_{z_2} := d_Z(z_1, z_2) \text{ denotes the distance between } z_1 \in Z \text{ and } z_2 \in Z.$$
$$B_\eta(z^*) := \left\{z \in Z \;\middle|\; |z|_{z^*} < \eta\right\} \text{ is the open ball around } z^* \in Z \text{ with radius } \eta.$$

Additionally, the trajectory of (3.3) *for a given initial state z_0 is denoted by $z(\cdot) = z(\cdot; z_0)$.*

Definition 3.7 (Asymptotic stability)
Let $z^* \in Z$ be an equilibrium point of (3.3).
(a) z^* is locally asymptotically stable $:\Leftrightarrow \exists \eta > 0, \beta \in \mathcal{KL}\ \forall z_0 \in B_\eta(z^*), n \in \mathbb{N}_0$:

$$|z(n; z_0)|_{z^*} \leq \beta(|z_0|_{z^*}, n).$$

(b) $z^* \in \mathcal{A}$ *is* asymptotically stable on a positive invariant set $\mathcal{A} \subset Z$ $:\Leftrightarrow \exists \beta \in \mathcal{KL}\ \forall z_0 \in \mathcal{A}, n \in \mathbb{N}_0$:

$$|z(n; z_0)|_{z^*} \leq \beta(|z_0|_{z^*}, n).$$

(c) $z^* \in \mathcal{A}$ *is* globally asymptotically stable $:\Leftrightarrow z^*$ *is asymptotically stable on* $\mathcal{A} = Z$.

This notion of stability contains two properties, the Lyapunov stability, and attraction; both are expressed by the $\mathcal{KL}$ function β. Again, for more details, see [11]. With Definition 3.7 in mind, the problem of proving stability of the closed-loop system (3.2) is tackled.

Definition 3.8 (Optimal value function)
Consider the cost functional J_N *in* (OCP_N). *The corresponding* optimal value function *is defined as*

$$V_N(z_0) := \inf_{u \in \mathcal{U}} J_N(z_0, u).$$

Definition 3.9 (Relaxed Lyapunov inequality)
The relaxed Lyapunov inequality *holds* $:\Leftrightarrow \exists \alpha \in]0, 1]\ \forall z \in Z$:

$$V_N(f(z, \mu_N(z))) \leq V_N(z) - \alpha\ l(z, \mu_N(z)), \tag{3.4}$$

where V_N *is the optimal value function.*

The relaxed Lyapunov inequality is a key ingredient to ensure stability. The interpretation is that the optimal value function has to decrease in time by some ratio of the stage costs in the current time step. If the stage costs are chosen carefully, this entails the state (slowly) approaching the desired state. To achieve this inequality, certain requirements are posed on the system and on the optimization horizon. One possible requirement is the exponential controllability of the system.

Definition 3.10 (Exponential controllability w.r.t. stage costs l)
The system (3.1) *is* exponentially controllable with respect to stage costs l
$:\Leftrightarrow \exists C \geq 1, \sigma \in]0, 1[\ \forall z_0 \in Z\ \exists u_z \in \mathcal{U}$:

$$l(z_{u_z}(n; z_0), u_z(n)) \leq C\sigma^n \underbrace{\min_{u \in U} l(z_0, u)}_{=: l^*(z_0)}, \quad n \in \mathbb{N}_0. \tag{3.5}$$

The second requirement is imposed on the MPC horizon N and is formulated in the following stability theorem.

Theorem 3.11 (Stability Theorem)
Let the controllability condition (3.5) *hold with* $C \geq 1$ *and* $\sigma \in]0, 1[$. *Choose* N *such that the* stability condition

$$\alpha_N := 1 - \frac{(\gamma_N - 1) \prod_{i=2}^{N} (\gamma_i - 1)}{\prod_{i=2}^{N} \gamma_i - \prod_{i=2}^{N} (\gamma_i - 1)} > 0$$

holds with

$$\gamma_i := C \sum_{n=0}^{i-1} \sigma^n = C \frac{1-\sigma^i}{1-\sigma}.$$

$\Rightarrow \forall z \in Z$: *The relaxed Lyapunov inequality* (3.4) *holds with* $\alpha = \alpha_N$.
Let, additionally, $\underline{\eta}, \overline{\eta} \in \mathcal{K}_\infty$ *satisfy*

$$\underline{\eta}(\|z - z^*\|) \leq l^*(z) \quad \text{and} \quad \overline{\eta}(\|z - z^*\|) \geq V_N(z), \ z \in Z$$

with $z^* \in Z$ *being an equilibrium point.*
$\Rightarrow$ *The trajectory generated by the closed-loop system* (3.2) *is asymptotically stable. In particular, it converges to the equilibrium* z^* *as* $n \to \infty$.

Proof. The first result follows directly from [11, Theorem 6.14] and [11, Proposition 6.17]. The second result is then a direct consequence of [11, Theorem 4.11]. □

3.3 MPC for the Fokker-Planck Equation

The various Fokker-Planck optimal control problems introduced in Section 2.3 all stemmed from the desire to find a control u such that, starting with an initial probability density function ρ, the stochastic process evolves such that a sequence of desired target probability densities is matched as good as possible at certain times. This task perfectly matches the idea of MPC. Therefore, the aim of this section is to explain the details on how to apply an MPC scheme to the Fokker-Planck optimal control problems (2.20)-(2.23) and (2.29)-(2.33). First, since the time in the Fokker-Planck equation is continuous, the PDE needs to be translated into a discrete system. This is done by sampling: The idea is to define a discrete system such that the trajectory of it coincides with the continuous solution of the PDE in some pre-defined sample times, cf. Figure 3.2. These systems are called sample data systems. For simplicity, only equidistant sampling times $t_n := nT_s, n \in \mathbb{N}_0$ for a given $T_s > 0$ are considered.

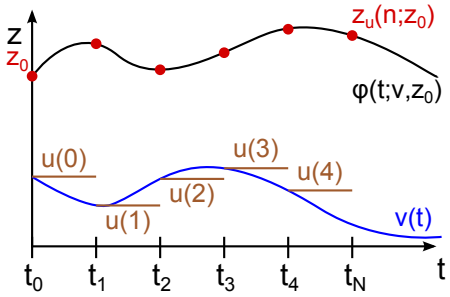


Figure 3.2: Sample data system

Let $\varphi(x, t; v, \rho)$ be the solution of the Fokker-Planck equation (2.7)-(2.9) with a control v. Furthermore, let $T_s > 0$. Define

$$z(n+1) := \varphi(x, T_s; u(n), z(n))$$

with $z(0) := \rho$. The discrete control u can then be defined as

$$u(n) := v(\cdot, t + nT_s)_{|[nT_s,(n+1)T_s]}$$

with $u \in L^\infty([0, T_E[, \mathbb{R})$ for (2.20)-(2.23) and $u \in L^\infty([0, T_E[, L^2(\Omega))$ for (2.29)-(2.33). Hence, the discrete time n corresponds to the continuous time nT_s, i.e.

$$z(n) = y(\cdot, nT_s) \in H_0^1(\Omega) =: Z.$$

Next, the stage costs l need to be defined. In Section 2.3, the idea was to partition the time interval $]0, T_E[$ and solve optimal control problems of type (2.20)-(2.23) on small time intervals, one by one, until the whole time interval $]0, T_E[$ is covered. This coincides with applying an MPC scheme on $]0, T_E[$ with the shortest possible horizon and l defined as follows:

$$l(z(n), u(n)) := \frac{1}{2} \|y(\cdot, nT_s) - y_d(\cdot, nT_s)\|^2_{L^2(\Omega)} + \frac{\lambda}{2}|v(nT_s)|^2. \tag{3.6}$$

As mentioned in Section 3.1, the shortest possible horizon length is $N = 2$. Hence, in each MPC optimization step (step 2 in Algorithm 3.2), the cost functional J_N in (OCP_N) results in

$$\begin{aligned}
J_2(z_0, \tilde{u}(\cdot)) &= \sum_{k=0}^{1} l(z_{\tilde{u}}(k; z_0), \tilde{u}(k)) = l(\underbrace{z_{\tilde{u}}(0; z_0)}_{=z_0=z(n)}, \underbrace{\tilde{u}(0)}_{=u(n)}) + l(\underbrace{z_{\tilde{u}}(1; z_0)}_{=z(n+1)}, \underbrace{\tilde{u}(1)}_{u(n+1)}) \\
&= \frac{1}{2} \|y(\cdot, nT_s) - y_d(\cdot, nT_s)\|^2_{L^2(\Omega)} + \frac{\lambda}{2}|v(\underbrace{nT_s}_{=t_n})|^2 \\
&+ \frac{1}{2} \|y(\cdot, (n+1)T_s) - y_d(\cdot, (n+1)T_s)\|^2_{L^2(\Omega)} + \frac{\lambda}{2}|v(\underbrace{(n+1)T_s}_{=t_n+T_s})|^2 \\
&= \frac{1}{2} \|y(\cdot, t_n) - y_d(\cdot, t_n)\|^2_{L^2(\Omega)} + \frac{\lambda}{2}|v(t_n)|^2 \\
&+ \frac{1}{2} \|y(\cdot, t_n + T_s) - y_d(\cdot, t_n + T_s)\|^2_{L^2(\Omega)} + \frac{\lambda}{2}|v(t_n + T_s)|^2,
\end{aligned}$$

which does not coincide with (2.20). However, the first term, $\frac{1}{2} \|y(\cdot, t_n) - y_d(\cdot, t_n)\|^2_{L^2(\Omega)}$, is constant because t_n is the current initial time, i.e. $y(\cdot, t_n)$ is a given. Constant terms in the cost functional do not change the optimum and can therefore be omitted. The last term, $\frac{\lambda}{2}|v(t_n+T_s)|^2$, is zero: It is always optimal to choose $\tilde{u}(1) = v(t_n+T_s) = 0$ because the control value $v(t_n + T_s)$ only influences the state y for $t \geq t_n + T_s$, but the currently considered time frame is $]t_n, t_n + T_s[$. Therefore, only

$$\frac{\lambda}{2}|v(t_n)|^2 + \frac{1}{2} \|y(\cdot, t_n + T_s) - y_d(\cdot, t_n + T_s)\|^2_{L^2(\Omega)} \tag{3.7}$$

is of relevance. In particular, the only parameter to optimize is $v(t_n) = \tilde{u}(0)$, a real value. Since this is exactly (2.20), the optimal control problem (OCP_N) defined by the above

discrete system and stage costs (3.6) is solved by solving the corresponding optimality system (2.48)-(2.50).
One aim of this thesis is to analyze the impact of the horizon length N. Increasing N while keeping the stage costs (3.6) leads to

$$\begin{aligned} J_N(z_0, \tilde{u}(\cdot)) &= \sum_{k=0}^{N-1} l(\underbrace{z_{\tilde{u}}(k; z_0)}_{=z(n+k)}, \underbrace{\tilde{u}(k)}_{=u(n+k)}) \\ &= \sum_{k=0}^{N-1} \left(\frac{1}{2} \left\| y(\cdot, (n+k)T_s) - y_d(\cdot, (n+k)T_s) \right\|^2_{L^2(\Omega)} + \frac{\lambda}{2} |v((n+k)T_s)|^2 \right) \\ &\approx \frac{1}{T_s} \int_{t_n}^{t_n+(N-1)T_s} \frac{1}{2} \left\| y(\cdot, t) - y_d(\cdot, t) \right\|^2_{L^2(\Omega)} + \frac{\lambda}{2} |v(t)|^2 \, dt, \end{aligned}$$

which is (2.24), short of the factor $\frac{1}{T_s}$. More precisely,

$$T_s \; J_N(z_0, \tilde{u}(\cdot)) = \int_{t_n}^{t_n+(N-1)T_s} \frac{1}{2} \left\| y(\cdot, t) - y_d(\cdot, t) \right\|^2_{L^2(\Omega)} + \frac{\lambda}{2} |v(t)|^2 \, dt + \mathcal{O}(T_s).$$

Since multiplying a cost functional by a real number does not change the optimum, the factor $\frac{1}{T_s}$ (or T_s) may as well be ignored. Therefore, for small T_s, the cost functional J_N approximates (2.24). Hence, the optimality system (2.52)-(2.54) is used to compute the optimal MPC feedback law μ_N.
Another aim of this work was to compare space-dependent controls with space-independent ones. Similar to (3.6), the stage costs are defined as

$$l(z(n), u(n)) := \frac{\alpha}{2} \left\| y(\cdot, nT_s) - y_d(\cdot, nT_s) \right\|^2_{L^2(\Omega)} + \frac{\lambda}{2} \left\| v(nT_s) \right\|^2_{L^2(\Omega)}. \tag{3.8}$$

Then, by the same reasoning as above, J_N with stage costs (3.8) is equivalent to (2.29) for $N = 2$ and approximates the cost functional (2.34) for higher N and small $T_s > 0$. Consequently, the optimal solution of (OCP_N) is obtained by solving the optimality system (2.56)-(2.58) in the former case and (2.60)-(2.62) in the latter.

Remark 3.12
If a terminal time T_E is fixed, employing an MPC scheme with $N > 2$ results in computing an optimal solution while considering the (desired) states after T_E. These, however, are not of interest. There are two ways to deal with this nuisance: Either change the horizon accordingly or leave it constant. The first approach results in a termination of the MPC algorithm as soon as the horizon is shortened for the first time. The reason behind this is simple: Since the reference trajectory and the predicted states up until T_E have already been taken into account in the computation of the optimal control sequence and since no

disturbances appear in the model, the whole sequence can be applied. The second approach on the other hand is more in line with the aim to obtain an asymptotically stable closed-loop system, cf. Section 3.2.

Remark 3.13
In the MPC literature, the minimal horizon length often is 2, despite the additional terms in the cost functional J_N that are either constant or zero. The upside of this notation is the ability to easily add so-called terminal costs to the stage costs in (OCP_N). These costs, however, are not the topic of this work. Furthermore, in the implementation of MPC for the Fokker-Planck equation, only the relevant terms appear in the cost functionals, cf. Chapter 5. Therefore, in the following chapters, the horizon parameter N is reinterpreted as the number of controls to be optimized in (OCP_N), rendering $N = 1$ the smallest possible value. In this case, the above cost functionals can be reformulated into

$$\begin{aligned} J_N(z_0, \tilde{u}(\cdot)) = \sum_{k=0}^{N-1} \Bigg(& \frac{1}{2} \left\| y(\cdot, (n+k+1)T_s) - y_d(\cdot, (n+k+1)T_s) \right\|^2_{L^2(\Omega)} \\ & + \frac{\lambda}{2} |v((n+k)T_s)|^2 \Bigg) \end{aligned} \tag{3.9}$$

and

$$\begin{aligned} J_N(z_0, \tilde{u}(\cdot)) = \sum_{k=0}^{N-1} \Bigg(& \frac{\alpha}{2} \left\| y(\cdot, (n+k+1)T_s) - y_d(\cdot, (n+k+1)T_s) \right\|^2_{L^2(\Omega)} \\ & + \frac{\lambda}{2} \|v((n+k)T_s)\|^2_{L^2(\Omega)} \Bigg), \end{aligned} \tag{3.10}$$

depending on whether the control is space-dependent or not. N still represents the horizon: A value of 1 means that only one step ahead is calculated, i.e. starting from $z_{\mu_1}(n; z_0)$ (or, equivalently, $y(\cdot, t_n)$), only $z_{\mu_1}(n+1; z_0)$ (or $y(\cdot, t_n + T_s)$) is computed.

One topic that has not been discussed in this section is the stability of the closed-loop system (3.2) for the Fokker-Planck equation and the above stage costs (3.6) and (3.8). The tools introduced in Section 3.2 – Theorem 3.11 in particular – provide a good starting point for a theoretical investigation. Using these tools, a similar analysis for the linear heat equation and a constant reference trajectory has been conducted in [2]. In this thesis, however, stability of the closed-loop system (3.2) remains an open question. The focus is on numerical results instead. On that account, the next chapter is about solving the Fokker-Planck optimal control problems numerically.

Chapter 4

Discretization of the Fokker-Planck Optimality Systems

In this chapter, the Fokker-Planck optimal control problems are tackled numerically. More precisely, it is discussed how to discretize the optimality systems (2.48)-(2.50), (2.52)-(2.54), (2.56)-(2.58), and (2.60)-(2.62). These systems each contain two PDEs that need to be solved. In some cases, the solution of the Fokker-Planck equation is known in closed form; for example, if the control u is simply a real value and the stochastic process at hand is the Ornstein-Uhlenbeck process, cf. Example 2.6. In the vast majority of practically relevant situations, however, the solution can only be obtained numerically. In this work, this is accomplished by a Finite Differences approach, which is presented in Section 4.1. Having solved both PDEs numerically, the (point-wise) variational inequality needs to hold. This leads to iterative optimization algorithms like the Projected Gradient method and a Newton method with a primal-dual active set strategy. Both are presented in Section 4.2.

4.1 Discretization of the Fokker-Planck Equation

The line of action in this section can be outlined as follows: First, the spatial domain $\Omega \subset \mathbb{R}$ is discretized by a grid Ω_h. On this grid, the derivatives w.r.t. x in the Fokker-Planck equation are approximated by Finite Differences, incorporating the boundary conditions into these approximations. Since the time is still continuous, this approach leads to a system of ordinary differential equations (ODEs) as large as the number of grid points. The ODE system is then discretized and solved using common ODE solvers like the Euler method.
For $\Omega =]x_L, x_R[\subset \mathbb{R}$, define the grid with a spatial mesh size $h > 0$ by

$$\Omega_h := \{x \in \mathbb{R} \mid x := x_j = j\ h, j \in \mathbb{Z}\} \cap \Omega.$$

It is assumed that h is chosen such that the boundaries of Ω coincide with the grid points, i.e.

$$\exists j_1, j_2 \in \mathbb{Z} : x_L = j_1\ h \text{ and } x_R = j_2\ h.$$

Ω_h is the set of interior grid points, Γ_h is the set of the two boundary points, i.e. $\Gamma_h = \{x_L, x_R\}$ and $\bar{\Omega}_h = \Omega_h \cup \Gamma_h$ is called the mesh.

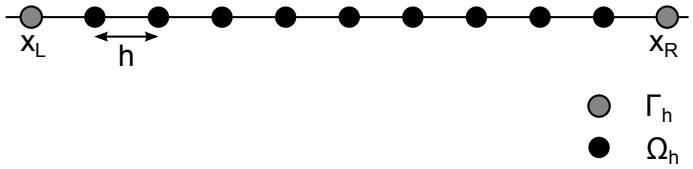


Figure 4.1: Exemplary grid

The next step is to approximate the spatial derivatives of functions f defined on Ω_h. The idea of Finite Differences is to approximate these derivatives by means of difference quotients. For example, $f'(x_j)$ with $x_j \in \Omega_h$ can be approximated by the forward difference

$$(\partial_x^+ f)(x_j) := \frac{f(x_{j+1}) - f(x_j)}{h}$$

or the backward difference

$$(\partial_x^- f)(x_j) := \frac{f(x_j) - f(x_{j-1})}{h},$$

cf. [12]. These approximations are of first order. A common approximation of $f''(x_j)$ of second order is given by the so-called three-point stencil

$$(\Delta_h f)(x_j) := (\partial_x^- \partial_x^+ f)(x_j) = \frac{f(x_{j+1}) - 2f(x_j) + f(x_{j-1})}{h^2}.$$

In the case of the Fokker-Planck equation (2.7), this is

$$\left(\Delta_h \left(\sigma^2 y\right)\right)(x_j, t) = \frac{\sigma(x_{j+1}, t)^2 y(x_{j+1}, t) - 2\sigma(x_j, t)^2 y(x_j, t) + \sigma(x_{j-1}, t)^2 y(x_{j-1}, t)}{h^2} \tag{4.1}$$

for a fixed time t. While (4.1) is suited for approximating the diffusion term $\partial_{xx}^2(\sigma^2 y)$ in the Fokker-Planck equation, one has to be sensitive about the advection term $\partial_x(b\ y)$, since the conservation of the probability must be preserved, cf. (2.10). To this end, with the help of [29], an approximation that fits these requirements was derived in [3] for time-independent drift functions b and the control u being a real value; see the respective references for details. The resulting formula for a viable approximation of $\partial_x\Big(b(x,u)\ y(x,t)\Big)$ for a fixed time t and a given control u can be written as

$$\begin{aligned}\left(\partial_x^C(b\ y)\right)(x_j, t) := \frac{1}{12h}\Big[&\big(b(x_{j-2}; u) + B\big)y(x_{j-2}, t)\\ &-4\big(2b(x_{j-1}; u) + B\big)y(x_{j-1}, t)\\ &+6B\ y(x_j, t)\\ &+4\big(2b(x_{j+1}; u) - B\big)y(x_{j+1}, t)\\ &-\big(b(x_{j+2}; u) - B\big)y(x_{j+2}, t)\Big],\end{aligned} \tag{4.2}$$

where

$$B := \max_{x\in\Omega} |b(x;u)| \tag{4.3}$$

for given u. In a discretized setting, this translates to

$$B = \max_{x\in\Omega_h} |b(x;u)|.$$

In both (4.1) and (4.2) the homogeneous boundary conditions (2.9), i.e. $y_{|\Sigma} = 0$, are taken into account. Whenever needed, functions are extended by 0 outside the mesh. This is no limitation because of the homogeneous boundary conditions and the vanishing fluxes at the boundaries resulting from the properties of a probability density function, cf. (2.10). For the three relevant stochastic processes, cf. Example 2.6, b does not depend on t directly. However, if u is a function of time, time does play a role. In these cases, (4.2) is extended in the following way:
First of all, once again, the time t is fixed; only the spatial discretization matters for now. B in (4.3) is calculated by

$$B(t) := \max_{x\in\Omega_h} |b(x,t;u(x,t))|.$$

Notice that the values of the control u are known on the grid Ω_h. The approximation of $\partial_x\Big(b(x,t;u(x,t))\ y(x,t)\Big)$ for a fixed time t is then given by

$$\begin{aligned}\big(\partial_x^C(b\ y)\big)(x_j,t) := \frac{1}{12h}\Big[&\big(b(x_{j-2},t;u(x_{j-1},t)) + B(t)\big)y(x_{j-2},t)\\ &-4\big(2b(x_{j-1},t;u(x_{j-1},t)) + B(t)\big)y(x_{j-1},t)\\ &+6B(t)\ y(x_j,t)\\ &+4\big(2b(x_{j+1},t;u(x_{j+1},t)) - B(t)\big)y(x_{j+1},t)\\ &-\big(b(x_{j+2},t;u(x_{j+1},t)) - B(t)\big)y(x_{j+2},t)\Big],\end{aligned} \tag{4.4}$$

In practice, this procedure yields good numerical results, cf. Chapter 6. With (4.1) and (4.4), the spatial derivatives are approximated on Ω_h, i.e.

$$\partial_t y(x_j,t) \approx \frac{1}{2}\Delta_h\left(\sigma(x_j,t)^2 y(x_j,t)\right) - \partial_x^C\Big(b(x_j,t;u(x_j,t))\ y(x_j,t)\Big) \tag{4.5}$$

for $x_j \in \Omega_h$ and fixed t. (4.5) with "=" defines an ODE on every grid point $x_j \in \Omega_h$. Since the solution on Σ is already given, only the inner grid points are of interest. Let these be numbered in ascending order, i.e. $x_1 < ... < x_M$. Define

$$\tilde{y}(t) := (\tilde{y}_1(t), ..., \tilde{y}_M(t))^T := (y(x_1,t), ..., y(x_M,t))^T.$$

Then "(4.5) with '=' for every grid point Ω_h" can be subsumed by the ODE system

$$\dot{\tilde{y}}(t) := \frac{d}{dt}\tilde{y}(t) = A(t)\tilde{y}(t) \tag{4.6}$$

with $A(t)$ defined by

$$A := \frac{1}{12h^2}\begin{pmatrix} -2\alpha_1 - 6hB & \alpha_2 + 4h\delta_2 & -h\epsilon_3 & & \\ \alpha_1 + 4h\gamma_1 & \ddots & \ddots & \ddots & \\ -h\beta_1 & \ddots & \ddots & \ddots & -h\epsilon_M \\ & \ddots & \ddots & \ddots & \alpha_M + 4h\delta_M \\ & & -h\beta_{M-2} & \alpha_{M-1} + 4h\gamma_{M-1} & -2\alpha_M - 6hB \end{pmatrix},$$

where the dependence on t has been omitted and

$$\begin{aligned} \alpha_j(t) &:= 6\sigma(x_j,t)^2, \\ \beta_j(t) &:= B(t) + b(x_j,t;u(x_j,t)), \\ \gamma_j(t) &:= B(t) + 2b(x_j,t;u(x_j,t)), \\ \delta_j(t) &:= B(t) - 2b(x_j,t;u(x_j,t)), \\ \epsilon_j(t) &:= B(t) - b(x_j,t;u(x_j,t)). \end{aligned}$$

The initial condition (2.8) becomes

$$\tilde{y}(t_0) = (\rho(x_1), ..., \rho(x_M))^T \tag{4.7}$$

with $t_0 = 0$. The initial value problem (4.6)-(4.7) can then be solved using standard algorithms, e.g. Runge-Kutta methods, cf. [10]. The most simple is the (explicit) Euler method: Given an initial value problem (4.6)-(4.7) and a time discretization step size $\delta t > 0$, the next step in time $\tilde{y}(t_{k+1})$ is calculated by

$$\tilde{y}(t_{k+1}) = \tilde{y}(t_k) + \delta t A(t_k)\tilde{y}(t_k) \tag{4.8}$$

for $t_k := t_0 + k \cdot \delta t$. This corresponds to using the forward difference approximation of the time derivative, which is only a first order approximation. A second order approximation in time and space is discussed in [4]. However, (4.8) in conjunction with (4.2), both used in [3], yield already good numerical results if δt and h are chosen carefully, cf. Remark 4.1. In particular, the conservation of probability seems to hold for all times.
The scheme can be employed analogously for the adjoint problems (2.49), (2.53), (2.57), and (2.61). With the time transformation described in Remark 2.28, only little adjustments need to be made: Since the coefficient functions σ^2 and b are now factors in front of the respective derivatives, the difference approximations (4.1) and (4.4) become

$$(\Delta_h y)(x_j,t) = \frac{y(x_{j+1},t) - 2y(x_j,t) + y(x_{j-1},t)}{h^2}$$

and

$$\begin{aligned} \left(\partial_x^C y\right)(x_j,t) := \frac{1}{12h}\big[&\big(1+1\big)y(x_{j-2},t) - 4\big(2+1\big)y(x_{j-1},t) + 6y(x_j,t) \\ &+4\big(2-1\big)y(x_{j+1},t) - \big(1-1\big)y(x_{j+2},t)\big] \\ = \frac{1}{12h}&\left[2y(x_{j-2},t) - 12y(x_{j-1},t) + 6y(x_j,t) + 4y(x_{j+1},t)\right]. \end{aligned}$$

Therefore, $A(t)$ in (4.6) reads

$$A := \frac{1}{12h^2}\begin{pmatrix} -2\alpha_1 + 3h\zeta_1 & \alpha_1 + 2h\zeta_1 & & & \\ \alpha_2 - 6h\zeta_2 & \ddots & \ddots & & \\ h\zeta_3 & \ddots & \ddots & \ddots & \\ & \ddots & \ddots & \ddots & \alpha_{M-1} + 2h\zeta_{M-1} \\ & & h\zeta_M & \alpha_M - 6h\zeta_M & -2\alpha_M + 3h\zeta_M \end{pmatrix},$$

where the dependence on t has been omitted and

$$\zeta_j(t) := 2b(x_j, t; u(x_j, t)).$$

In case of (2.53) and (2.61), the respective right-hand side is added to the right-hand side in (4.6). For example, the ODE system in case of (2.53) is given by

$$\dot{\tilde{y}}(t) = A(t)\tilde{y}(t) + (\tilde{y}(t) - \tilde{y}_d(t)),$$

where $\tilde{y}_d(t)$ is defined analogously to $\tilde{y}(t)$.

Remark 4.1
To ensure convergence of the numerical scheme as a whole, the discretization step sizes h and δt have to be chosen carefully. In particular, the ratio of δt to h^2 is important, cf. [27]. For the three stochastic processes considered in this work, the values for h and δt were adapted from [3].

4.2 Nonlinear Optimization Algorithms

Once both the original IBVP and the adjoint problem are solved by means of the algorithms presented in Section 4.1, the variational inequality in the respective optimality systems (2.48)-(2.50), (2.52)-(2.54), (2.56)-(2.58), and (2.60)-(2.62) can be evaluated. With the help of Lemma 2.29, each variational inequality can be replaced by its point-wise equivalent. Then, in order to verify the inequality, the gradient of the reduced cost functional, $\nabla\hat{J}(u)$, is evaluated (point-wise). This leads to common nonlinear optimization techniques. A simple method that is easy to implement is the Gradient, or Steepest Descent method. Starting from an initial value u_0, the next iterative is computed using the negative gradient. Due to restrictions on u, i.e. $u \in U_{ad}$, a projection onto U_{ad} is needed, leading to the Projected Gradient method:

Algorithm 4.2 (Projected Gradient)
Let u_0 with $u_a \leq u_0 \leq u_b$ and set $k := 0$. While some stopping criteria are violated (see below), do

1. *Calculate the state y_k for the control u_k.*
2. *Calculate the corresponding adjoint state p_k.*

3. *Calculate the search direction* $d_k := -\nabla \hat{J}(u_k)$ *(using* y_k, p_k*).*
4. *(Optional) Obtain a step length* s_k *from one of the linesearch methods below. Alternatively, set* $s_k := 1$*.*
5. *Set* $u_{k+1} := \mathbb{P}_{[u_a,u_b]}\{u_k + s_k d_k\}$ *and set* $k := k + 1$*.*

End while.

Remark 4.3
Even though step 4 in Algorithm 4.2 is marked as optional, a linesearch needs to be performed when trying to prove convergence of the algorithm, see below. However, in practice, this comes at the cost of computation time. Therefore, in the implementation, the user decides whether to perform a linesearch or not.

There are several possible stopping criteria and methods to obtain a suitable step length. Typical stopping criteria include (a) a maximum number of steps, i.e.

$$k \geq k_{max} \tag{4.9}$$

and (b) a vanishing gradient, i.e.

$$\|\nabla \hat{J}(u_k)\|_U = 0.$$

In practice, the latter condition becomes

$$\|\nabla \hat{J}(u_k)\|_U < \varepsilon \tag{4.10}$$

with a small $\varepsilon > 0$. However, one cannot expect the gradient to vanish in general because of the control constraints. To deal with active control constraints, the following stopping criterion was introduced in [15, Ch. 9]:

$$\left\|u_k - \mathbb{P}_{[u_a,u_b]}\{u_k + d_k\}\right\|_U \leq \tau_r \left\|u_0 - \mathbb{P}_{[u_a,u_b]}\{u_0 + d_0\}\right\|_U + \tau_a \tag{4.11}$$

with $0 < \tau_r < 1$ and $0 < \tau_a \ll 1$.
Finding a suitable rule to determine a "good" step length is no easy task and requires numerical experience. The purpose of the step length is to not increase u_k too much and miss the optimum, but at the same time, to not choose a step length so little that no progress is visible. The Armijo condition

$$\hat{J}(u_k + s_k d_k) \leq \hat{J}(u_k) + \delta\ s_k \left(\nabla \hat{J}(u_k), d_k\right)_U \tag{4.12}$$

together with the Wolfe condition

$$\left(\nabla \hat{J}(u_k + s_k d_k), d_k\right)_U > \sigma \left(\nabla \hat{J}(u_k), d_k\right)_U, \tag{4.13}$$

where $0 < \delta < \sigma < \frac{1}{2}$, ensure convergence of the Steepest Descent method if, for given $\alpha \in]0,1[$, $s_k \in \left\{\alpha^l \mid l \in \mathbb{N}_0\right\}$ is the largest step length that satisfies (4.12) and (4.13), cf. [6, p. 43].[1]

[1] One does not necessarily need to choose the largest step length. However, a "substantial" reduction in the objective function must be achieved, cf. [22] for details.

Another possible condition used in [15, Ch. 9] is given by

$$J(\mathbb{P}_{[u_a,u_b]}\{u_k + s_k d_k\}) - J(u_k) \leq -\frac{\tilde{\sigma}}{s_k}\left\|u_k - \mathbb{P}_{[u_a,u_b]}\{u_k + s_k d_k\}\right\|_U^2, \tag{4.14}$$

where $\tilde{\sigma} \in]0,1[$. As above, the condition is subsequently tested for $s_k := \alpha^l$ for a chosen $\alpha \in]0,1[$ with increasing $l \in \mathbb{N}_0$ until (4.14) is satisfied.
A typical feature of the Projected Gradient method is its good progress in reducing the cost functional in the initial iterations, while slowing down significantly in later iterations [13]. To counteract this, another method is introduced – the (semi-smooth) Newton method with primal-dual active set strategy. This technique was presented e.g. in [13] and [17] and works as follows. First, rewrite

$$\left(\nabla\hat{J}(\bar{u}), u - \bar{u}\right)_U$$

in (2.69) using Lagrange multipliers in the following, equivalent form:

$$\nabla\hat{J}(\bar{u}) + \mu_b - \mu_a = 0 \tag{4.15a}$$

$$\begin{aligned} \mu_a \geq 0, \quad u_a - \bar{u} \leq 0, \quad \mu_a^T(u_a - \bar{u}) = 0 \\ \mu_b \geq 0, \quad \bar{u} - u_b \leq 0, \quad \mu_b^T(\bar{u} - u_b) = 0 \end{aligned} \tag{4.15b}$$

Just as the variational inequality, (4.15) is understood in a point-wise almost everywhere sense in the domain $\mathcal{E}$ where the control is tracked, e.g. the space-time cylinder $\mathcal{E} := Q$ in the case of (2.60)-(2.62). (4.15) can be reformulated by introducing $\mu := \mu_b - \mu_a$ and merging (4.15b) into one equation:

$$\nabla\hat{J}(\bar{u}) + \mu = 0 \tag{4.16a}$$

$$\max\{0, \mu + c\,(\bar{u} - u_b)\} + \min\{0, \mu + c\,(\bar{u} - u_a)\} - \mu = 0 \tag{4.16b}$$

with $c > 0$. Again, min and max are understood in a point-wise sense. The constant c does not change the fact that (4.16) is equivalent to (4.15), cf. [13]. However, the introduced constant is very beneficial: Equation (4.16b) is not differentiable in the classical sense. Therefore, classical Newton methods are theoretically inapplicable to solve (4.16). However, (4.16b) has the so-called Newton differentiability property for $c = \lambda$, where λ is the regularization parameter in the cost functional of the optimal control problem. Hence, one can apply a so-called semi-smooth Newton method. To this end, the Jacobian of (4.16) is needed. It depends on the subsets of $\mathcal{E}$ on which the max or min in (4.16b) is attained [13]. These subsets form the so-called active and inactive sets. At an iterate (u_k, μ_k) they are defined by

$$\begin{aligned} \mathcal{A}_k^+ &:= \{x \in \mathcal{E} \mid \mu_k + c\,(u_k - u_b) > 0\}, \quad \mathcal{A}_k := \mathcal{A}_k^+ \cup \mathcal{A}_k^-, \\ \mathcal{A}_k^- &:= \{x \in \mathcal{E} \mid \mu_k + c\,(u_k - u_a) < 0\}, \quad \mathcal{I}_k := \mathcal{E} \setminus \mathcal{A}_k. \end{aligned}$$

The Newton direction $(\delta u, \delta\mu)^T$ for (4.16b) at (u_k, μ_k) is then given by

$$\begin{pmatrix} \nabla^2\hat{J}(u_k) & I \\ c\,\mathbb{1}_{\mathcal{A}_k} & -\mathbb{1}_{\mathcal{I}_k} \end{pmatrix} \begin{pmatrix} \delta u \\ \delta\mu \end{pmatrix} = -\begin{pmatrix} \nabla\hat{J}(u_k) + \mu_k \\ c\,\mathbb{1}_{\mathcal{A}_k^+}(u_k - u_b) + c\,\mathbb{1}_{\mathcal{A}_k^-}(u_k - u_a) - \mathbb{1}_{\mathcal{I}_k}\mu_k \end{pmatrix},$$

which can be reformulated into a symmetric form

$$\begin{pmatrix} \nabla^2 \hat{J}(u_k) & \mathbb{1}_{\mathcal{A}_k} \\ \mathbb{1}_{\mathcal{A}_k} & 0 \end{pmatrix} \begin{pmatrix} \delta u \\ \delta \mu_{|\mathcal{A}_k} \end{pmatrix} = - \begin{pmatrix} \nabla \hat{J}(u_k) + \mu_k \\ \mathbb{1}_{\mathcal{A}_k^+}(u_k - u_b) + \mathbb{1}_{\mathcal{A}_k^-}(u_k - u_a) \end{pmatrix}. \tag{4.17}$$

In the latter formulation, μ_{k+1} is set to zero on $\mathcal{I}_k$. The system (4.17) is then solved iteratively, yielding the new search direction. This results in the following algorithm.

Algorithm 4.4 (Newton method with primal-dual active set strategy)
Let u_0 be the initial guess and let μ_0. Set $k := 0$. While some stopping criteria are violated (see below), do

1. *Calculate the state y_k for the control u_k.*
2. *Calculate the corresponding adjoint state p_k.*
3. *Evaluate the reduced gradient $\nabla \hat{J}(u_k)$ (using y_k, p_k).*
4. *Compute the active sets $\mathcal{A}_k$.*
5. *Solve the Newton system* (4.17) *iteratively.*
6. *(Optional) Obtain a step length s_k from one of the above linesearch methods. Alternatively, set $s_k := 1$.*
7. *Set*

$$\begin{aligned} u_{k+1} &:= u_k + s_k\ \delta u, \\ \mu_{k+1|\mathcal{A}_k} &:= u_{k|\mathcal{A}_k} + s_k\ \delta u_{|\mathcal{A}_k}, \\ \mu_{k+1|\mathcal{I}_k} &:= 0, \\ k &:= k + 1. \end{aligned}$$

End while.

A reasonable stopping criterion is

$$\left\| - \begin{pmatrix} \nabla \hat{J}(u_k) + \mu_k \\ \mathbb{1}_{\mathcal{A}_k^+}(u_k - u_b) + \mathbb{1}_{\mathcal{A}_k^-}(u_k - u_a) \end{pmatrix} \right\|_U < \varepsilon, \tag{4.18}$$

which is used in combination with a maximum number of steps (4.9) in the implementation.

Remark 4.5

(a) Often one tries to avoid the computation of the Hessian $\nabla^2 \hat{J}(u_k)$. Instead, it is approximated to reduce the computation time. This leads to (semi-smooth) Quasi-Newton methods, one of which is the so-called Broyden-Fletcher-Goldfarb-Shanno (BFGS) method. For details regarding the BFGS algorithm, see [22].

(b) In Algorithm 4.4, the linesearch is really not needed if the Hessian $\nabla^2 \hat{J}(u_k)$ is known. However, when employing the BFGS algorithm, linesearch methods do play a role. Therefore, in the implementation, the decision whether to perform a linesearch is delegated to the user.

(c) Newton methods usually only converge locally, but offer a higher convergence rate than e.g. the Projected Gradient method. Both techniques can be combined in order to exploit the benefits of each algorithm. This is done by first employing the Projected Gradient method and after some iterations, switch to the (Quasi-)Newton approach, e.g. the BFGS method.

(d) It has not been specified how to solve the linear system (4.17)*. To this end, standard techniques like the Conjugated Gradient (CG) method are employed, cf. [14].*

Remark 4.6

In the algorithms above, information about the gradient is required. An efficient way to obtain this information is by solving two PDEs, the original IBVP and the corresponding adjoint problem, as stated in the respective algorithms. However, it is also possible to compute the gradient by analyzing the effects of small disturbances in each component of u. For u being a real value, i.e. if u consists of only one component, the gradient is approximated as follows:

$$\nabla \hat{J}(u) \approx \frac{J(u) - J(u - h_u)}{h_u}. \tag{4.19}$$

Since this approach entails solving two original IBVPs for each component of u, it is very inefficient, especially if u is space- (and time-)dependent. Even so, it does not require knowledge of the adjoint problem. Therefore, it can be used to test whether the adjoint problem is indeed correct. Also, one could always resort to it at the expense of computing time.

Chapter 5

PDE-MPC

The program "PDE-MPC" is the practical centerpiece of this thesis. It is written in C++ and is used to solve the Fokker-Planck optimal control problems introduced in Section 2.3 numerically by means of Model Predictive Control, cf. Chapter 3 and Section 3.3 in particular. Due to its modular structure, it is not restricted to these problems. The motivation behind choosing such a structure was to provide a convenient way to expand the program in several manners. For instance, PDEs other than the Fokker-Planck equation might be of interest. In addition, different algorithms might be more in line with certain problems. Therefore, one should not be confined to the already implemented algorithms.
In this chapter, the emphasis is on the structure of "PDE-MPC" and on the program flow. If one is interested in a description of every single method and its arguments, a documentation is provided along with the source code on the enclosed CD-ROM.

5.1 Basic Structure and Program Flow

The program flow already became apparent from the previous chapters: Employing the MPC Algorithm 3.2, one needs to solve the optimal control problem (OCP_N) on a certain time interval, implement the first value of the calculated optimal control sequence, move the considered time interval, solve (OCP_N) on this new interval, and so on, until a terminal time T_E is reached. The time interval is determined by the horizon parameter N, which the user must choose. If N is not the shortest possible horizon, (desired) states after T_E might be considered in the computation of an optimal control sequence, cf. Remark 3.12. In this case, N is still left constant. The MPC algorithm is implemented in the `MPC` class. Note that since (OCP_N) has not been specified yet, the implementation is independent of the optimal control problem and the PDE in particular.
To be able to deal with general PDE-constrained optimal control problems, an abstract `Optimizer` object is used as a placeholder. To this end, `ProjectedGradient` and `BFGS` are available, both subclasses of `Optimizer`. As the name suggests, `ProjectedGradient` is a realization of the Projected Gradient method, cf. Algorithm 4.2. The stopping criteria used are: (4.9) in conjunction with (4.10) if no control constraints are active and (4.9) in

conjunction with (4.11) if some control constraints are. `BFGS` consists of the BFGS method, a Quasi-Newton method based on Algorithm 4.4, where the Hessian is only approximated, cf. Remark 4.5(a). In this case, the stopping criterion is determined by (4.18) and a maximum number of steps (4.9). The two linesearch strategies (4.12)-(4.13) and (4.14) are implemented in both classes. For practical reasons, a maximum number of linesearch steps is added. It is up to the user whether to perform the linesearch or not. Furthermore, it is possible to mix the two algorithms, cf. Remark 4.5(c).
The abstract `Optimizer` class allows to select a nonlinear optimization algorithm at run time. Different as the suitable algorithms may be, most of them, if not all, require information about the gradient. To this end, the derivation of the optimality systems in Subsection 2.3.2 illustrated two things: One, the gradient of the reduced cost functional, $\nabla \hat{J}(u)$, depends not only on the cost functional, but also on the PDE at hand. Two, once the formula is known, one needs to solve the original IBVP and the corresponding adjoint problem in order to evaluate it. Consequently, the optimizer has to delegate the evaluation of the gradient. First, the formula for $\nabla \hat{J}(u)$ needs to be known. This information is retrieved from the abstract `Model` class, which is introduced below. Second, the two PDEs are to be solved.
An alternative approach to obtain information about the gradient has been introduced in Remark 4.6: Instead of using the formula for $\nabla \hat{J}(u)$ and solving the adjoint problem, one could simply analyze the effects of small disturbances in each component of the control, cf. Remark 4.6 and (4.19). This resort has been implemented in both the `ProjectedGradient` and the `BFGS` classes. In this case, two original IBVPs need to be solved for each component of the control. Either way, solutions of PDEs need to be obtained.
One possible approach to solve these PDEs has been introduced in Section 4.1. The idea is to only discretize the spatial domain, e.g. by Finite Differences, while the time is kept continuous at first. This approach results in a (possibly large) system of ordinary differential equations. Assuming the spatial discretization for a given time t is known, it remains to solve the ODE system. To this end, `OdeSolver`, another abstract class, is introduced. Each realization thereof consists of a method to solve such a system. The explicit Euler method (4.8) and the implicit Euler method, cf. [10], are implemented in the `ExplEuler` and `ImplEuler` classes, respectively.
In contrast, the discretization in space is handled by the `Model` class. As the name suggests, the `Model` class consists of a certain model, including the PDE at hand, e.g. the Fokker-Planck equation. Since the spatial discretization depends on the PDE and the considered spatial domain, it makes sense to move this job to the `Model` class. In addition, all information regarding (MPC) cost functionals, possible control and state constraints and – as mentioned above – the form of the gradient of the reduced cost functional are expected to be found in the `Model` class or rather in its subclasses. Currently, only the Fokker-Planck optimal control problems considered in this work are implemented in the `FokkerPlanck1D` class, a subclass of `Model`. Here, the cost functional is determined by the relevant terms in either one of the stage costs (3.6) or (3.8). The resulting cost functionals are given by (3.9) for space-independent control and (3.10) in the case of space-dependent control. As stated in Section 3.3, the resulting optimal control problems of type (OCP_N) coincide or at least approximate the optimal control problems (2.20)-(2.23), (2.24)-(2.28), (2.29)-(2.33), and (2.34)-(2.38).

Thus, the corresponding optimality systems derived in Subsection 2.3.2 are solved and the related reduced gradients (2.70)-(2.73) are used in the above optimization algorithms. The derivatives in these gradients are approximated by either a forward or a backward difference. All occurring L^2 norms and all integrals in general are computed by trapezoidal formulas.
Furthermore, depending on the chosen algorithms, it might be necessary to solve a linear system, e.g. the Newton system (4.17). Also, a linear system needs to be solved if the implicit Euler method is used. To this end, direct and iterative solvers can be implemented as derived classes of the abstract `LesSolver` class. For now, only the Conjugate Gradient (CG) Method is available, cf. [14]. It is implemented in the `CGMethod` class. A preconditioned variant of the CG Method, `PCGMethod`, has been added for testing purposes. Given a linear system $Ax = b$ with $A \in \mathbb{R}^{n \times n}$, the preconditioner matrix equals the diagonal matrix D^{-1} with $D_{jj} = A_{jj}, j = 1, ..., n$.
Lastly, for the sake of convenience, the `FPMPCConfig` class encapsules all parameters and settings that are needed by the currently implemented classes, e.g. model parameters, MPC settings, discretization step sizes, error tolerances, and many more. The advantage is having all (relevant) parameters in one place. The downside is the loss of consistency regarding the modular structure. However, since the `FPMPCConfig` class is not required at all in order for the other classes to function, its presence does not influence the expandability of the program.
The program generates several files in which for example the state y, the control u, and the adjoint state p are stored. For more details, see Table 5.1.
Figure 5.1 depicts all currently existing classes and their subclasses in the program on the left. It also illustrates the program flow described above. On the right, one can find the main purpose(s) of the respective classes. The `FPMPCConfig` class is separated since it is a special auxiliary class that is not necessarily needed.

Filename	Description
adjTrajectory.txt	Contains the values of the adjoint state p on the mesh.
config.txt	Stores the relevant configuration parameters for further use.
config-detailed.txt	Stores the relevant configuration parameters in a more readable form.
control.txt	Contains the control values u on the mesh.
controlcosts.txt	Stores the control costs, which are a part of the stage costs, for each implemented MPC step.
stagecosts.txt	Stores the stage costs for each implemented MPC step.
statecosts.txt	Stores the state costs, which are a part of the stage costs, for each implemented MPC step.
totalcosts.txt	Contains the (weighted) sum of all implemented state-, control-, and stage costs.
trajectory.txt	Contains the values of the state y on the mesh.
xCoords.txt	Contains the spatial discretization of Ω.

Table 5.1: Files generated by PDE-MPC.

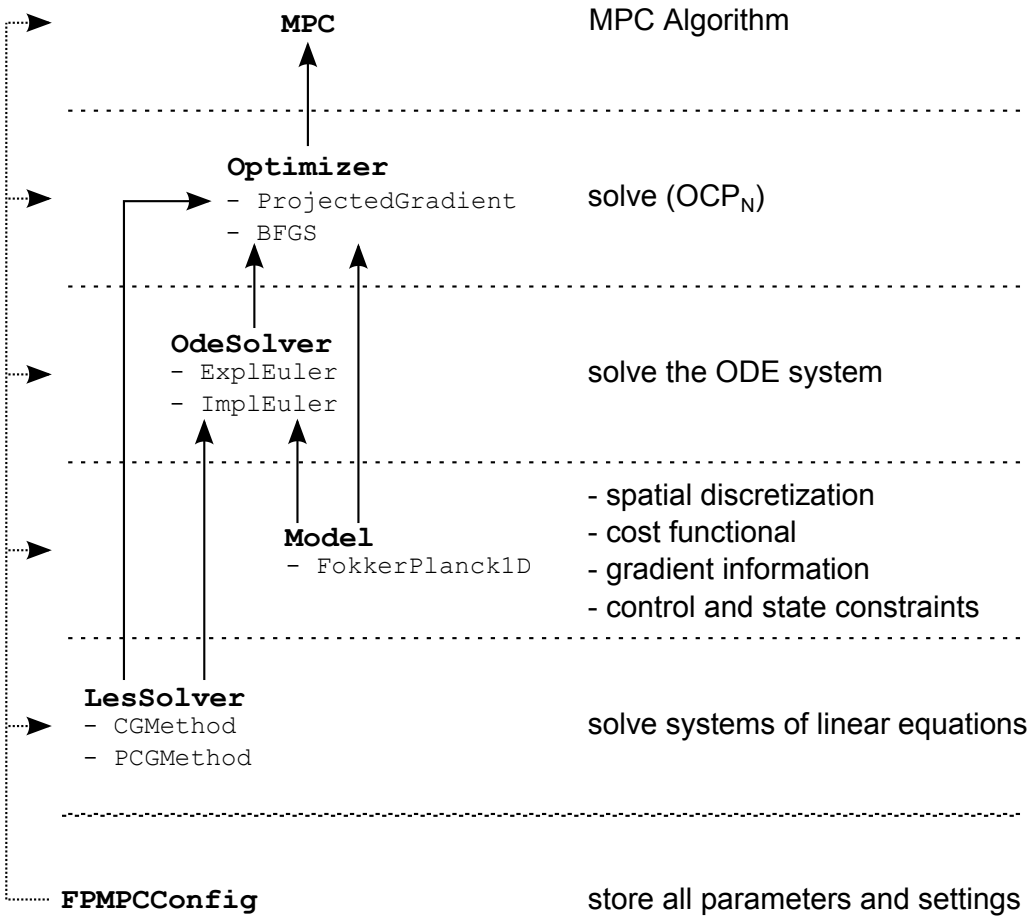


Figure 5.1: Program structure

Chapter 6

Numerical Results

In this chapter, numerical simulations are performed to answer the questions posed at the very beginning of this thesis: First, what impact does the MPC horizon have on the control of the Fokker-Planck equation? And second, how does a space-dependent control affect the results?
As explained in Chapter 2, the Fokker-Planck equation is used to model the evolution of a probability density function of certain stochastic processes. The three stochastic processes considered in this work are the same processes that are considered in [3]. They were introduced in Example 2.6: the Ornstein-Uhlenbeck Process with Additive Control, the Geometric-Brownian Process with Additive Drift Control, and the Shiryaev Process. Since the results presented in [3] motivated this analysis, a comparison of the numerical results in this work to those in [3] is self-evident. To this end, in a first step, the numerical results in [3] are reproduced in Section 6.1. The optimal control problem considered in [3] is given by (2.20)-(2.23). This corresponds to employing an MPC scheme with the cost functional (3.9) and $N = 1$, cf. Remark 3.13.[1] In a next step, the effects of increasing N are analyzed for space-independent control. Afterwards, space-dependent control is treated in Section 6.2.
All numerical results were obtained by the program "PDE-MPC" introduced in Chapter 5. There are some parameters that do not change throughout this chapter, unless stated otherwise. These are summarized in Table 6.1. The parameter tol_{opt} is the error tolerance in the optimization algorithms, i.e. the ε in (4.10) and (4.18). Unless stated otherwise, the gradient $\nabla\hat{J}(u)$ is computed by solving the original IBVP and, subsequently, the adjoint problem. If the gradient is computed by analyzing small disturbances in the control instead, then h_u is the corresponding "step size", cf. (4.19). The parameters τ_r and τ_a appear in (4.11).

6.1 Space-independent Control

As the title suggests, space-independent control of the Fokker-Planck equation is tackled in this section. The MPC cost functional is given by (3.9) in this case.

[1] Note that the notation suggested in Remark 3.13 is used in this chapter.

Parameter	Value
tol_{opt}	1e-5
tol_{les}	1e-12
h_u	1e-6
τ_r	1e-5
τ_a	1e-5

Table 6.1: Standard parameter values for the numerical simulations.

6.1.1 The Ornstein-Uhlenbeck Process with Additive Control

The Fokker-Planck equation describing the evolution of the probability density function of the Ornstein-Uhlenbeck process with additive control is given by

$$\begin{aligned} 0 &= \partial_t y(x,t) - \frac{1}{2}\partial_{xx}^2\left(\tilde{\sigma}^2 y(x,t)\right) + \partial_x\big((-\mu x + u)\; y(x,t)\big) \\ &= \partial_t y(x,t) - \frac{1}{2}\tilde{\sigma}^2\partial_{xx}^2 y(x,t) + \partial_x\big((-\mu x + u)\; y(x,t)\big) \end{aligned} \tag{6.1}$$

with $\mu, \tilde{\sigma} > 0$, cf. Example 2.6. If the MPC horizon N is the shortest possible, i.e. $N = 1$, then the control u is a real value in each MPC optimization step. For $N > 1$, the control is a function of time, i.e. $u = u(t)$.[2] Table 6.2 depicts the standard parameter values for the Ornstein-Uhlenbeck process. Unless specifically mentioned, these values were used in the numerical simulations.

At first, N is set to 1 to see whether the results in [3] can be reproduced. This is of particular interest since the results in [3] were obtained without solving the corresponding optimality system. Instead, a bisection minimization scheme was employed. Due to u being a simple real value, this method is viable and does not require the knowledge of the gradient of the (reduced) cost functional or knowledge of the adjoint problem. In this work, however, all results were computed using nonlinear optimization algorithms, cf. Chapters 4 and 5. Cases in which $\nabla\hat{J}(u)$ is obtained by solving the original IBVP and, subsequently, the adjoint problem, serve as a test whether the derived optimality system is correct.

Figure 6.1 depicts the computed probability density function y, the initial PDF $y_0 = \rho$, and the desired PDF y_d for $N = 1$. The corresponding optimal control function u is illustrated in Figure 6.2. The adjoint state p is shown in Figure 6.3. Notice that by receding the horizon, technically, the adjoint state is only defined on time intervals $[t_n, t_n + T_s]$. Since on each time interval, there is a new adjoint problem, there are at least two possible values for $p(\cdot, t_n)$ as n increases. In Figure 6.3 and in all following figures that show the adjoint state, only the state that was used to compute $u^*(0)$ in the MPC algorithm is illustrated, cf. Algorithm 3.2. Detailed results can be found in Table 6.3.

While the state y follows the desired state y_d in time fairly well, it never reaches the desired height. Instead, it is much more diffused at all times. The initial PDF and the desired PDF

[2] The MPC algorithm yields an optimal control sequence, which can be interpreted as a piecewise constant function in time.

Parameter	Value
μ	1
$\tilde{\sigma}$	0.8
Ω	]-5, 5[
h	0.05
δt	1.6e-3
u_a	-4
u_b	4
λ	0.1
T_E	5
T_s	0.5
$\rho(x)$	$\frac{1}{\sqrt{2\pi \cdot 0.1^2}} \exp\left(-\frac{x^2}{2 \cdot 0.1^2}\right)$
$y_d(x,t)$	$\frac{\exp\left(-\frac{[x - 2\sin(\pi t/5)]^2}{2 \cdot 0.2^2}\right)}{\sqrt{2\pi \cdot 0.2^2}}$

Table 6.2: Standard parameter values for the Ornstein-Uhlenbeck process with additive control.

are both Gaussian distributions. The computed PDF looks like a Gaussian distribution as well, only with a higher variance. Since only images of the numerical results are provided in [3], only qualitative statements can be made. Comparing the state and the control figures, it seems that the solutions coincide. Therefore, it appears the optimality system (2.48)-(2.50) is correct and solving it leads to a (local) minimum.

t	u	$\frac{1}{2}\|y(\cdot, t+T_s) - y_d(\cdot, t+T_s)\|^2_{L^2(\Omega)}$	$\frac{\lambda}{2}\|u\|^2$
0	1.28441	0.231499	0.082486
0.5	1.67205	0.312326	0.139788
1	1.91661	0.352848	0.183669
1.5	1.99161	0.368928	0.198325
2	1.88367	0.362372	0.177412
2.5	1.5885	0.340495	0.126167
3	1.12412	0.315046	0.0631827
3.5	0.538068	0.296092	0.0144758
4	-0.103586	0.289337	0.000536506
4.5	-0.729877	0.296354	0.026636

Table 6.3: Optimal control of the Ornstein-Uhlenbeck process for $N = 1$: Detailed results.

In a next step, the horizon N is to 2. The main difference is that now, the optimality system

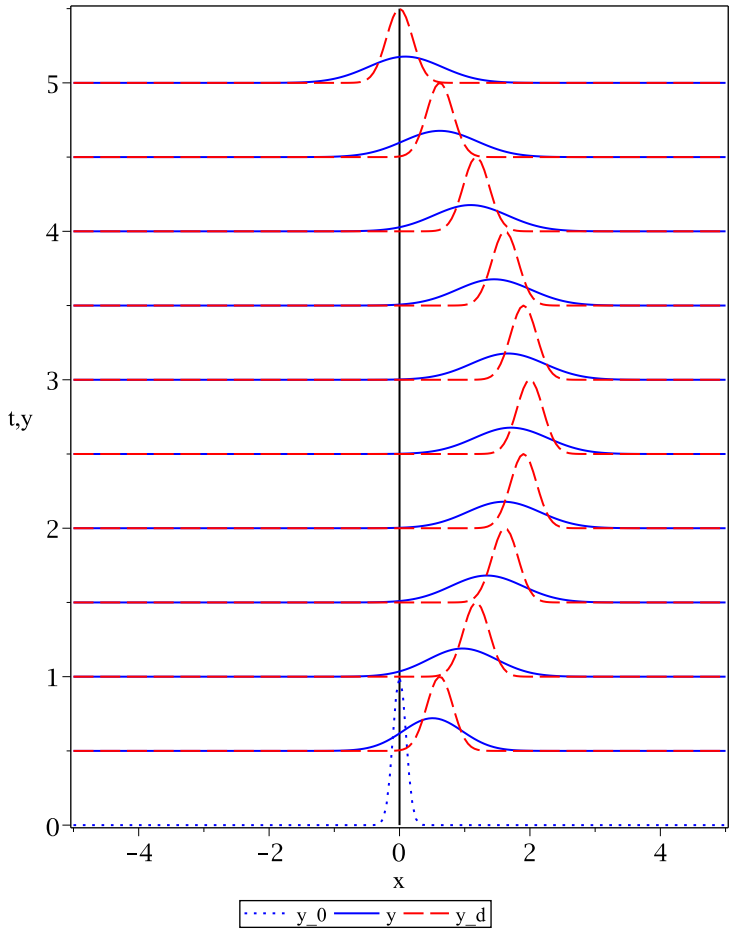


Figure 6.1: Optimal control of the Ornstein-Uhlenbeck process for $N = 1$: The computed PDF y, the initial PDF y_0, and the desired PDF y_d.

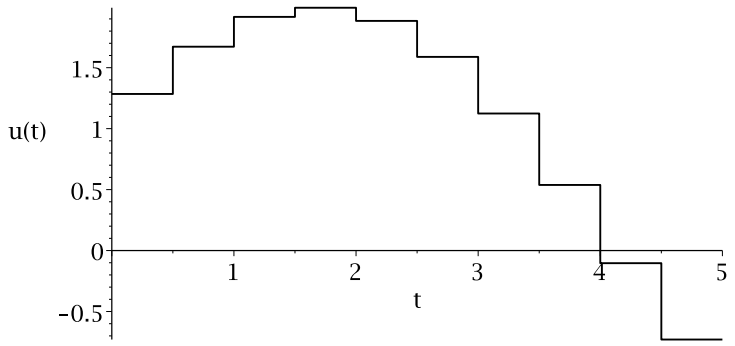


Figure 6.2: Optimal control of the Ornstein-Uhlenbeck process for $N = 1$: The optimal control function.

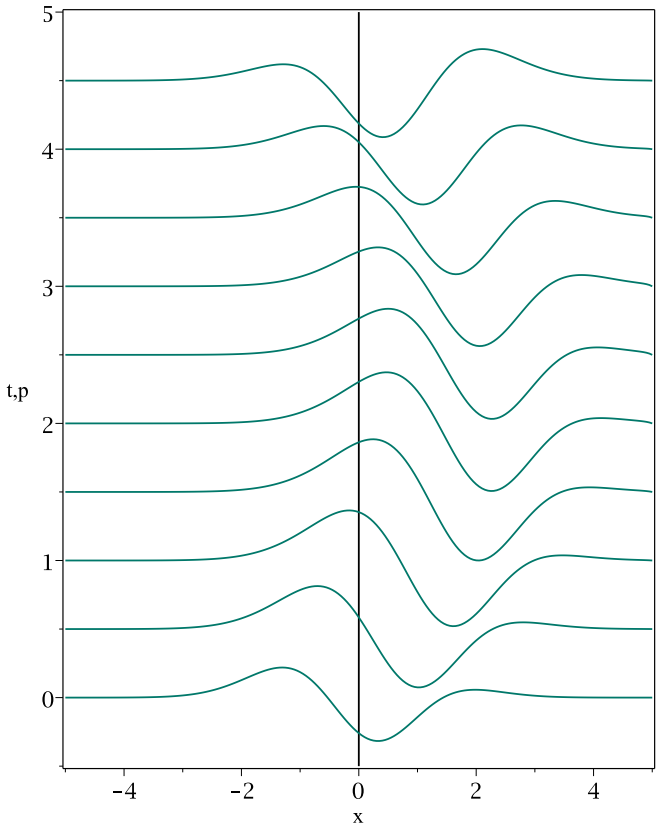


Figure 6.3: Optimal control of the Ornstein-Uhlenbeck process for $N = 1$: The adjoint state p ($\max |p(x,t)| = 0.320421$).

(2.52)-(2.54) has to be solved. Figures 6.4-6.6 show the computed PDF y, the initial PDF y_0, the desired PDF y_d, the computed optimal control function u and the adjoint state p.

t	u	$\frac{1}{2}\|y(\cdot,t+T_s) - y_d(\cdot,t+T_s)\|^2_{L^2(\Omega)}$	$\frac{\lambda}{2}\|u\|^2$
0	2.08881	0.278171	0.218156
0.5	1.7981	0.266235	0.161657
1	2.12477	0.283338	0.225732
1.5	2.1403	0.297415	0.229044
2	1.93228	0.309518	0.186685
2.5	1.51752	0.319868	0.115143
3	0.950163	0.326336	0.0451404
3.5	0.297442	0.325674	0.00442359
4	-0.373778	0.316994	0.0069855
4.5	-1.00231	0.303888	0.0502309

Table 6.4: Optimal control of the Ornstein-Uhlenbeck process for $N = 2$: Detailed results.

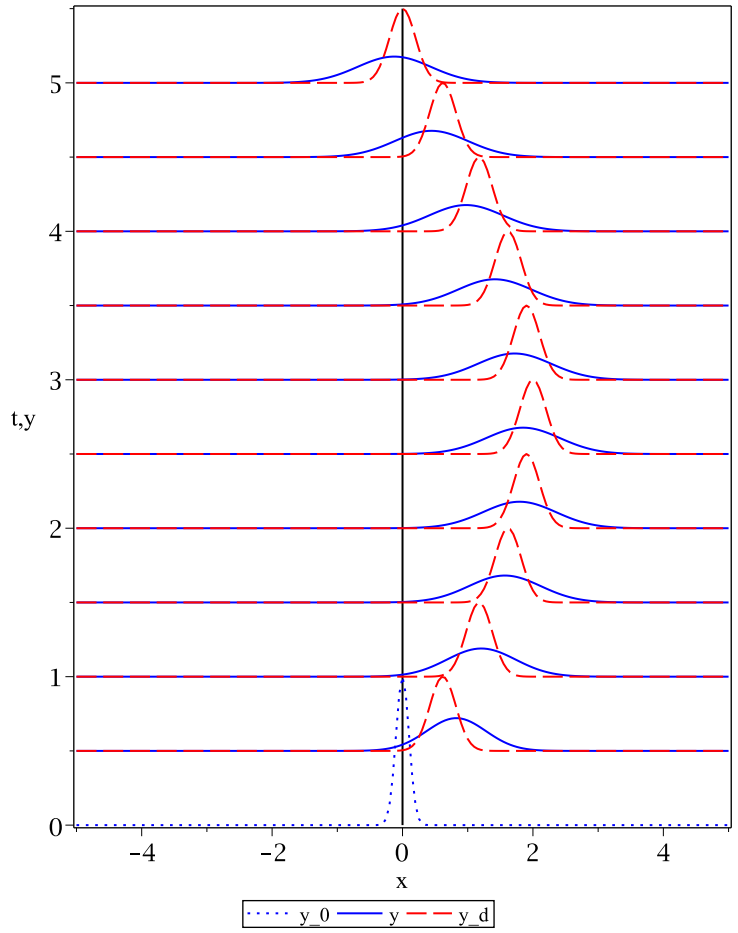


Figure 6.4: Optimal control of the Ornstein-Uhlenbeck process for $N = 2$: The computed PDF y, the initial PDF y_0, and the desired PDF y_d.

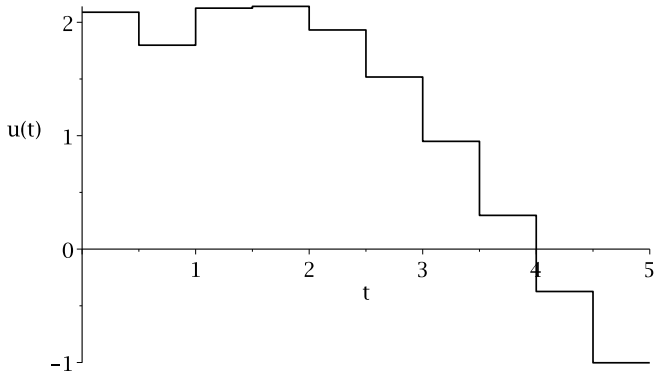


Figure 6.5: Optimal control of the Ornstein-Uhlenbeck process for $N = 2$: The optimal control function.

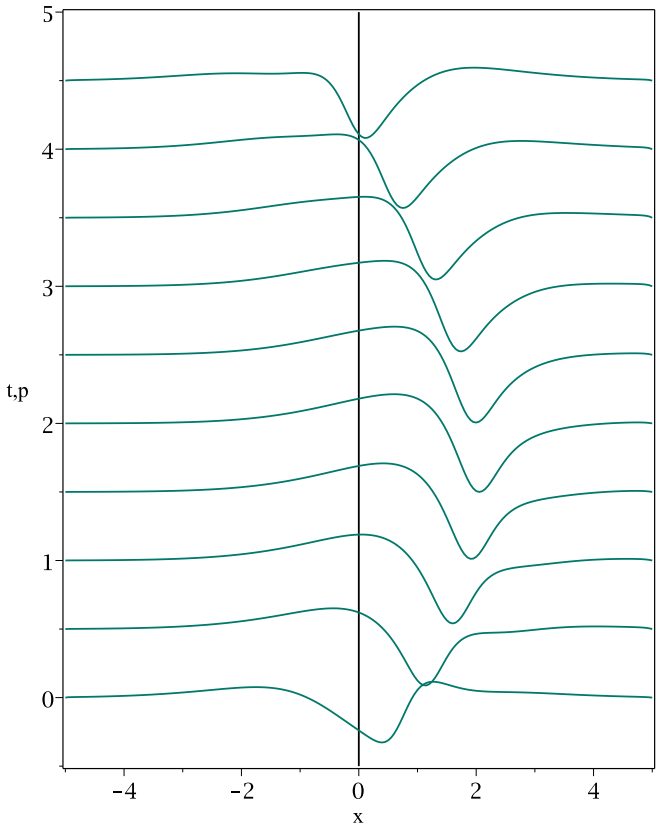


Figure 6.6: Optimal control of the Ornstein-Uhlenbeck process for $N = 2$: The adjoint state p ($\max |p(x,t)| = 0.360579$).

Comparing Figure 6.5 to Figure 6.2, one finds that the computed optimal controls look different in the first few steps. However, this does not lead to better results. On the contrary, comparing the norms in Table 6.4 to those in Table 6.3 shows that the shorter horizon $N = 1$ yields a much better performance. This seems surprising at first, as one usually expects better results with increasing N. The problem is probably twofold: The first problem is the discrepancy between the cost functional assumed in the optimality system and the actual cost functional. As mentioned in Section 3.3, the MPC cost functional is a good approximation of the one in the optimality system for small values of T_s. A value of 0.5 seems too high. The second problem lies in the implementation of the optimality system: In the adjoint PDE in (2.53), $y - y_d$ appears on the right-hand side. Since the PDE is defined in Q, the right-hand side needs to be updated as often as the left-hand side in the numerical scheme, e.g. every δt units in time. However, it is only updated every T_s units in time for technical reasons. Since $T_s = 0.5$ is a relatively high value, one cannot expect a good approximation.
A remedy is to compute the gradient directly by analyzing the effects of perturbed controls, cf. Remark 4.6. Doing so yields better results compared to solving the optimality system, cf. Table 6.5. In that table, "adj." indicates that the adjoint problem and the optimality

system were used, while "num." stands for the alternative computation of the gradient. In order to compare the results, the total costs are evaluated. They are given by the sum over all stage costs of actually measured states and implemented controls. The total costs are then weighted by T_s. This renders possible a comparison of different numerical simulations with different sampling rates T_s. The weighted total costs are denoted by $C_{total,w}$.

N	Computation of the gradient	$C_{total,w}$
1	adj.	2.08899
2	adj.	2.13532
2	num.	2.03251
5	adj.	2.16759
5	num.	2.03337

Table 6.5: Optimal control of the Ornstein-Uhlenbeck process: Weighted total costs for various N.

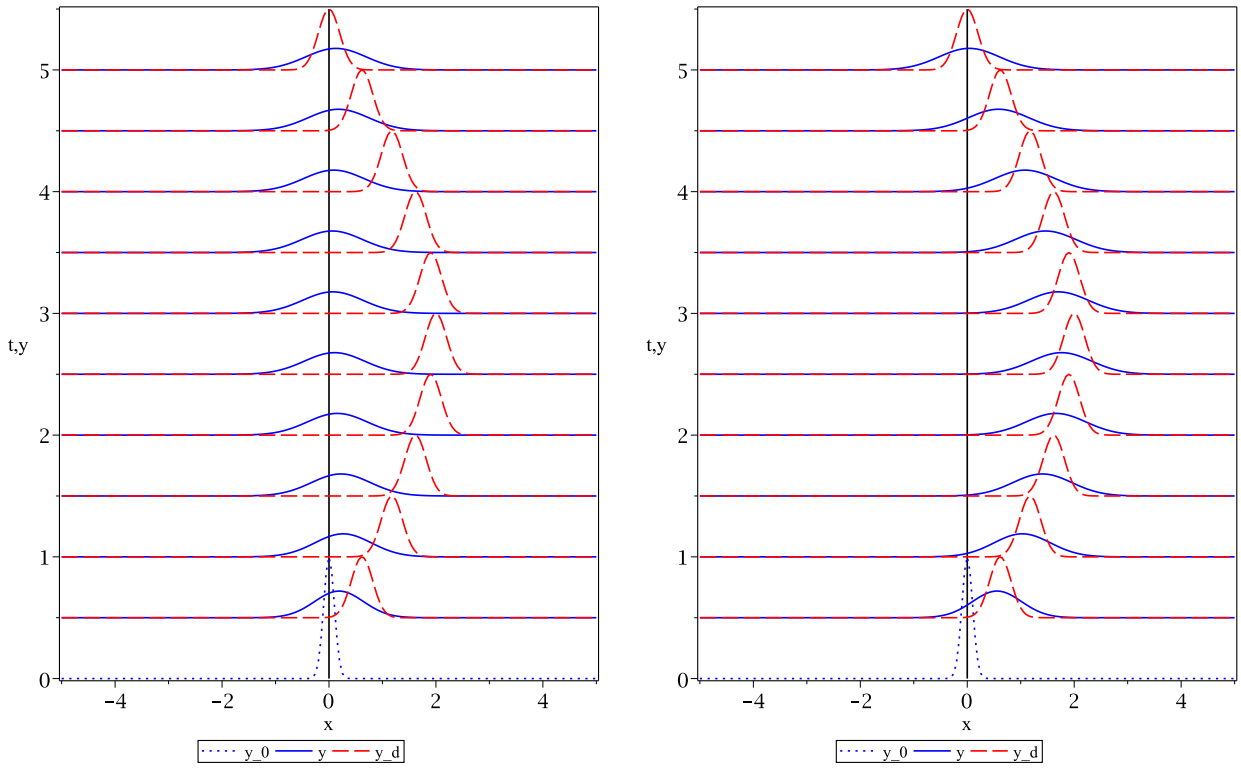
The slight increase in the costs from $N = 2$ to $N = 5$ is plausible, since desired states after the terminal time T_E are taken into account when computing the optimal control sequence, cf. Remark 3.12. Again, the effect is noticeable since the value of T_s is high.
Lowering T_s should mitigate the effects of a bad approximation of the cost functional and the effects of considering the (desired) states after T_E. The following results show that this expectation is justified. Table 6.6 depicts the weighted total costs for various combinations of sampling rates T_s and horizons N.

T_s	N	Computation of the gradient	$C_{total,w}$
0.05	1	adj.	3.71696
0.05	1	num.	3.72166
0.05	10	adj.	1.99609
0.05	10	num.	1.99282
0.0625	8	adj.	1.99772
0.0625	8	num.	1.99471
0.1	5	adj.	2.0081
0.1	5	num.	2.00564
0.25	2	adj.	2.04803
0.25	2	num.	2.03402

Table 6.6: Optimal control of the Ornstein-Uhlenbeck process: Weighted total costs for various N and T_s.

First of all, it is notable that the method of computing the gradient does not influence the results distinctly for lower T_s, i.e. $T_s \leq 0.1$. This is to be expected, since the approximation

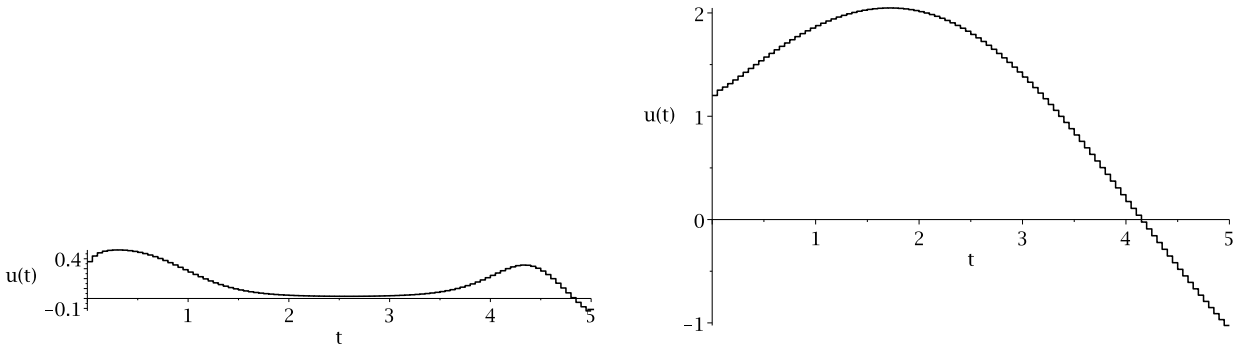
of the cost functional becomes better and the right-hand side of the adjoint PDE is updated more often.
Furthermore, the expected MPC behavior surfaces: While for $T_s = 0.05$ and $N = 1$, the results are useless, the situation changes with higher N, cf. Figure 6.7. Note that only the states at times $t_n = 0.5n, n = 0, ..., 10$ are shown in this figure. Intermediate states are omitted where necessary. The same holds true for all following figures.
Another interesting result emerges when looking at Table 6.6: Except for $N = 1$, $T_s \cdot N = 0.5$, i.e. a time frame of 0.5 is covered in each MPC step. In all of these cases, decent and similar results are obtained. It seems that the combination of T_s and N is important, rather than just the horizon N. The slight improvements for increasing N and lowering T_s probably stem from the possibility to adjust the control values more often, cf. Figure 6.8.



(a) $N = 1$ (b) $N = 10$

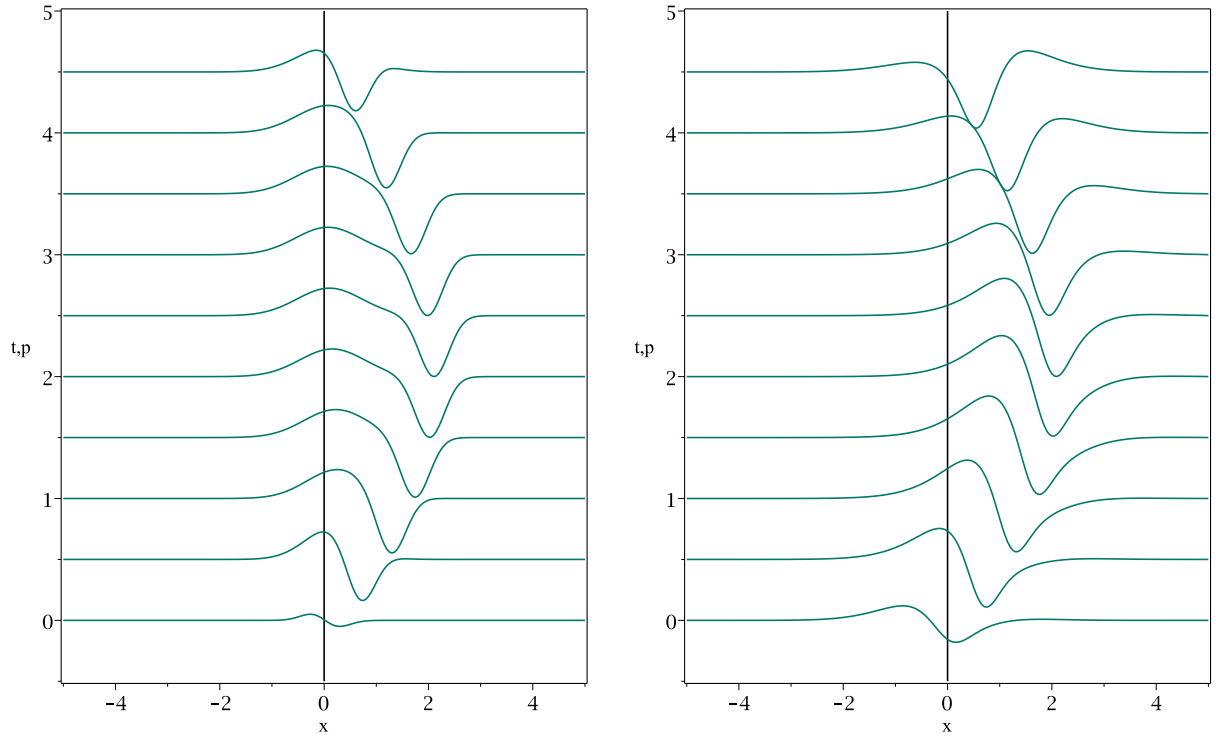
Figure 6.7: Optimal control of the Ornstein-Uhlenbeck process for $T_s = 0.05$ and varying N: The computed PDF y, the initial PDF y_0, and the desired PDF y_d.

In conclusion, decent results are obtained in cases where $T_s \cdot N \geq 0.5$ and T_s is sufficiently low such that the cost functional of the corresponding optimality system is approximated well enough and the right-hand side of the adjoint equation is updated often enough. Increasing the horizon while decreasing T_s at the same time yields the best results. Also, it is reasonable to assume that the optimality systems (2.48)-(2.50) and (2.52)-(2.54) are correct and solving these systems yields a (local) minimum. At the same time, increasing the horizon does



(a) $N = 1$ (b) $N = 10$

Figure 6.8: Optimal control of the Ornstein-Uhlenbeck process for $T_s = 0.05$ and varying N: The optimal control function.



(a) $N = 1, \max|p(x,t)| = 1.49361$ (b) $N = 10, \max|p(x,t)| = 0.252011$

Figure 6.9: Optimal control of the Ornstein-Uhlenbeck process for $T_s = 0.05$ and varying N: The adjoint state p.

not solve the problem of the computed PDF exhibiting a too high variance compared to the desired PDF. Since the control determines the shifting of the state, a space-dependent control likely improves the results tremendously by shifting every single point individually. Whether this proves true is discussed in Section 6.2.

6.1.2 Geometric-Brownian Process with Additive Drift Control

The second stochastic process examined in this work is the Geometric-Brownian process with additive drift control. The corresponding Fokker-Planck equation is given by

$$\begin{aligned} 0 &= \partial_t y(x,t) - \frac{1}{2}\partial_{xx}^2\left((\tilde{\sigma}x)^2 y(x,t)\right) + \partial_x\big((\mu+u)x\ y(x,t)\big) \\ &= \partial_t y(x,t) - \frac{1}{2}\tilde{\sigma}^2\partial_{xx}^2\left(x^2 y(x,t)\right) + (\mu+u)\partial_x\big(x\ y(x,t)\big) \end{aligned} \tag{6.2}$$

with $\mu \in \mathbb{R}$ and $\tilde{\sigma} > 0$, cf. Example 2.6. For $N = 1$ the control u is a real value in each MPC optimization step; for $N > 1$ it is a function of time, i.e. $u = u(t)$. Unless state otherwise, the parameter values used in the numerical simulations in this subsection are given in Table 6.7. Here, both the initial PDF ρ and the desired PDF y_d are log-normal distributions.

Parameter	Value
μ	1
$\tilde{\sigma}$	0.1
Ω	$]0,\ 25[$
h	0.0625
δt	2.4752e-4
u_a	-1
u_b	0.5
λ	0.1
T_E	2.5
T_s	0.25
$\rho(x)$	$\dfrac{1}{x\sqrt{2\pi\cdot 0.1^2}}\exp\left(-\dfrac{(\log(x)-0.8)^2}{2\cdot 0.1^2}\right)$
$y_d(x,t)$	$\dfrac{\exp\left(-\dfrac{[\log(x)-1-\sin(\pi t/5)]^2}{2\cdot 0.1^2}\right)}{x\sqrt{2\pi\cdot 0.1^2}}$

Table 6.7: Standard parameter values for the Geometric-Brownian process with additive drift control.

The procedure and the notation are the same as in the Ornstein-Uhlenbeck case. First, $N = 1$ is considered. The results are depicted in Figures 6.10-6.12 and Table 6.8.
Again, the state y follows the desired state y_d decently. With time, however, the diffusion becomes a problem. Not only is the height of the desired PDF not met, but also the computed PDF diffuses in time much faster than desired. At $t = T_E = 2.5$, it is questionable whether the computed PDF meets the user's requirements. Also, notice the jump in the control values in the first two steps. Without a good enough initial guess for the first two control values, the BFGS algorithm yields worse results in terms of costs. Therefore, in each MPC step,

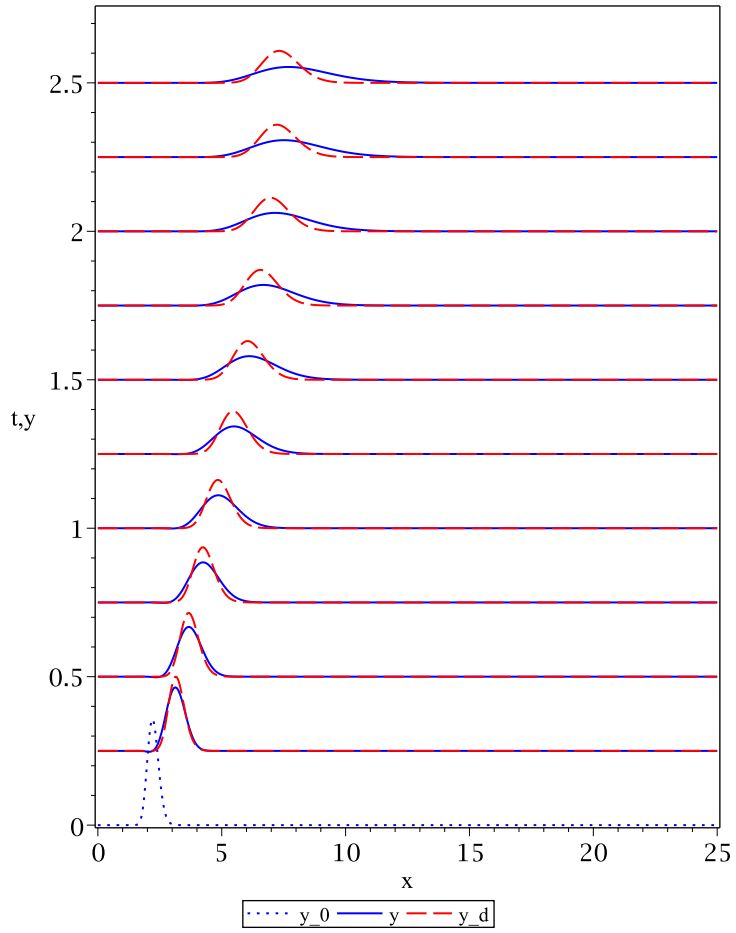


Figure 6.10: Optimal control of the Geometric-Brownian process for $N = 1$: The computed PDF y, the initial PDF y_0, and the desired PDF y_d.

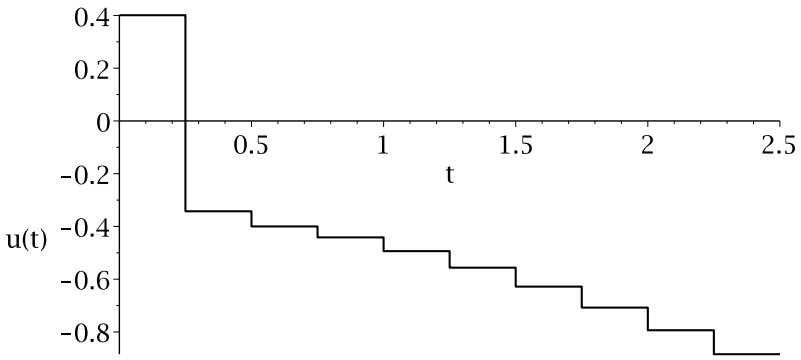


Figure 6.11: Optimal control of the Geometric-Brownian process for $N = 1$: The optimal control function.

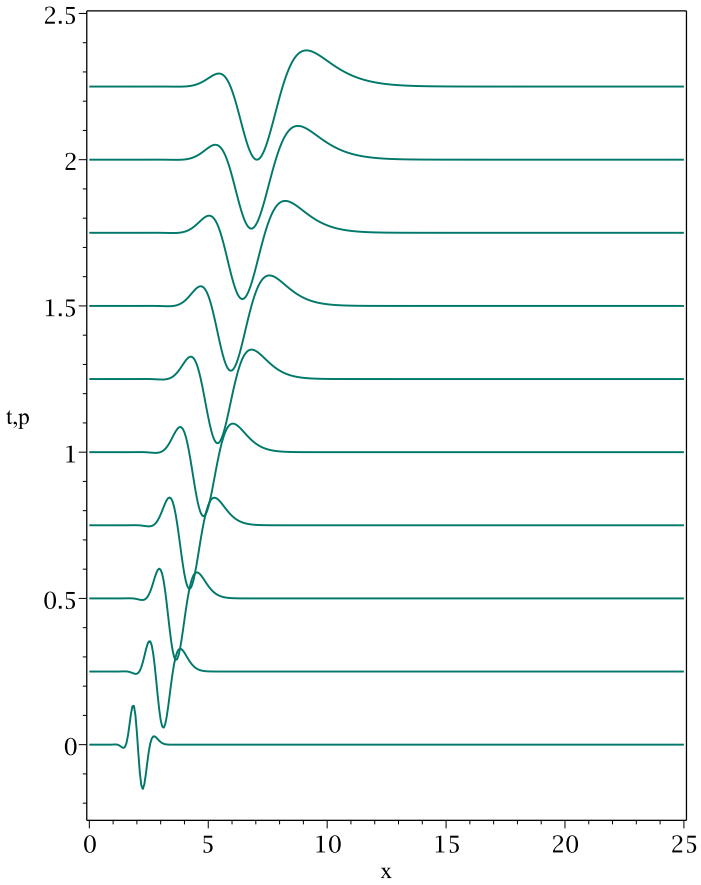


Figure 6.12: Optimal control of the Geometric-Brownian process for $N = 1$: The adjoint state p ($\max |p(x,t)| = 0.234792$).

t	u	$\frac{1}{2}\|y(\cdot, t+T_s) - y_d(\cdot, t+T_s)\|^2_{L^2(\Omega)}$	$\frac{\lambda}{2}\|u\|^2$
0	0.400988	0.00920641	0.00803957
0.25	-0.342515	0.0155605	0.00586584
0.5	-0.399993	0.0212408	0.00799971
0.75	-0.44157	0.0255997	0.00974919
1	-0.493676	0.0291417	0.0121858
1.25	-0.556354	0.0323978	0.0154765
1.5	-0.62822	0.035872	0.019733
1.75	-0.70778	0.0400275	0.0250476
2	-0.793652	0.0452798	0.0314942
2.25	-0.884473	0.0519928	0.0391146

Table 6.8: Optimal control of the Geometric-Brownian process for $N = 1$: Detailed results.

the first few iterations are computed using the Projected Gradient method. Even though the results as a whole might seem disappointing, compared to the figures in [3], the results are qualitatively identical.
In a next step, the horizon is increased. Figures 6.13-6.15 illustrate the numerical results; detailed numbers are found in Table 6.9.

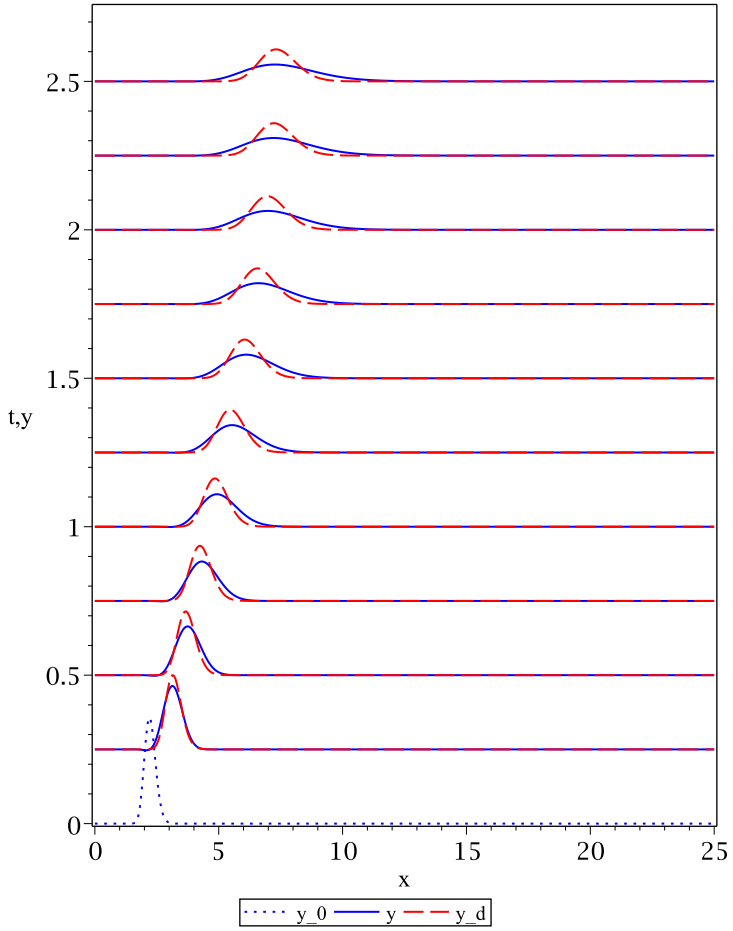


Figure 6.13: Optimal control of the Geometric-Brownian process for $N = 2$: The computed PDF y, the initial PDF y_0, and the desired PDF y_d.

When comparing Figure 6.13 to Figure 6.10, the target PDF at time T_E is reached somewhat better in the $N = 2$ case, since the peak values of both the target and the computed PDF are attained at about the same x. In contrast to the Ornstein-Uhlenbeck process, the control does not change significantly when switching from $N = 1$ to $N = 2$, cf. Figure 6.14 and Table 6.9. Notice, however, a lower value of T_s is used in this experiment. This and the fact that the costs for the Geometric-Brownian process are generally low (< 0.1) might mitigate the approximation errors. In addition, the system dynamics play an important role: The control u affects the drift term μ additively, i.e. the shift of the computed PDF depends on the value of $\mu + u$. Since $\mu = 1$, the uncontrolled Fokker-Planck equation would shift to the

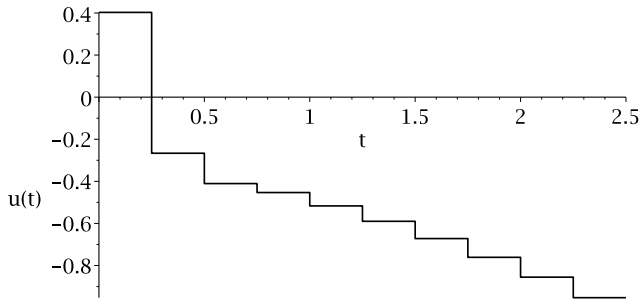


Figure 6.14: Optimal control of the Geometric-Brownian process for $N = 2$: The optimal control function.

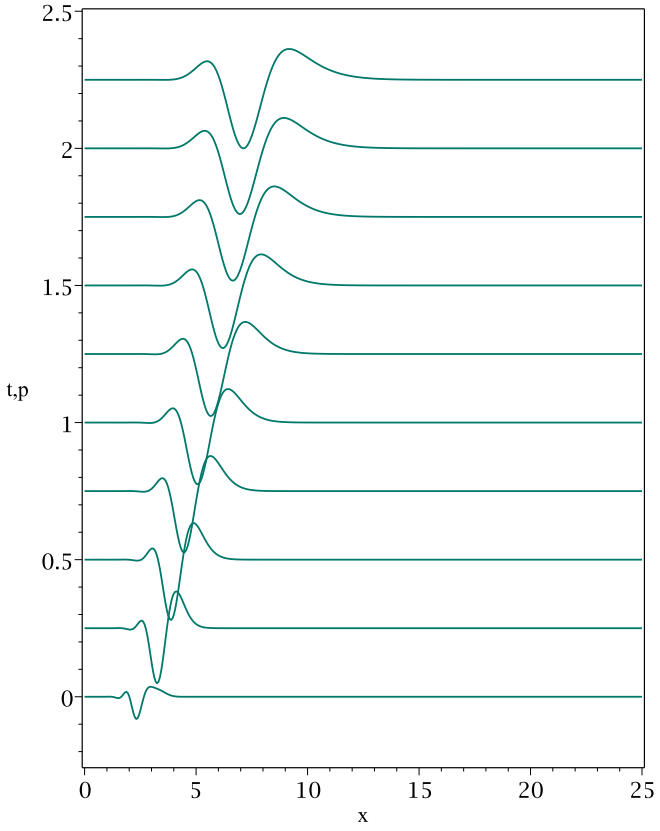


Figure 6.15: Optimal control of the Geometric-Brownian process for $N = 2$: The adjoint state p ($\max |p(x,t)| = 0.108432$).

right by its own. This falls into place with the desired PDF, which also shifts to the right as t increases, cf. e.g. Figure 6.10. Therefore, only little adjustments are needed in order to improve the tracking of y_d. Thus, the right-hand side in the adjoint equation is small, such that approximation errors become hardly noticeable. On that account, notice that the maximum absolute value of p is 0.108432 for $N = 2$. Table 6.10 backs this claim: The

t	u	$\frac{1}{2}\|y(\cdot, t+T_s) - y_d(\cdot, t+T_s)\|^2_{L^2(\Omega)}$	$\frac{\lambda}{2}\|u\|^2$
0	0.402697	0.00911289	0.00810825
0.25	-0.2665	0.0217662	0.00355111
0.5	-0.410663	0.0257687	0.00843221
0.75	-0.453266	0.0289298	0.0102725
1	-0.516805	0.0308815	0.0133544
1.25	-0.589932	0.0323257	0.017401
1.5	-0.671984	0.0336894	0.0225781
1.75	-0.761056	0.0352797	0.0289603
2	-0.855321	0.0373284	0.0365787
2.25	-0.952884	0.0400399	0.0453994

Table 6.9: Optimal control of the Geometric-Brownian process for $N = 2$: Detailed results.

weighted total costs are almost the same, regardless of whether the gradient was computed by solving the optimality system or not. As in the previous example, results obtained by solving the optimality system yield higher weighted total costs. In this case, however, the difference is hardly noticeable. Similar to the Ornstein-Uhlenback process, the raise in N does not necessarily lead to lower total costs. Instead, higher costs are observed due to the anticipation of events after T_E and the consideration thereof in the computation of the optimal control sequence, cf. Remark 3.12.

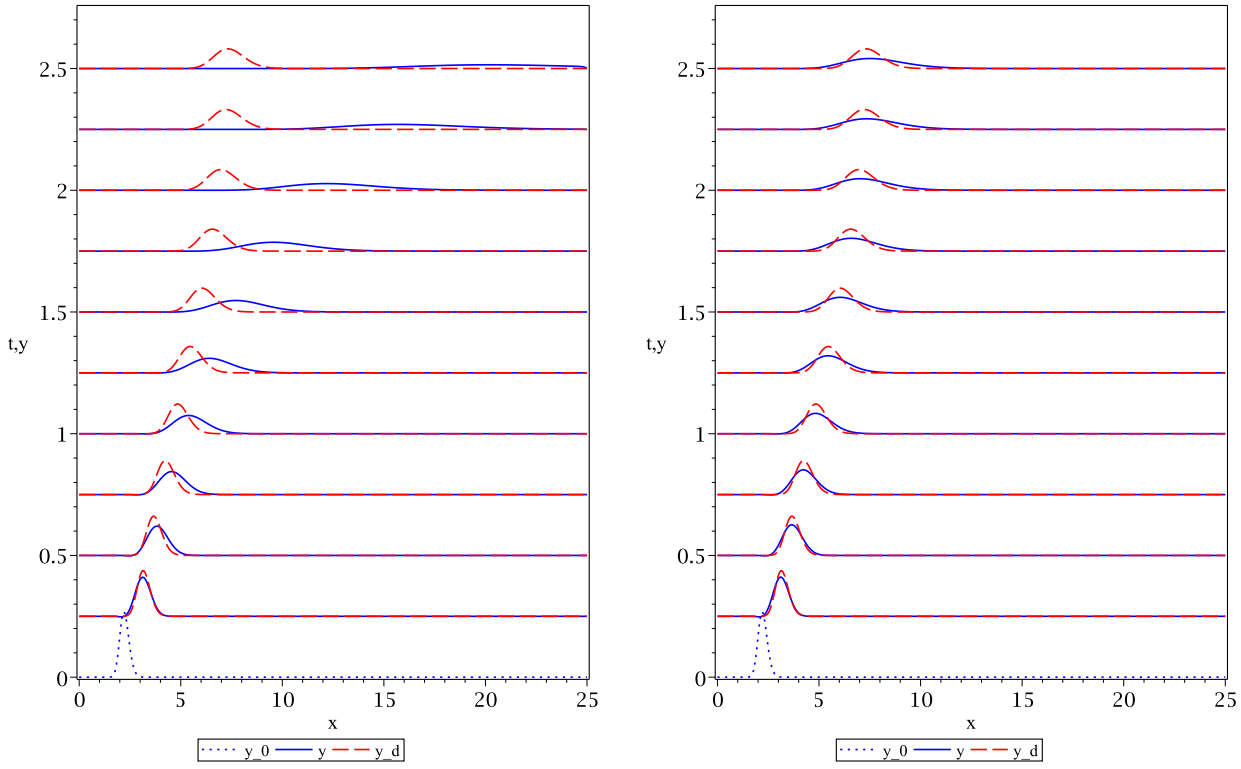
N	Computation of the gradient	$C_{total,w}$
1	adj.	0.120256
2	adj.	0.12244
2	num.	0.115704
5	adj.	0.123297
5	num.	0.118729

Table 6.10: Optimal control of the Geometric-Brownian process: Weighted total costs for various N.

In conclusion, it seems hardly worthwhile to increase N for $T_s = 0.25$. However, this changes when lowering T_s. To this end, the weighted total costs for various combinations of sampling rates T_s and horizons N are depicted in Table 6.11.
There are several observations to be made. First, the total costs for $N = 1$ and $T_s = 0.025$ are much higher than for $N = 10$ and unchanged T_s, i.e. increasing the MPC horizon is sensible. Figures 6.16-6.18 show the differences in the state, the control, and the adjoint state between $N = 1$ and $N = 10$. For $N = 1$ and $t_n \geq 1.25$ the computed PDF does not resemble at all the shape of the desired PDF. Second, the best results are gained for $N = 2$ and $T_s = 0.125$, i.e. increasing the horizon any further while lowering T_s does not pay off. This behavior is

T_s	N	Computation of the gradient	$C_{total,w}$
0.025	1	adj.	0.428949
0.025	1	num.	0.427031
0.025	10	adj.	0.161421
0.025	10	num.	0.161393
0.03125	8	adj.	0.159497
0.03125	8	num.	0.159361
0.05	5	adj.	0.154168
0.05	5	num.	0.153631
0.125	2	adj.	0.138042
0.125	2	num.	0.134475

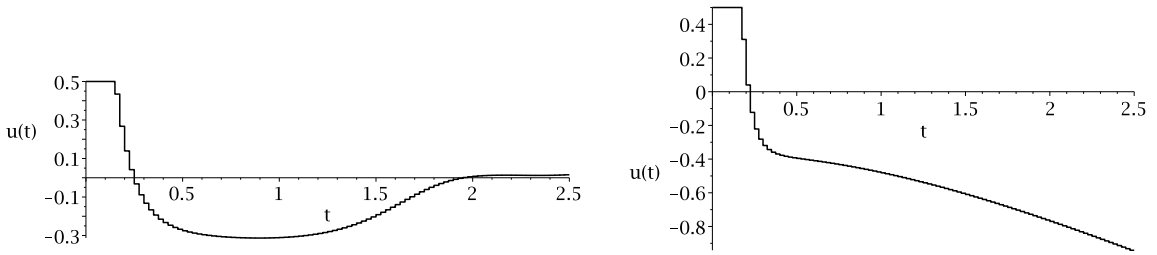
Table 6.11: Optimal control of the Geometric-Brownian process: Weighted total costs for various N and T_s.



(a) $N = 1$ (b) $N = 10$

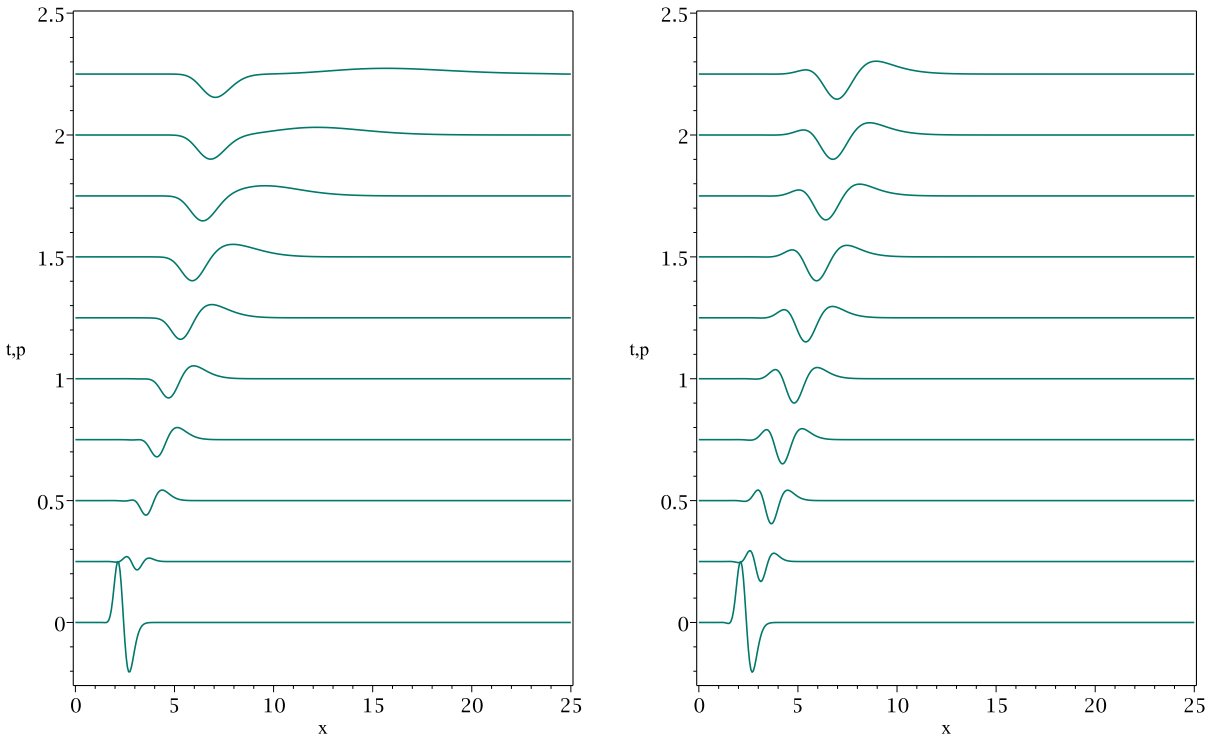
Figure 6.16: Optimal control of the Geometric-Brownian process for $T_s = 0.025$ and varying N: The computed PDF y, the initial PDF y_0, and the desired PDF y_d.

strange and might occur due to numerical reasons, as the absolute values of the costs are close to zero and the changes minimal. Furthermore, it is seemingly more important to cover



(a) $N = 1$ (b) $N = 10$

Figure 6.17: Optimal control of the Geometric-Brownian process for $T_s = 0.025$ and varying N: The optimal control function.



(a) $N = 1, \max|p(x,t)| = 1.41499$ (b) $N = 10, \max|p(x,t)| = 0.14504$

Figure 6.18: Optimal control of the Geometric-Brownian process for $T_s = 0.025$ and varying N: The adjoint state p.

a certain time interval in the MPC algorithm rather than just increasing the horizon, just as in the Ornstein-Uhlenbeck case. The last observation is that it virtually does not matter whether to solve the optimality system or to calculate the gradient differently. Therefore, once again, it is reasonable to assume that the two optimality systems given by (2.20)-(2.23) and (2.24)-(2.28) are correct and solving these systems yields a (local) minimum.
As before, the diffusion of the PDF is a problem that cannot be solved by space-independent control. Whether a space-dependent control is a solution is discussed in Section 6.2.

6.1.3 The Shiryaev Process

The Shiryaev process is the third stochastic process considered in this thesis. Plugging in the corresponding coefficient functions b and σ from Example 2.6 into the Fokker-Planck equation yields

$$\begin{aligned} 0 &= \partial_t y(x,t) - \frac{1}{2}\partial_{xx}^2\left((\tilde{\sigma}x)^2 y(x,t)\right) + \partial_x\big((u+\mu x)\ y(x,t)\big) \\ &= \partial_t y(x,t) - \frac{1}{2}\tilde{\sigma}^2\partial_{xx}^2\left(x^2 y(x,t)\right) + \partial_x\big((u+\mu x)\ y(x,t)\big) \end{aligned} \tag{6.3}$$

with $\mu \in \mathbb{R}, \tilde{\sigma} > 0$, and u being either a real value in each MPC optimization step if $N = 1$ or a function of time, i.e. $u = u(t)$, if $N > 1$. The standard values used for the numerical experiments are almost the same as in the Geometric-Brownian case, cf. Table 6.7. The modifications are depicted in Table 6.12.

Parameter	Value
μ	0.5
u_b	3

Table 6.12: Standard parameter values for the Shiryaev process differing from Table 6.7.

First, N is set to 1 in order to reproduce the results in [3]. Figures 6.19-6.21 illustrate the computed PDF y, the initial PDF y_0, the desired PDF y_d, the optimal control u, and the adjoint state p. Qualitatively, the results coincide with the ones in [3]. Detailed information about the control values and the stage costs are found in Table 6.13.

t	u	$\frac{1}{2}\|y(\cdot, t+T_s) - y_d(\cdot, t+T_s)\|^2_{L^2(\Omega)}$	$\frac{\lambda}{2}\|u\|^2$
0	1.93507	0.0498337	0.187224
0.25	0.666244	0.0117792	0.0221941
0.5	0.386913	0.0102034	0.00748506
0.75	0.238202	0.0117858	0.00283701
1	0.10297	0.0133056	0.000530143
1.25	-0.0301808	0.0174462	4.55439e-005
1.5	-0.151709	0.0311941	0.00115078
1.75	-0.239592	0.0635277	0.00287022
2	-0.259613	0.11857	0.00336994
2.25	-0.193784	0.184277	0.00187761

Table 6.13: Optimal control of the Shiryaev process for $N = 1$: Detailed results.

In addition to the diffusion problems that are familiar by now, the computed PDF moves too far to the right for $t \geq 2$. Also, the control exhibits a noticeable jump from $t_n = 0$

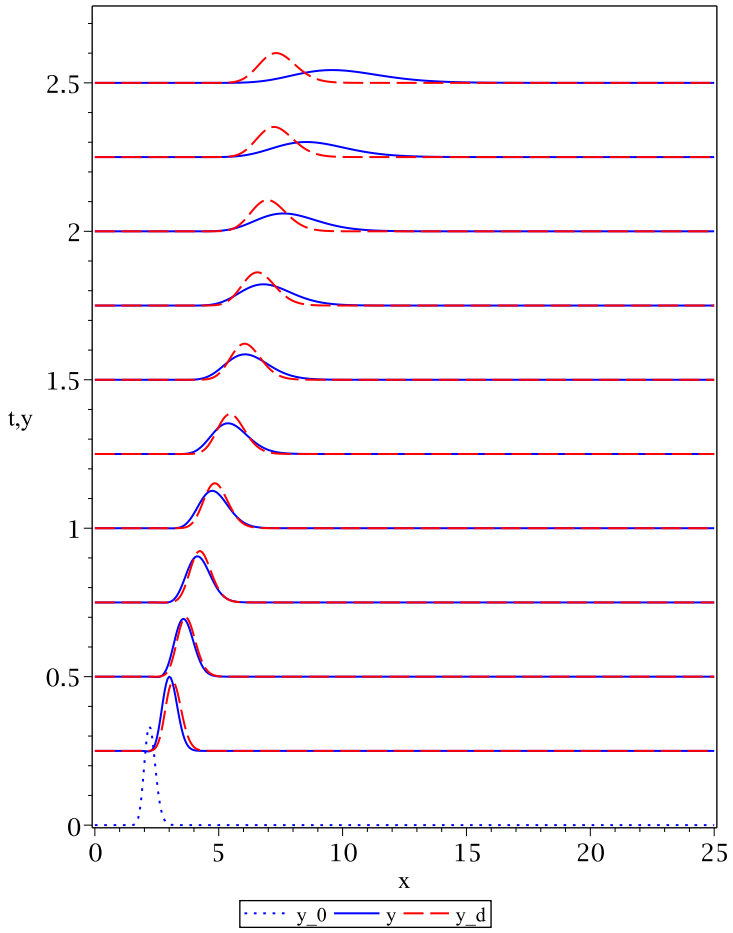


Figure 6.19: Optimal control of the Shiryaev process for $N = 1$: The computed PDF y, the initial PDF y_0, and the desired PDF y_d.

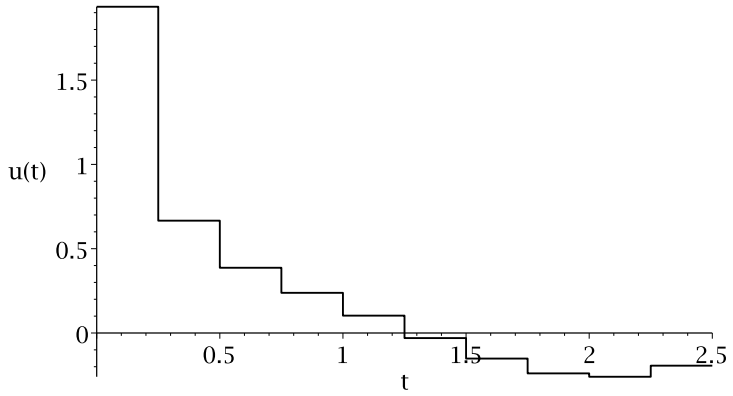


Figure 6.20: Optimal control of the Shiryaev process for $N = 1$: The optimal control function.

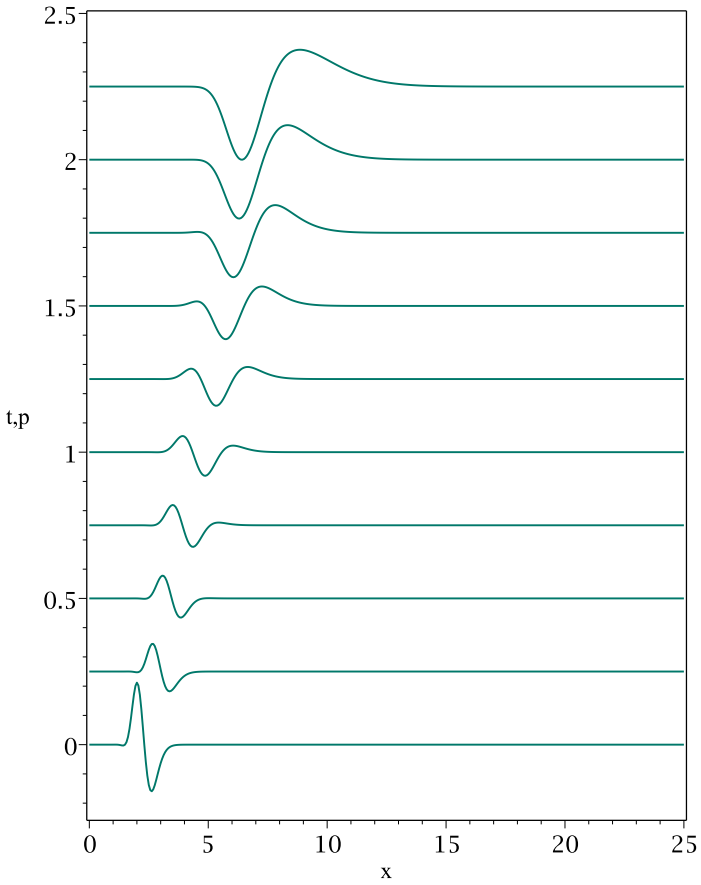


Figure 6.21: Optimal control of the Shiryaev process for $N = 1$: The adjoint state p ($\max |p(x,t)| = 0.413842$).

to $t_n = 0.25$. In the numerical experiments, the time frames $[0, 0.25]$ and $[2, 2.5]$ take particularly long to optimize. This is resembled by Figures 6.19 and 6.20: Starting with the initial distribution ρ, a suitable control needs to be found in order to move the computed PDF to the right such that the desired PDF is reached. Once y is near the desired state y_d, it seems relatively easy to stay near since the desired PDF does not change very much. After a while, though, the diffusion makes it increasingly difficult to attain the shape of the desired PDF. Whether an increased horizon provides a remedy is examined next.

The numerical results illustrated in Figures 6.22-6.24 show that increasing the horizon to $N = 2$ leads to a significantly worse performance. This is particularly striking since this did not occur in the Geometric-Brownian case, where the same initial distribution and target PDF were used. Consequently, the answer must lie in the different system dynamics. The control u still influences the drift, but in a different way, as it is not factored by x anymore. This results in much higher values of p than in the Geometric-Brownian case, especially for $t = 0$, cf. Figures 6.24 and 6.15. Therefore, the same argument applies as in the Ornstein-Uhlenbeck process: For $N = 2$ and $T_s = 0.25$, the approximation errors are too big. Numerical

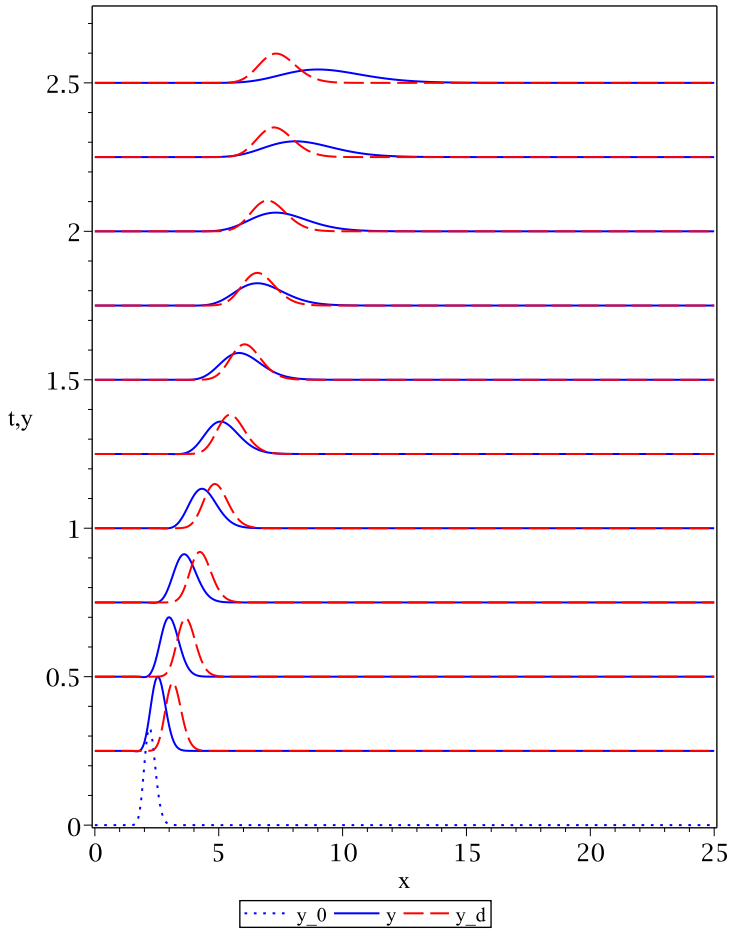


Figure 6.22: Optimal control of the Shiryaev process for $N = 2$: The computed PDF y, the initial PDF y_0, and the desired PDF y_d.

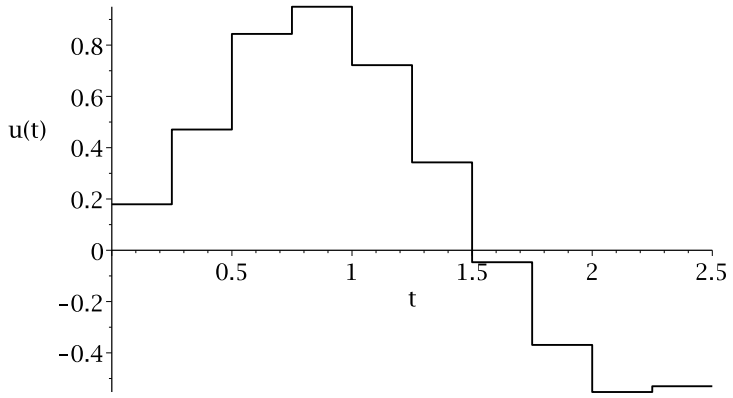


Figure 6.23: Optimal control of the Shiryaev process for $N = 2$: The optimal control function.

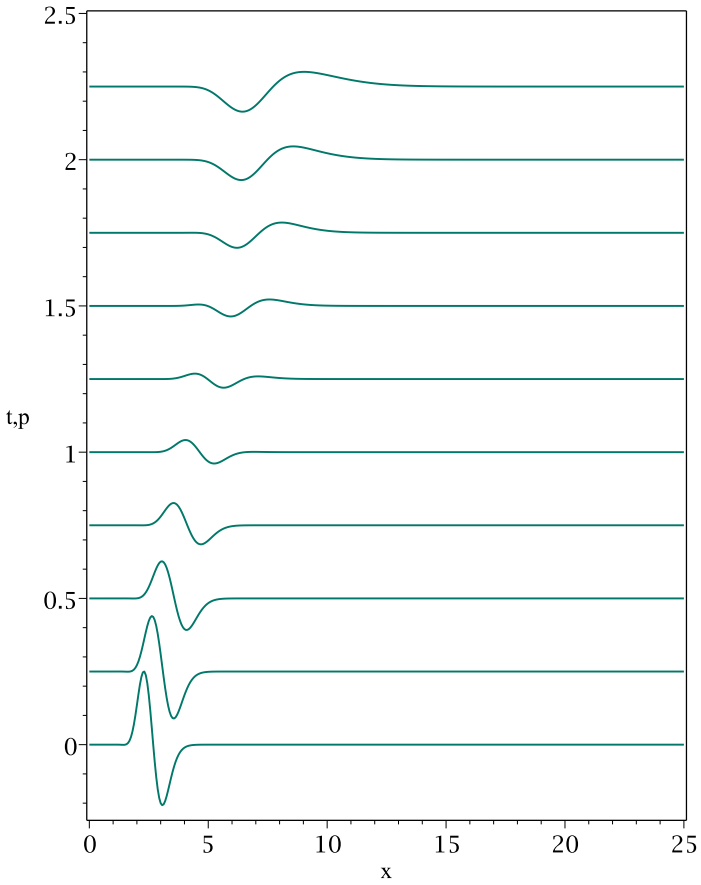


Figure 6.24: Optimal control of the Shiryaev process for $N = 2$: The adjoint state p ($\max |p(x,t)| = 0.484314$).

t	u	$\frac{1}{2}\|y(\cdot, t+T_s) - y_d(\cdot, t+T_s)\|^2_{L^2(\Omega)}$	$\frac{\lambda}{2}\|u\|^2$
0	0.179319	0.601528	0.00160777
0.25	0.470791	0.444718	0.0110822
0.5	0.843601	0.264584	0.0355832
0.75	0.949546	0.123439	0.0450819
1	0.721821	0.0494279	0.0260513
1.25	0.34269	0.0211148	0.00587184
1.5	-0.0465469	0.0187774	0.000108331
1.75	-0.369046	0.037421	0.00680974
2	-0.552653	0.0798171	0.0152713
2.25	-0.530164	0.144643	0.0140537

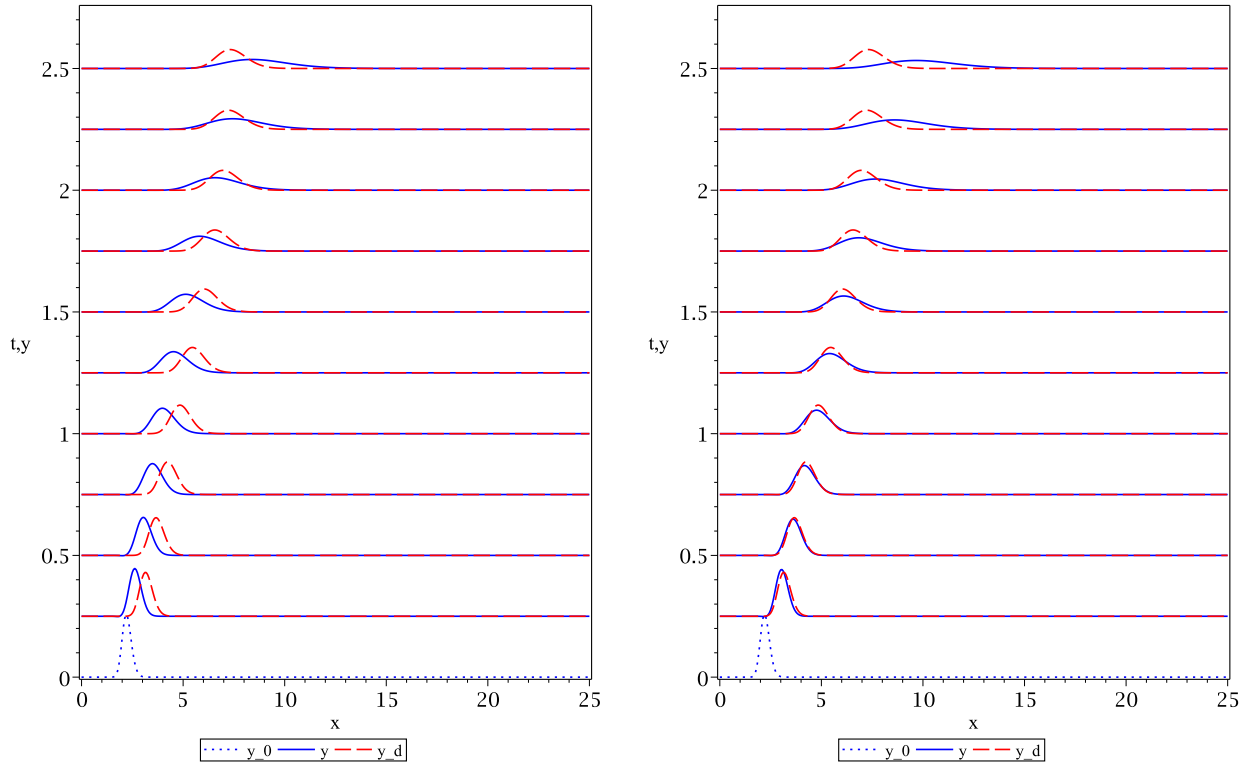
Table 6.14: Optimal control of the Shiryaev process for $N = 2$: Detailed results.

simulations performed without computing the gradient with the help of the adjoint state result in much better performance, cf. Table 6.15.

N	Computation of the gradient	$C_{total,w}$
1	adj.	0.185377
2	adj.	0.486748
2	num.	0.175844
5	adj.	0.223223
5	num.	0.167452

Table 6.15: Optimal control of the Shiryaev process: Weighted total costs for various N.

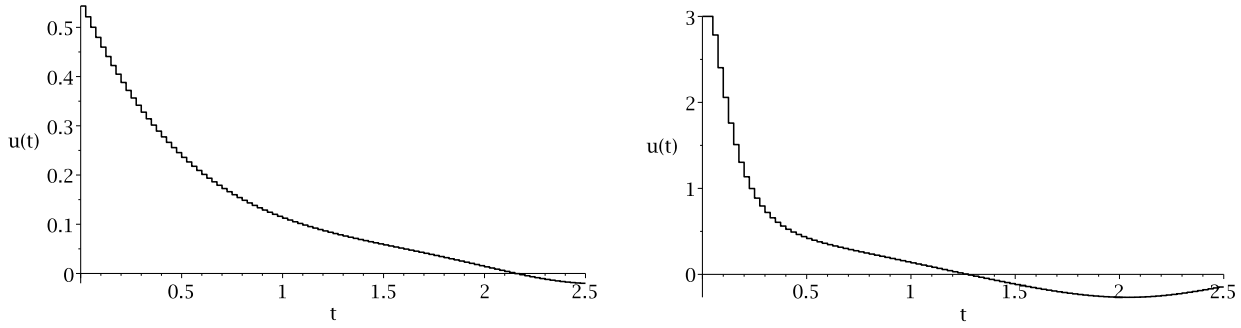
In conclusion, increasing the horizon for $T_s = 0.25$ does not pay off. Yet, as expected, this is different for lower values of T_s. Figures 6.25-6.27 display the numerical results for $T_s = 0.025$ and varying horizons N. Further results are depicted in Table 6.16.



(a) $N = 1$ (b) $N = 10$

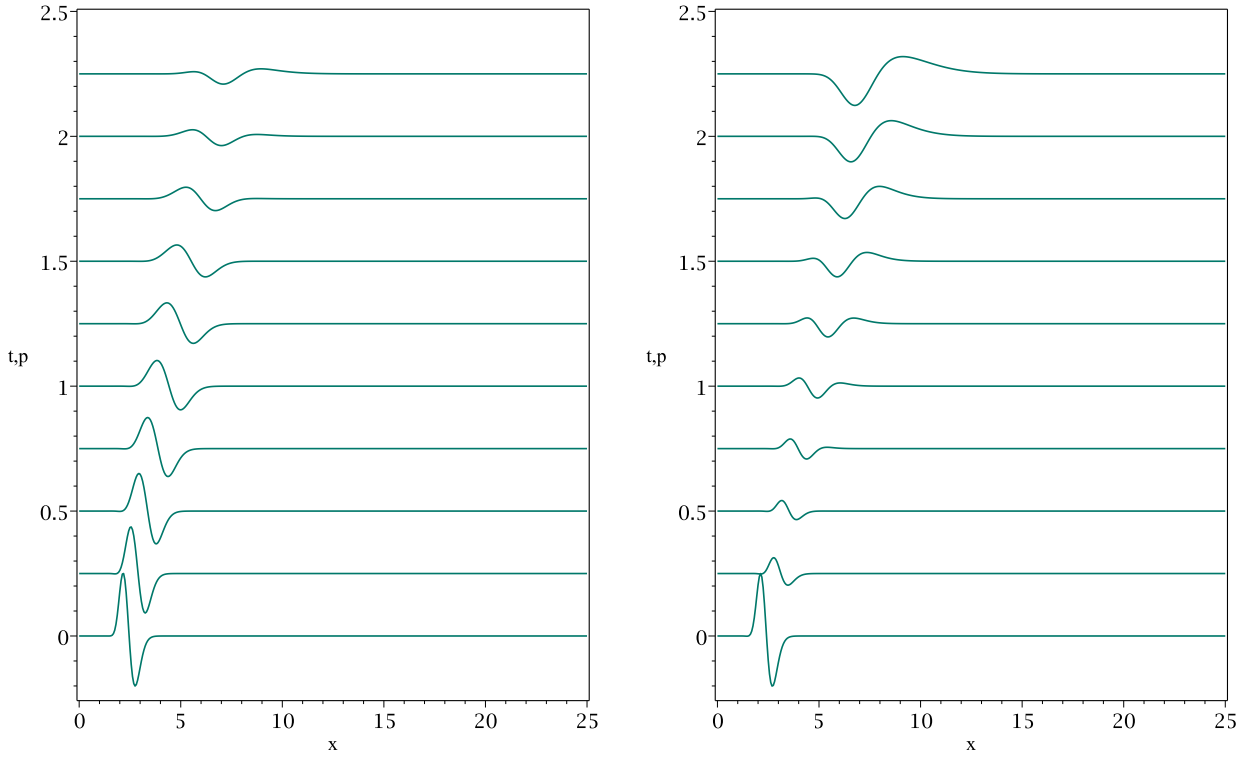
Figure 6.25: Optimal control of the Shiryaev process for $T_s = 0.025$ and varying N: The computed PDF y, the initial PDF y_0, and the desired PDF y_d.

The same arguments as in the Ornstein-Uhlenbeck case apply here: Smaller values of T_s lead to smaller approximation errors of the cost functional and at the same time, the right-



(a) $N = 1$ (b) $N = 10$

Figure 6.26: Optimal control of the Shiryaev process for $T_s = 0.025$ and varying N: The optimal control function.



(a) $N = 1, \max|p(x,t)| = 1.53539$ (b) $N = 10, \max|p(x,t)| = 0.195778$

Figure 6.27: Optimal control of the Shiryaev process for $T_s = 0.025$ and varying N: The adjoint state p.

hand side in the adjoint equation is updated more often. Increasing N up to a certain point for low T_s yields a better performance. After a certain threshold, however, no further improvements are noticeable. The covering of a certain time frame seems more important than the value of N by itself, i.e. the discretization of the control function plays only a minor

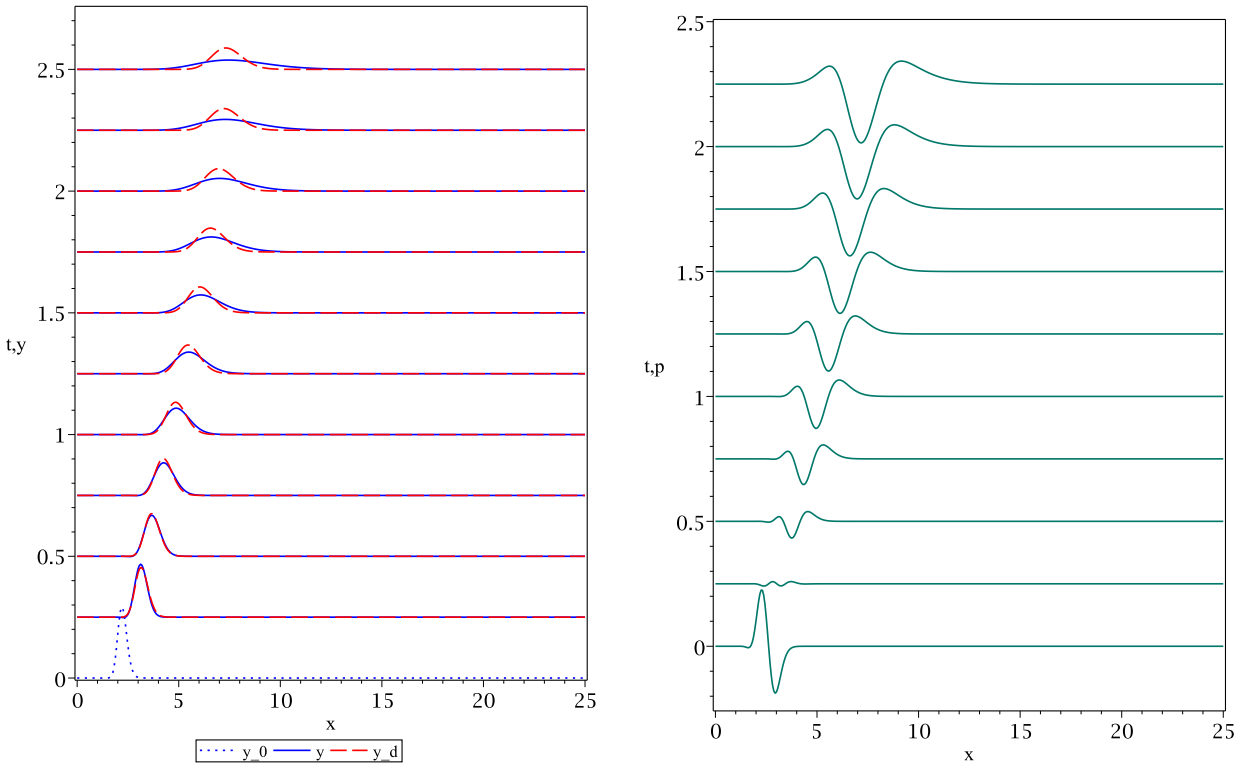
T_s	N	Computation of the gradient	$C_{total,w}$
0.025	1	adj.	0.59592
0.025	1	num.	0.63888
0.025	10	adj.	0.214732
0.025	10	num.	0.214846
0.03125	8	adj.	0.213354
0.03125	8	num.	0.213124
0.05	5	adj.	0.210722
0.05	5	num.	0.208683
0.125	2	adj.	0.23012
0.125	2	num.	0.194129

Table 6.16: Optimal control of the Shiryaev process: Weighted total costs for various N and T_s.

role. Furthermore, as long as T_s is sufficiently low, solving the optimality systems is a viable way to get (locally) optimal controls.
Yet, even for $T_s = 0.025$ and $N = 10$, the results are underwhelming. For $t \geq 2$, the computed PDF differs too much from the desired PDF y_d, cf. Figure 6.25. In the previous examples, at least the peak values were roughly at the same x. In order to negate the drifting to the right at the end, a low control value is needed. Apparently, it is too expensive to employ lower values of u. The solution is to reduce the regularization parameter from $\lambda = 0.1$ to $\lambda = 0.001$ and to extend the admissible controls, i.e. setting $u_a = -3$ instead of -1. The results in the case of $N = 2$ and $T_s = 0.125$ are illustrated in Figures 6.28-6.29. It is shown that with this configuration, the peak values are approximately at the same position. As in the previous examples, the diffusion of the PDF is an unsolvable problem for space-independent control. The following section will shine light on whether a space-dependent control is the solution to this problem.

6.2 Space-dependent Control

In this section, the control in each MPC optimization step is either a function of space, i.e. $u = u(x)$, or a function of time and space, i.e. $u = u(x,t)$, depending on whether $N = 1$ or $N > 1$. All three processes considered in this work exhibited the same property for space-independent control: a too high variance, i.e. compared to the desired PDFs, the computed PDFs were too widespread. Since the control influences the drift of the PDF, controlling the drift in every single point individually sounds promising. In this case, the MPC cost functional is given by (3.10). The parameter α in the new cost functional is set to 1 throughout this section, unless stated otherwise.
Notice that in this case, one should determine the gradient $\nabla \hat{J}(u)$ by solving the adjoint problem whenever possible, as it is much more efficient than analyzing perturbed controls.



(a) y, y_0, y_d (b) $p, \max|p(x,t)| = 0.0715941$

Figure 6.28: Optimal control of the Shiryaev process for $N = 2, T_s = 0.125, \lambda = 0.001$, and $u_a = -3$: The computed PDF y, the initial PDF y_0, the desired PDF y_d, and the adjoint state p.

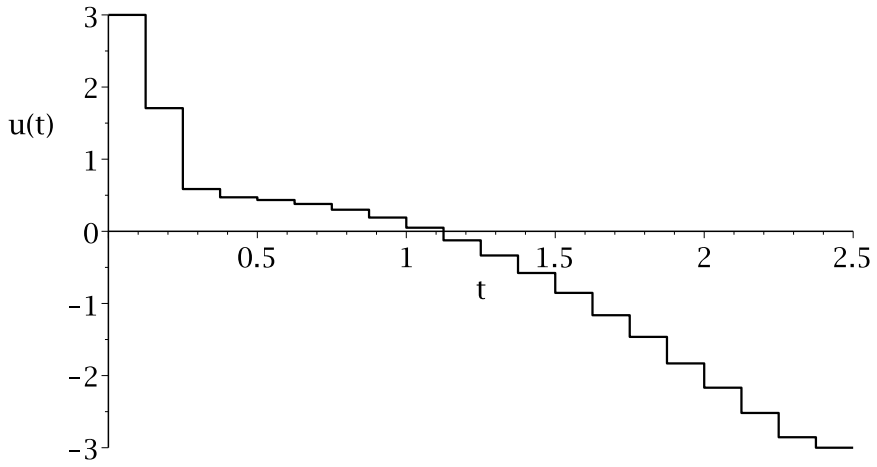
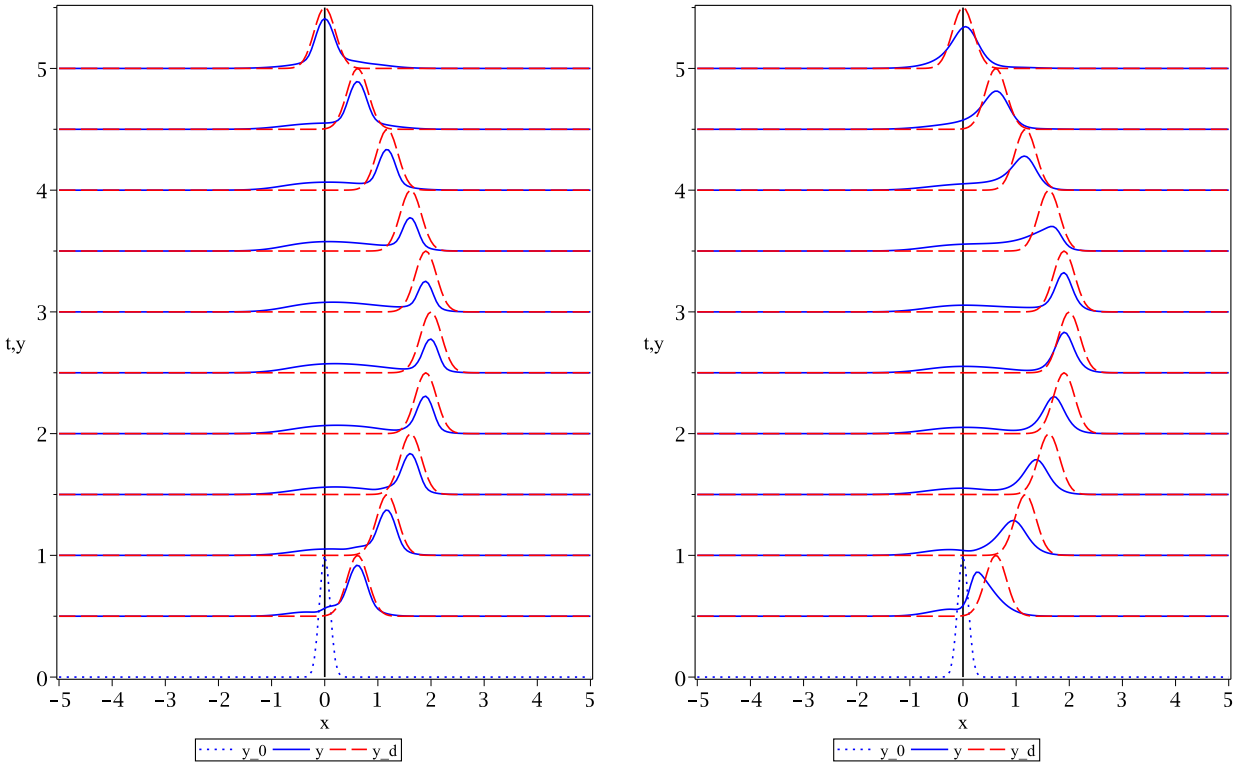


Figure 6.29: Optimal control of the Shiryaev process for $N = 2, T_s = 0.125, \lambda = 0.001$, and $u_a = -3$: The optimal control function.

6.2.1 The Ornstein-Uhlenbeck Process with Additive Control

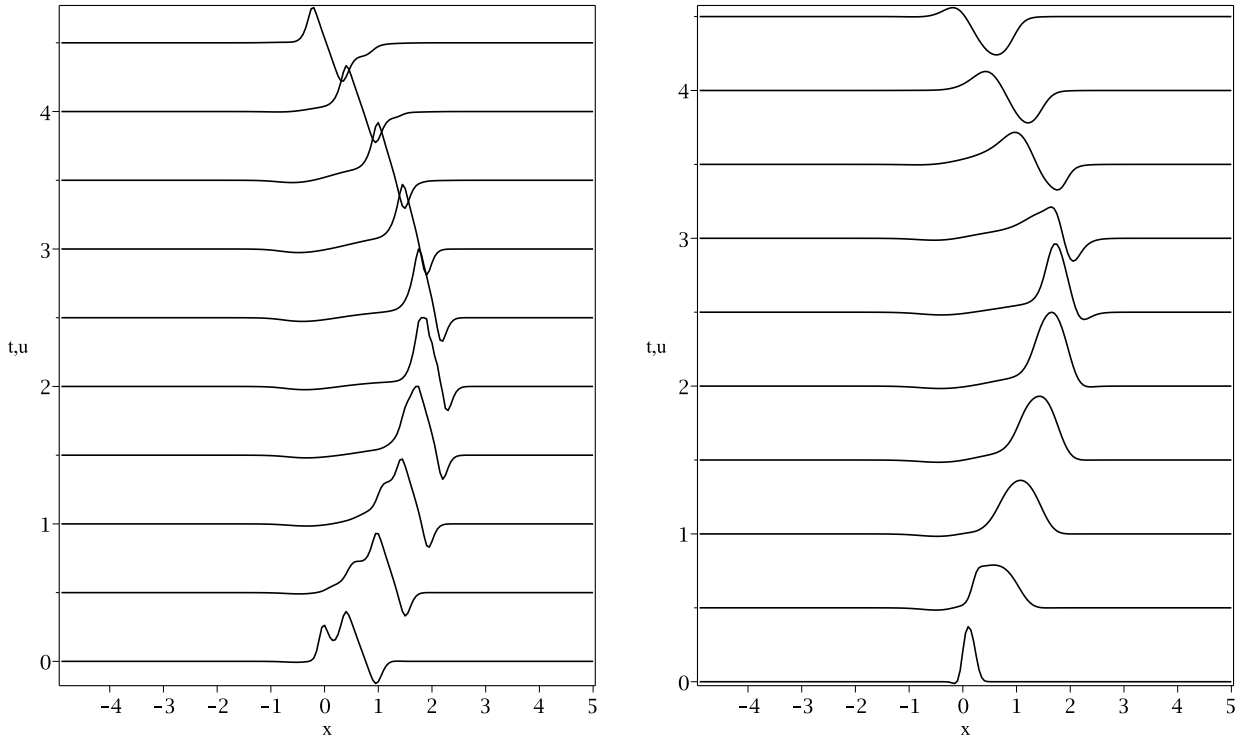
The Fokker-Planck equation for the Ornstein-Uhlenbeck process and space-dependent control is the same as for space-independent control, i.e. it is given by (6.1). Unless stated otherwise, the default values of the occurring parameters are the same as in the space-independent case, cf. Table 6.2. Following the structure of the previous numerical experiments, N is set to 1 first and increased afterwards.



(a) $N = 1$ (b) $N = 2$

Figure 6.30: Optimal space-dependent control of the Ornstein-Uhlenbeck process for varying N: The computed PDF y, the initial PDF y_0, and the desired PDF y_d.

Figures 6.30-6.32 depict the computed PDF y, the initial PDF y_0, the desired PDF y_d, the control u, and the adjoint state p for $N = 1$ and $N = 2$, respectively. As expected, a space-dependent control does influence the diffusion: For $N = 1$ and $t = T_E$, the computed PDF looks more similar to the desired one than in the space-independent case, cf. e.g. Figure 6.2. However, it is debatable whether for interim times a space-dependent control yields better results, since a second, smaller peak emerges. In addition, the same problems as in the space-independent case are encountered when setting $N = 2$: For $T_s = 0.5$, the discrepancy between the MPC cost functional and the cost functional in the optimality system that is solved in order to obtain gradient information is too high. This highly affects the computation of the optimal control. Self-evidently, detailed control values in this case



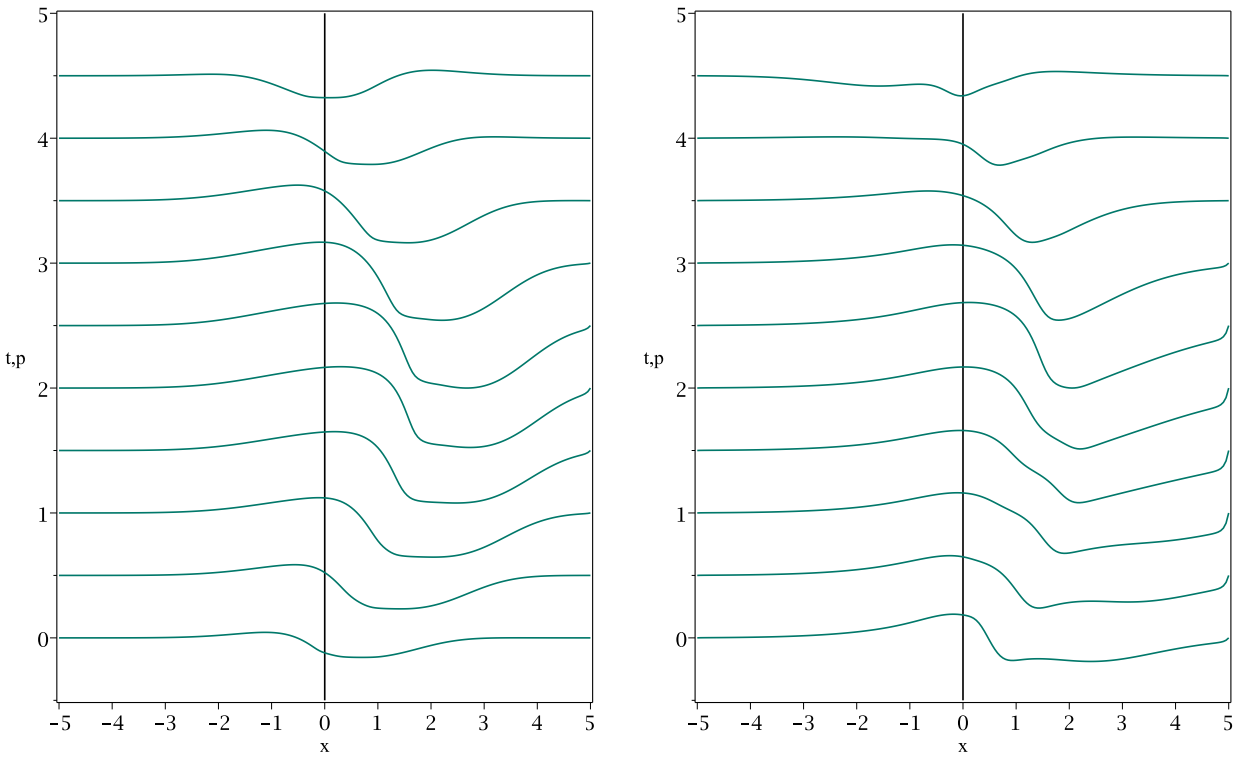
(a) $N = 1, \max |u(x,t)| = 4$ (b) $N = 2, \max |u(x,t)| = 3.83484$

Figure 6.31: Optimal space-dependent control of the Ornstein-Uhlenbeck process for varying N: The optimal control function.

cannot be provided in a table. However, looking at Figure 6.31, differences in the control when N is increased to 2 are easily observed.

Since space-dependent control was not considered in [3] or [4], there are no reference values that could be used for comparison purposes. However, Figure 6.30 shows that solving the optimality systems yields a control that is likely to be a (local) minimum, since the MPC algorithm obviously tries to compute a control sequence such that the desired trajectory is reached as good as possible. For testing purposes, additional simulations were run where the gradient information has been obtained by analyzing the effects of small changes in the control, cf. Remark 4.6. The results are presented in Table 6.17 and Figures 6.33-6.34.

First of all, Table 6.17 shows that obtaining the gradient by solving the adjoint problem yields higher costs, even for $N = 1$. However, the higher differences for $N > 1$ support the reasoning above, i.e. the discrepancy between the cost functionals seems too high. The increase in cost from 1.82967 for $N = 1$ to 2.0627 for $N = 2$ backs this argument. Also, it always pays off to increase the horizon if the gradient is not obtained by solving the adjoint problem, even when future states after T_E influence the computation. However, $N = 2$ seems to provide already good results, cf. Figure 6.33, and further increase in N leads to only slight reduction in the total costs. With these numbers, however, one might wonder whether it is viable at all to use the derived optimality systems. To this end, a comparison of



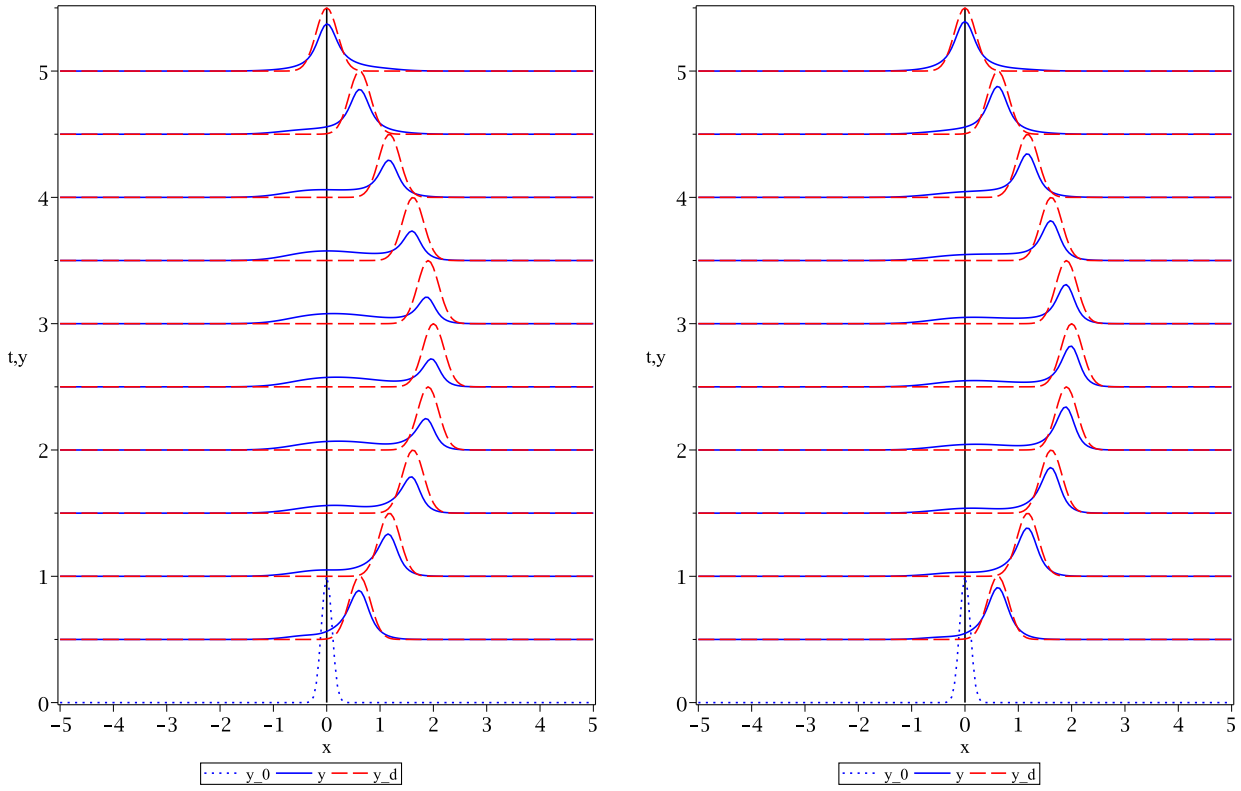
(a) $N = 1, \max |p(x,t)| = 0.752818$ (b) $N = 2, \max |p(x,t)| = 0.565058$

Figure 6.32: Optimal space-dependent control of the Ornstein-Uhlenbeck process for varying N: The adjoint state p.

N	Computation of the gradient	$C_{total,w}$
1	adj.	1.82967
1	num.	1.73789
2	adj.	2.0627
2	num.	1.44323
5	adj.	1.94317
5	num.	1.36755

Table 6.17: Optimal space-dependent control of the Ornstein-Uhlenbeck process: Weighted total costs for various N.

Figure 6.33 to Figure 6.30 shows that the qualitative results coincide for $N = 1$. Therefore, the optimality system (2.56)-(2.58) seems suitable. To conclude the same for the optimality system (2.60)-(2.62), smaller values of T_s are needed in order to mitigate the above discussed errors. Hence, in a next step, T_s is reduced from 0.5 to 0.05.
Figures 6.35-6.37 illustrate the numerical results for $T_s = 0.05$, $N = 10$, and varying α. The parameter α is introduced in order to reduce the control costs relatively to the state



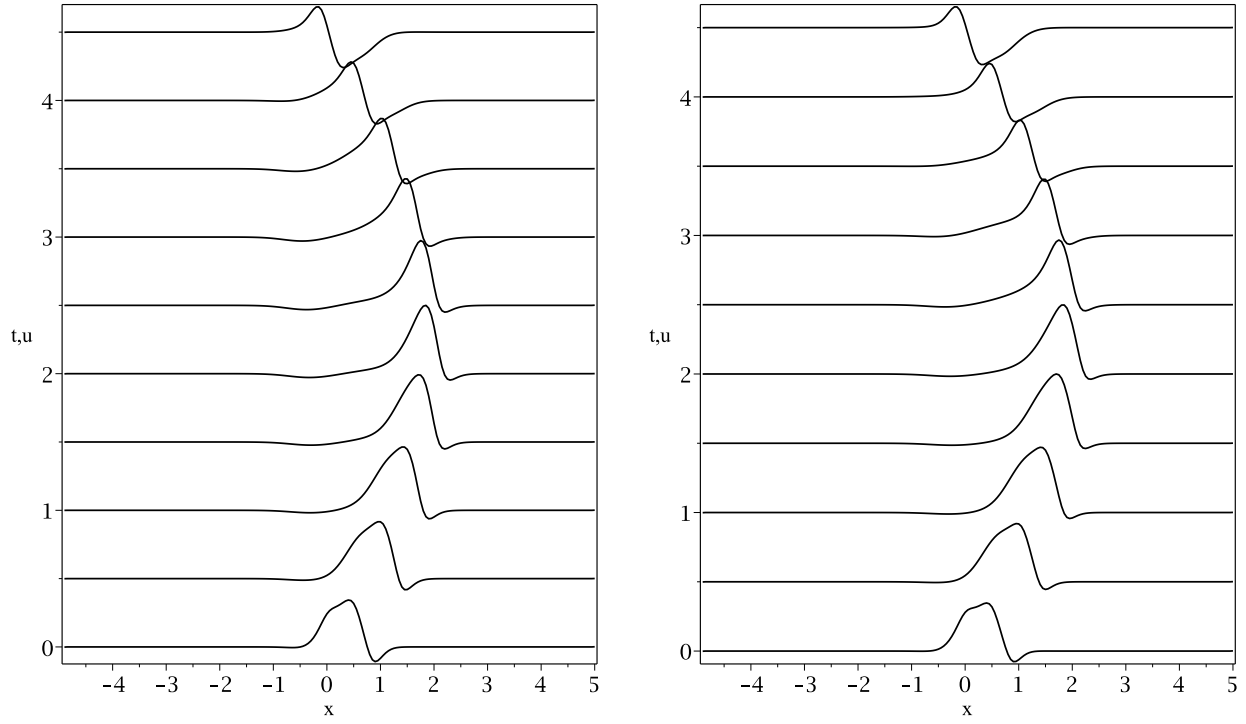
(a) $N = 1$ (b) $N = 2$

Figure 6.33: Optimal space-dependent control of the Ornstein-Uhlenbeck process for varying N (num.): The computed PDF y, the initial PDF y_0, and the desired PDF y_d.

costs. This can also be achieved by lowering the value of the regularization parameter λ. However, in practice, handling a low value of λ proves to be difficult for the optimization algorithms, since ultimately, this procedure is equivalent to evaluating the corresponding projection formulas, e.g. (2.65), in which λ appears in the denominator. Therefore, instead of reducing λ by a factor of 10, α is multiplied by this factor. For $\alpha = 1$, the results are qualitatively the same as for $(T_s, N) = (0.5, 1)$, implicating that solving the optimality system (2.60)-(2.62) is indeed viable. Increasing α to 10 yields an even better tracking of the desired PDF, cf. Figure 6.35.

Numerical simulations for low values of T_s and $N = 1$ have been omitted, since it has already been shown in Section 6.1 that for low values of T_s, increasing the horizon N improves the results significantly.

On another note, in all situations in which the box constraints of the control are not active, i.e. $\max |u(x,t)| < 4$, the resulting control functions appear to be smooth w.r.t. x, cf. Figures 6.31, 6.34, and 6.36. In particular, one can assume that $\partial_x u$ exists in the classical sense, which is required in order to prove existence and uniqueness of solutions of the Fokker-Planck equation for space-dependent control, cf. Subsection 2.2.1. If, in addition, u is Lipschitz-continuous w.r.t. x, then applying this control in the stochastic setting is meaningful as well,



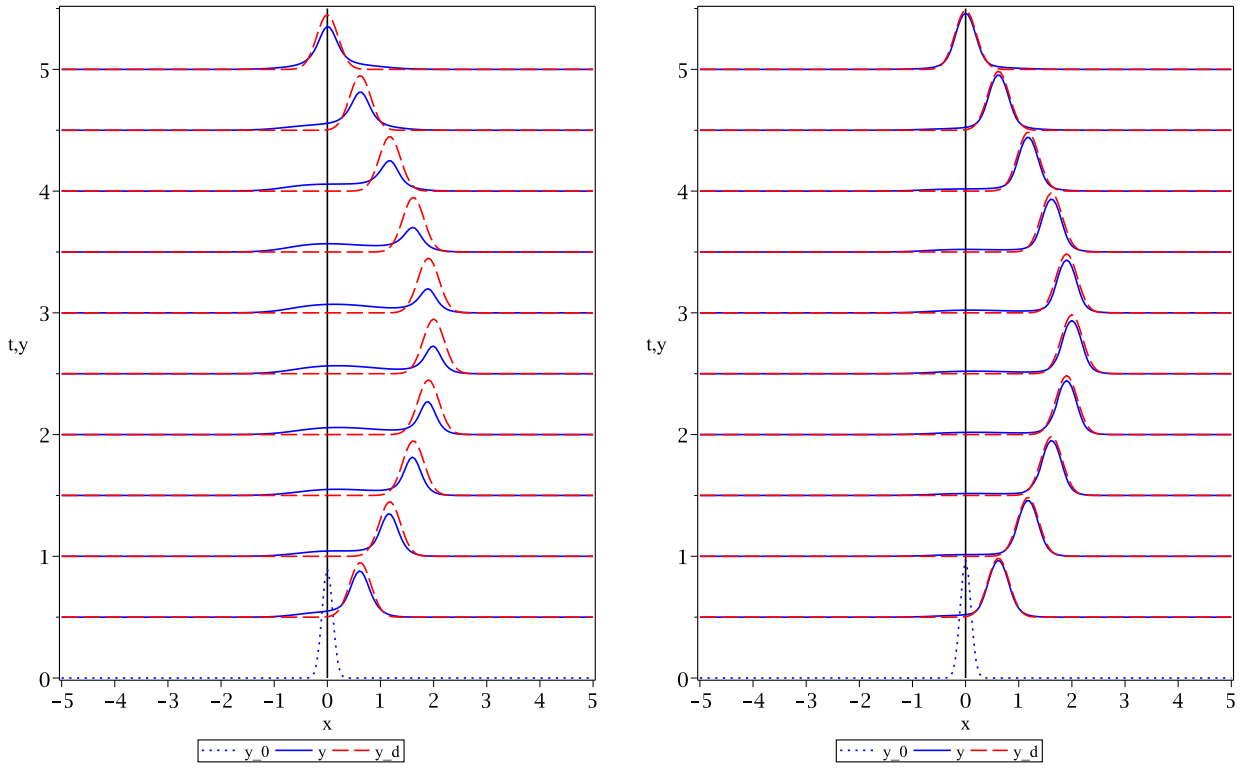
(a) $N = 1, \max|u(x,t)| = 3.0315$ (b) $N = 2, \max|u(x,t)| = 3.26975$

Figure 6.34: Optimal space-dependent control of the Ornstein-Uhlenbeck process for varying N (num.): The optimal control function.

cf. Assumption 2.5 and Example 2.6.

To conclude the numerical experiments for the Ornstein-Uhlenbeck process, a last test is undertaken in order to see how effective the control can be. It has been mentioned several times now that covering a certain time interval is more important than the value of the horizon itself. Therefore, in the following setting, the parameter T_s is set to its default value 0.5 and the horizon N is set to 1 to put this statement to a test. From Figure 6.35 it seems that the limiting factor for a very good tracking are the control costs. To remove this limit, λ is set to 0.01 and α is set to 100. In this experiment, the gradient information is obtained by analyzing the effects of small changes in the control. The results are depicted in Figures 6.38 and 6.39. The control in Figure 6.39 exhibits a strange behavior at the right boundary, probably the result of some numerical effects, e.g. the use of first order approximations of derivatives by Finite Differences. However, Figure 6.38 shows no observable deviation from the desired PDF. Although for standard parameter values, increasing the horizon N from 1 to 2 is beneficial (when not solving the adjoint equation in order to obtain gradient information), it seems that $N = 1$ already guarantees satisfactory results in this setting.

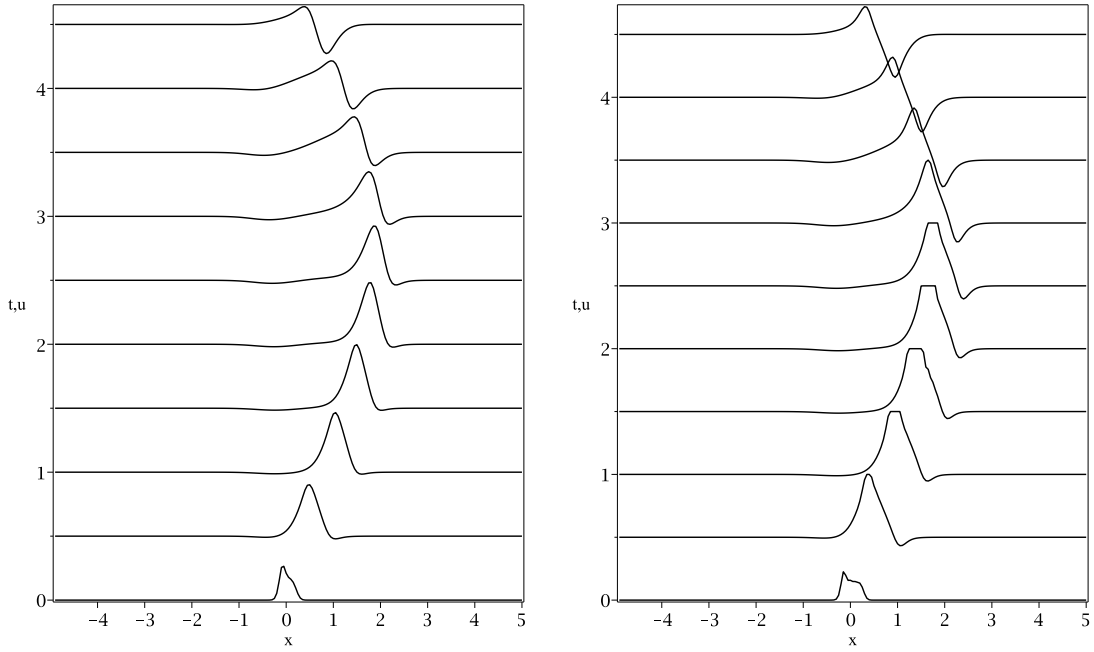
In conclusion, space-dependent control is superior to space-independent control. The desired PDF can be tracked with virtually no deviation. Also, obtaining the optimal control by solving the optimality systems is a viable and efficient option, keeping the computation



(a) $\alpha = 1$ (b) $\alpha = 10$

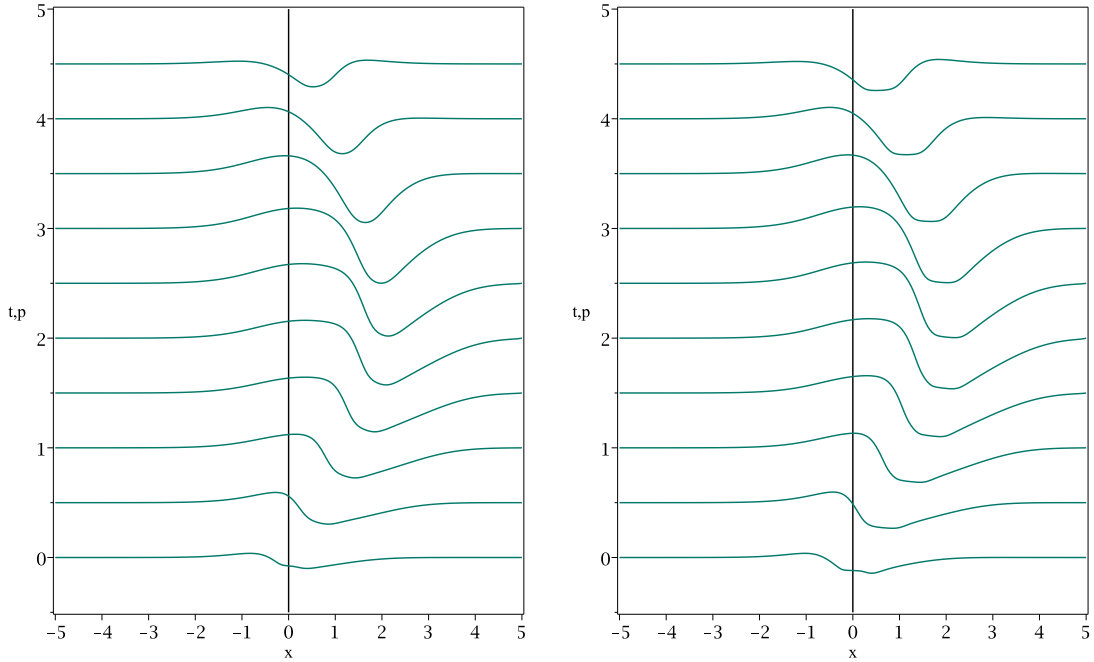
Figure 6.35: Optimal space-dependent control of the Ornstein-Uhlenbeck process for $N = 10$ and varying α: The computed PDF y, the initial PDF y_0, and the desired PDF y_d.

times at acceptable levels.



(a) $\alpha = 1, \max|u(x,t)| = 3.8357$ (b) $\alpha = 10, \max|u(x,t)| = 4$

Figure 6.36: Optimal space-dependent control of the Ornstein-Uhlenbeck process for $N = 10$ and varying α: The optimal control function.



(a) $\alpha = 1, \max|p(x,t)| = 0.41233$ (b) $\alpha = 10, \max|p(x,t)| = 1.11771$

Figure 6.37: Optimal space-dependent control of the Ornstein-Uhlenbeck process for $N = 10$ and varying α: The adjoint state p.

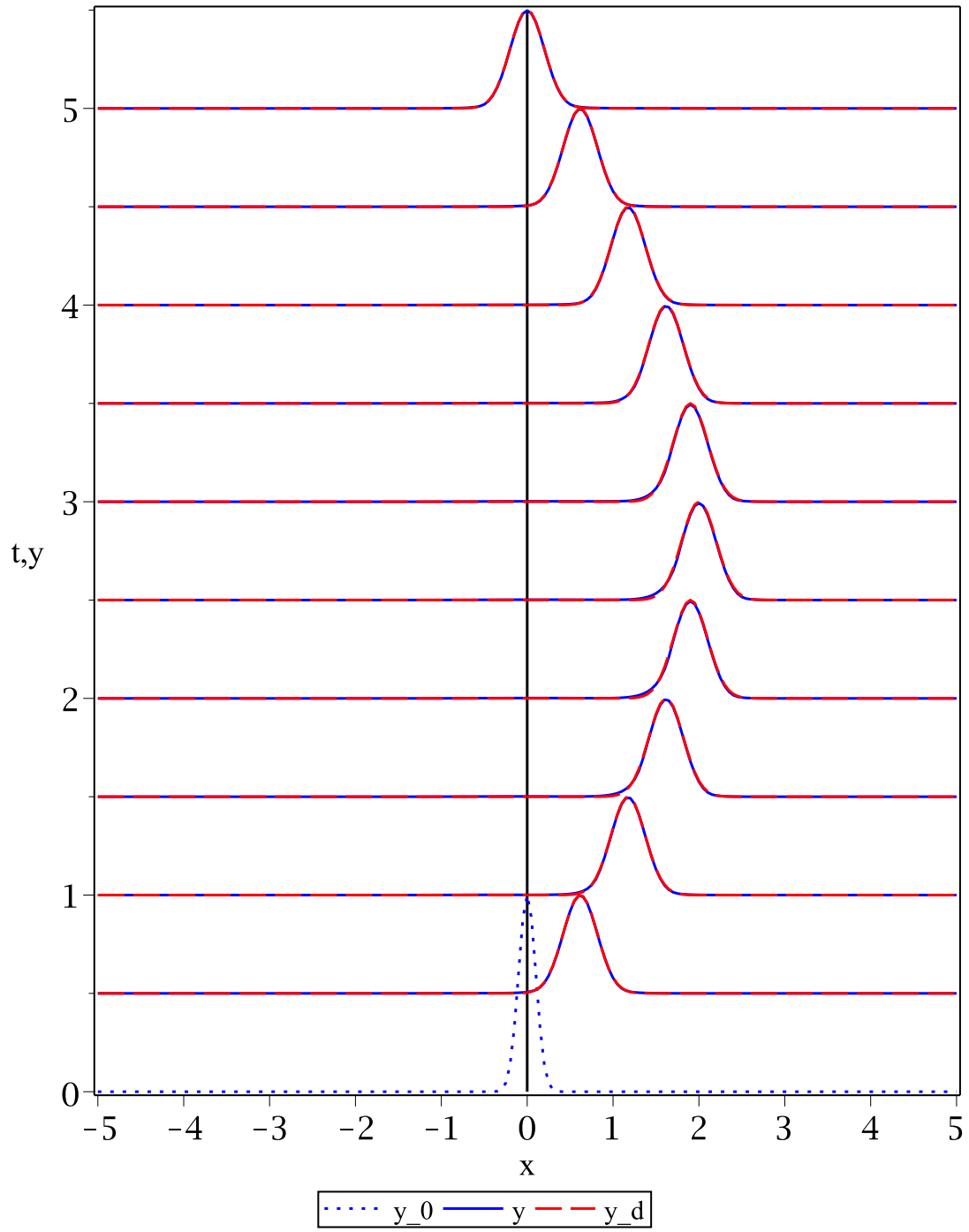


Figure 6.38: Optimal space-dependent control of the Ornstein-Uhlenbeck process for $N = 1, \alpha = 100$, and $\lambda = 0.01$ (num.): The computed PDF y, the initial PDF y_0, and the desired PDF y_d.

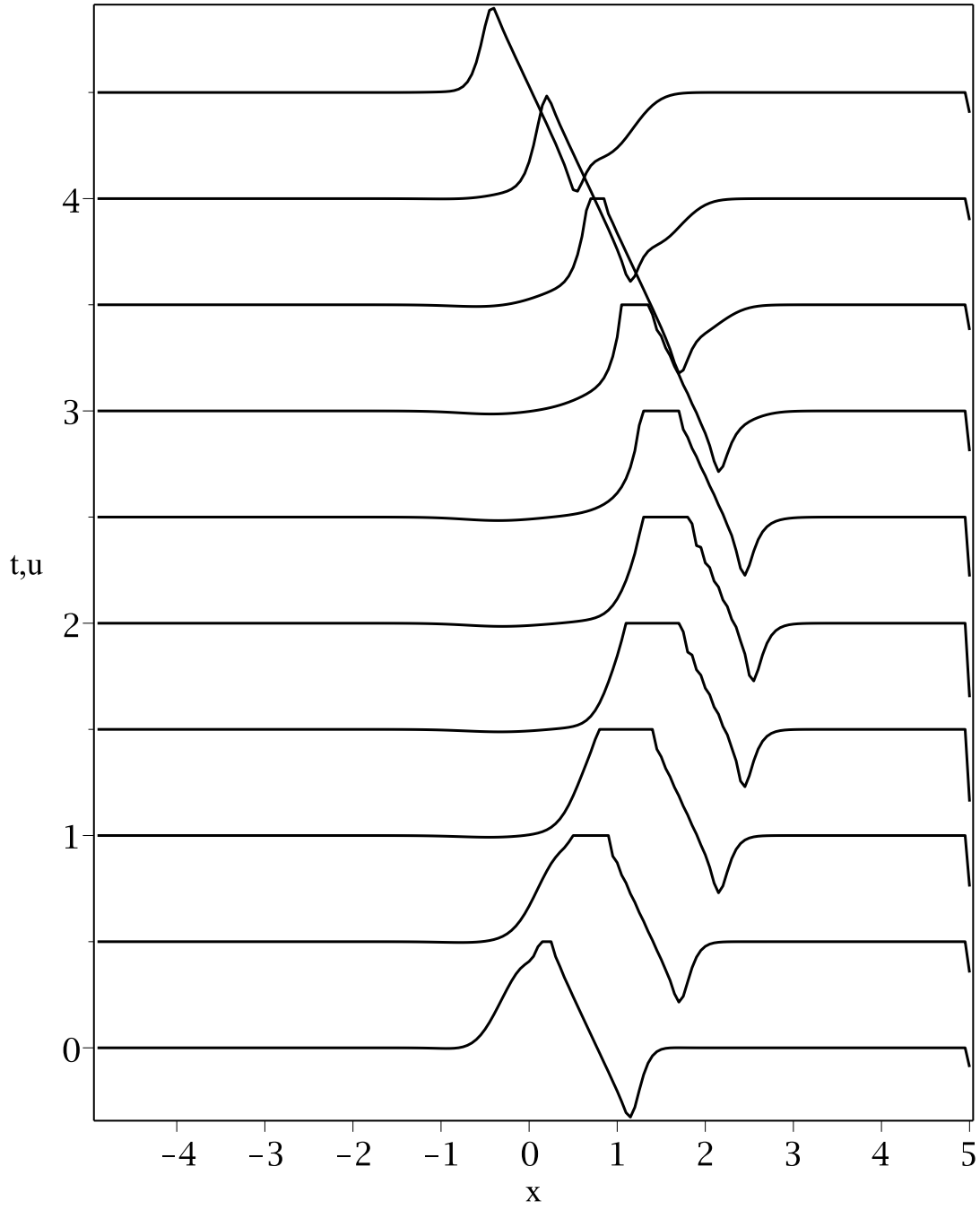


Figure 6.39: Optimal space-dependent control of the Ornstein-Uhlenbeck process for $N = 1, \alpha = 100$, and $\lambda = 0.01$ (num.): The optimal control function ($\max |u(x,t)| = 4$).

6.2.2 Geometric-Brownian Process with Additive Drift Control

The Fokker-Planck equation for the Geometric-Brownian process and space-dependent control is given by

$$\begin{aligned} 0 &= \partial_t y(x,t) - \frac{1}{2}\partial_{xx}^2\left((\tilde{\sigma}x)^2 y(x,t)\right) + \partial_x\big((\mu+u)x\ y(x,t)\big) \\ &= \partial_t y(x,t) - \frac{1}{2}\tilde{\sigma}^2\partial_{xx}^2\left(x^2 y(x,t)\right) + \partial_x\big((\mu+u)x\ y(x,t)\big) \end{aligned} \tag{6.4}$$

with $\mu \in \mathbb{R}$ and $\tilde{\sigma} > 0$. Notice that in contrast to (6.2), $\mu + u$ cannot be factored out. Unless specifically mentioned, the default values of the occurring parameters are the same as in the space-independent case, cf. Table 6.7.

This process is much harder to handle numerically than the Ornstein-Uhlenbeck process. In fact, employing the BFGS algorithm with the first few iterations being computed by the Projected Gradient method often results in errors when trying to obtain gradient information by solving the adjoint equation. This happened for various configurations and linesearch parameters. Therefore, the first numerical results are gained by computing the gradient of the cost functional by means of Remark 4.6. The results for $N = 1$ and $N = 2$ are depicted in Figures 6.40 and 6.41. In case of $N = 1$ the error tolerance opt_{tol} needed to be reduced from $1e-5$ to $5e-5$ in order to achieve convergence (within a reasonable number of iterations). In both cases, the desired probability density function y_d is tracked well for $t \geq 1.5$. Afterwards, the computed PDF y deviates more and more from the desired one. Furthermore, it is unclear whether conservation of probability is ensured at all times. In fact, this is not the case:

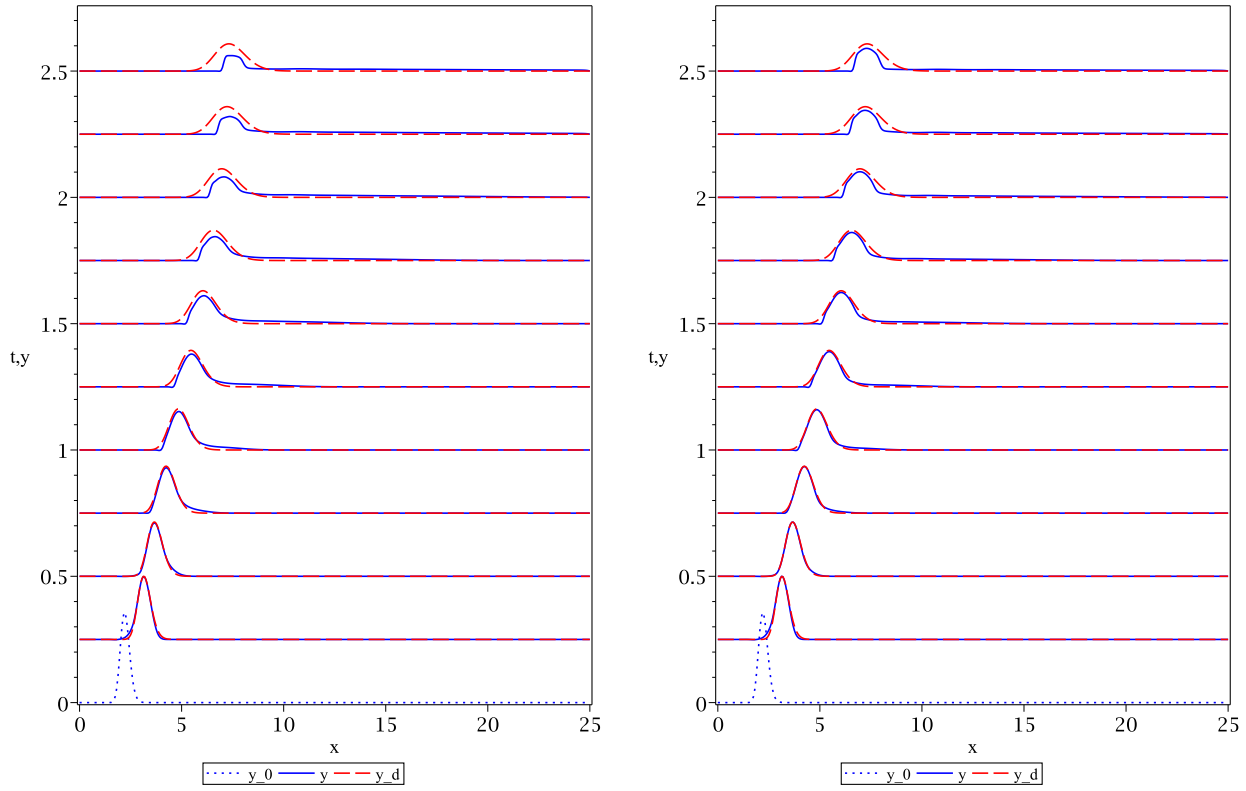
$$\int_\Omega y(x,T_E)\,dx \approx \begin{cases} 0.8337715555 & \text{if } N = 1, \\ 0.8875854545 & \text{if } N = 2. \end{cases}$$

A larger domain Ω would help ensuring this property. However, it would not improve the tracking results. Yet, increasing the horizon seems to help improve these results somewhat, cf. Table 6.18. An additional simulation for $N = 5$ has been run with a reduced error tolerance $tol_{opt} = 5e-5$. It shows a small raise in the costs compared to $N = 2$, which is probably due to the consideration of the desired PDF at times $t > T_E$.

To examine the effects for various α and lower T_s values, the gradient needs to be obtained efficiently, i.e. by solving the adjoint problem. A combination of the BFGS method and the Projected Gradient method did not converge. Thus, solely the Projected Gradient method is employed. In particular, the stopping criteria are different, cf. Chapter 5. Figures 6.42-6.44 show the computed PDF y, the initial PDF y_0, the target PDF y_d, the computed optimal control u, and the adjoint state p for $T_s = 0.025$, $N = 10$, and various α values.

Although Figures 6.42(a) and 6.40(a) look alike, the conservation of probability is satisfied in the former case:

$$\int_\Omega y(x,T_E)\,dx \approx \begin{cases} 0.9999999795 & \text{if } \alpha = 1, \\ 1.000000369 & \text{if } \alpha = 10. \end{cases}$$



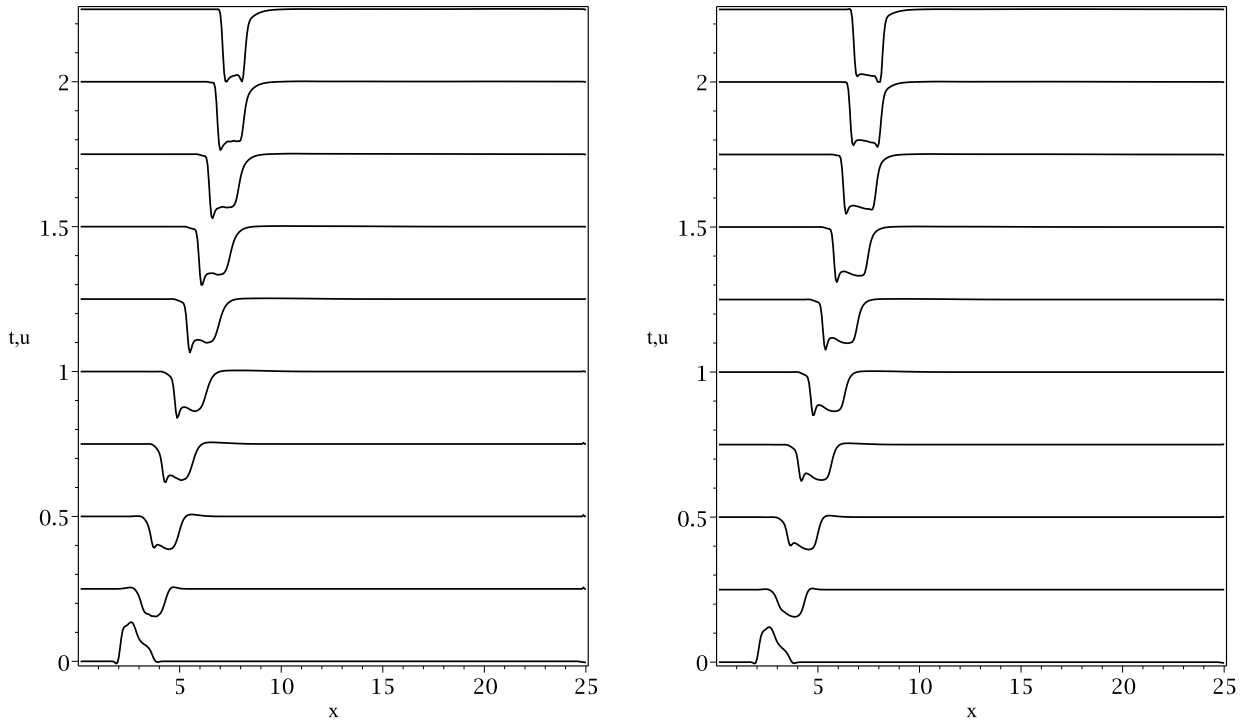
(a) $N = 1$ (b) $N = 2$

Figure 6.40: Optimal space-dependent control of the Geometric-Brownian process for varying N (num.): The computed PDF y, the initial PDF y_0, and the desired PDF y_d.

N	Computation of the gradient	$C_{total,w}$
1	adj.	n/a
1	num.	0.125933
2	adj.	n/a
2	num.	0.103104
5	adj.	n/a
5	num.	0.106761

Table 6.18: Optimal space-dependent control of the Geometric-Brownian process: Weighted total costs for various N.

Furthermore, these figures indicate once more that the derived optimality system (2.60)-(2.62) is correct and solving it yields reasonable results. As in the Ornstein-Uhlenbeck process, the control costs limit the tracking of the desired PDF. By weighting the state costs by a high enough factor, tracking results can be improved significantly, such that the computed PDF barely deviates from the desired one, cf. Figure 6.42(b). Compared

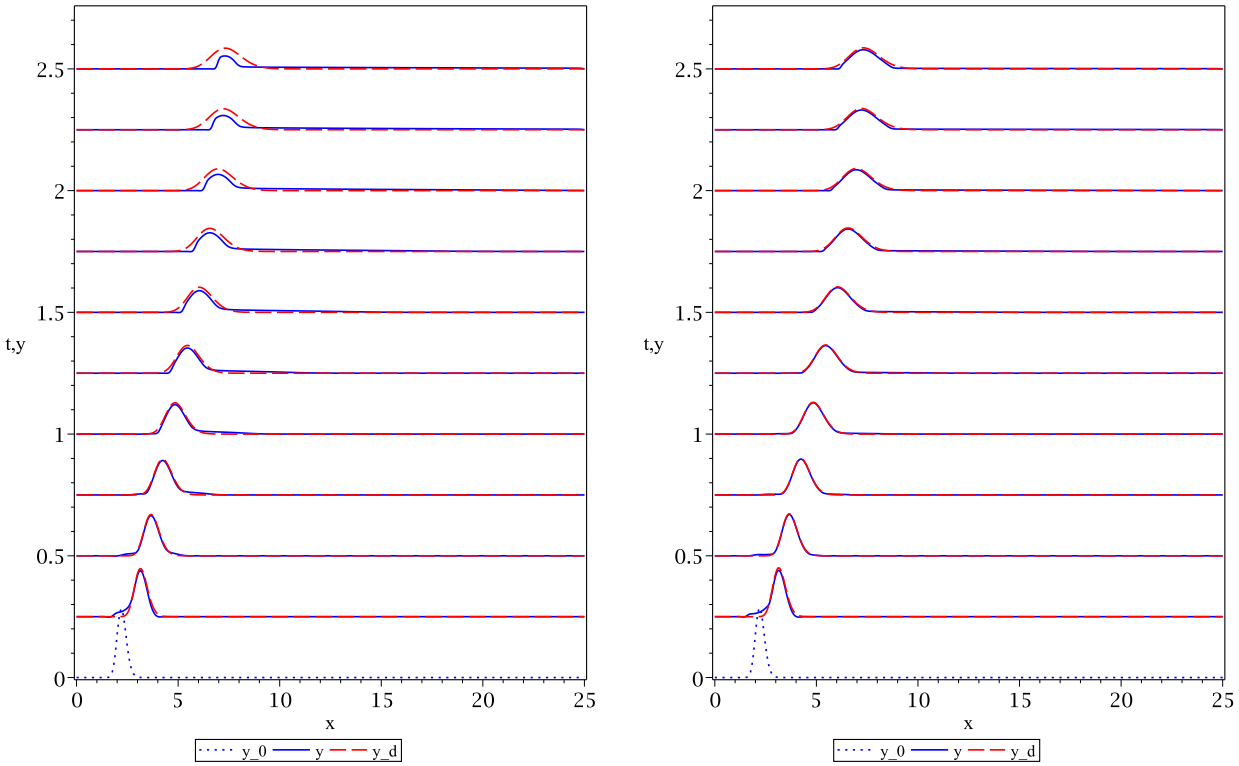


(a) $N = 1, \max |u(x,t)| = 0.920015$ (b) $N = 2, \max |u(x,t)| = 1$

Figure 6.41: Optimal space-dependent control of the Geometric-Brownian process for varying N (num.): The optimal control function.

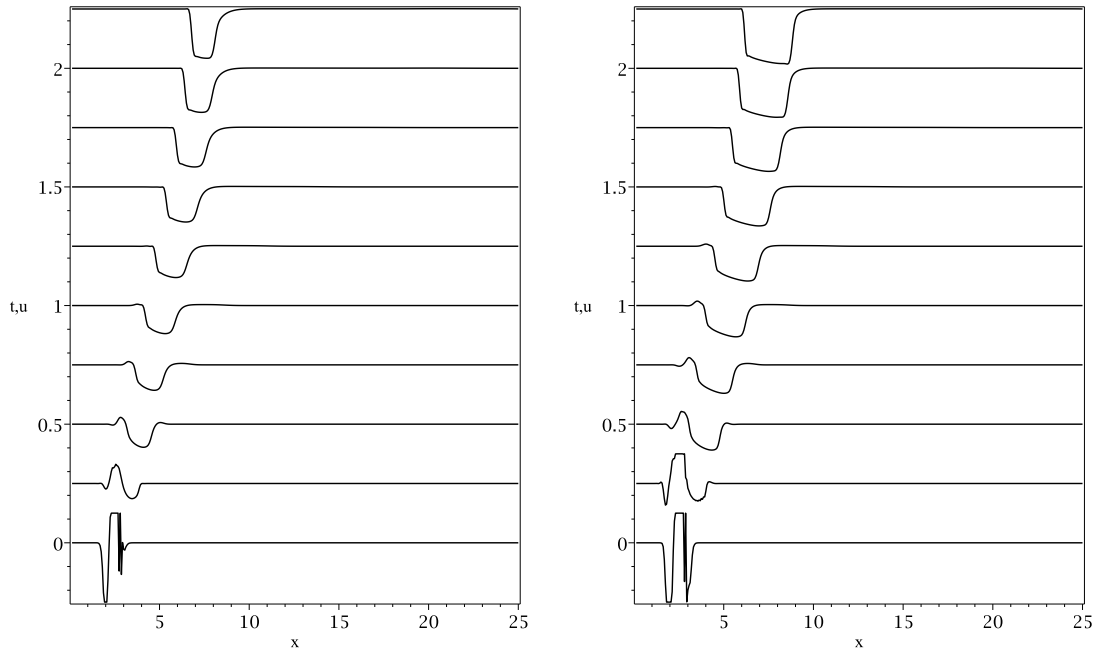
to space-independent control, much better tracking is achieved, cf. e.g. Figure 6.16(b).[3] The controls in Figure 6.43 highlight the (numerical) difficulties of the Geometric-Brownian process. Especially for $t = 0$, the control seems irregular, not even continuous. As time passes, however, the control function seems almost smooth. This is resembled by the state in Figure 6.42(b): for $t = 0.25$, the desired PDF is not attained as well as for later times. Maybe the results can be improved further by more sophisticated algorithms and linesearch strategies in particular. However, with these kinds of control functions, it is questionable whether they can be used in a stochastic setting as they are clearly not Lipschitz-continuous, at least not for $t = 0$. The steep descents at $t = T_E$ add nothing good to that cause. Yet, looking from the PDE-constrained optimization perspective, satisfactory results have been obtained.

[3] A numerical experiment with space-independent control and $\alpha = 10$ resulted in virtually the same computed PDF as in Figure 6.16(b).



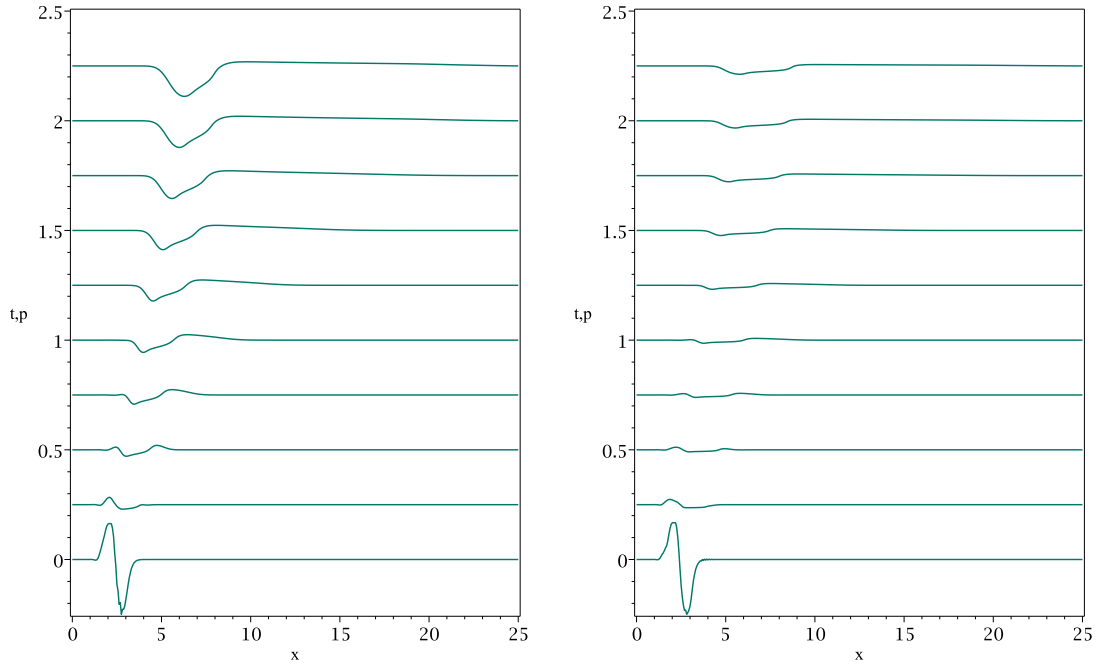
(a) $\alpha = 1$ (b) $\alpha = 10$

Figure 6.42: Optimal space-dependent control of the Geometric-Brownian process for $N = 10$ and varying α: The computed PDF y, the initial PDF y_0, and the desired PDF y_d.



(a) $\alpha = 1, \max|u(x,t)| = 1$ (b) $\alpha = 10, \max|u(x,t)| = 1$

Figure 6.43: Optimal space-dependent control of the Geometric-Brownian process for $N = 10$ and varying α: The optimal control function.

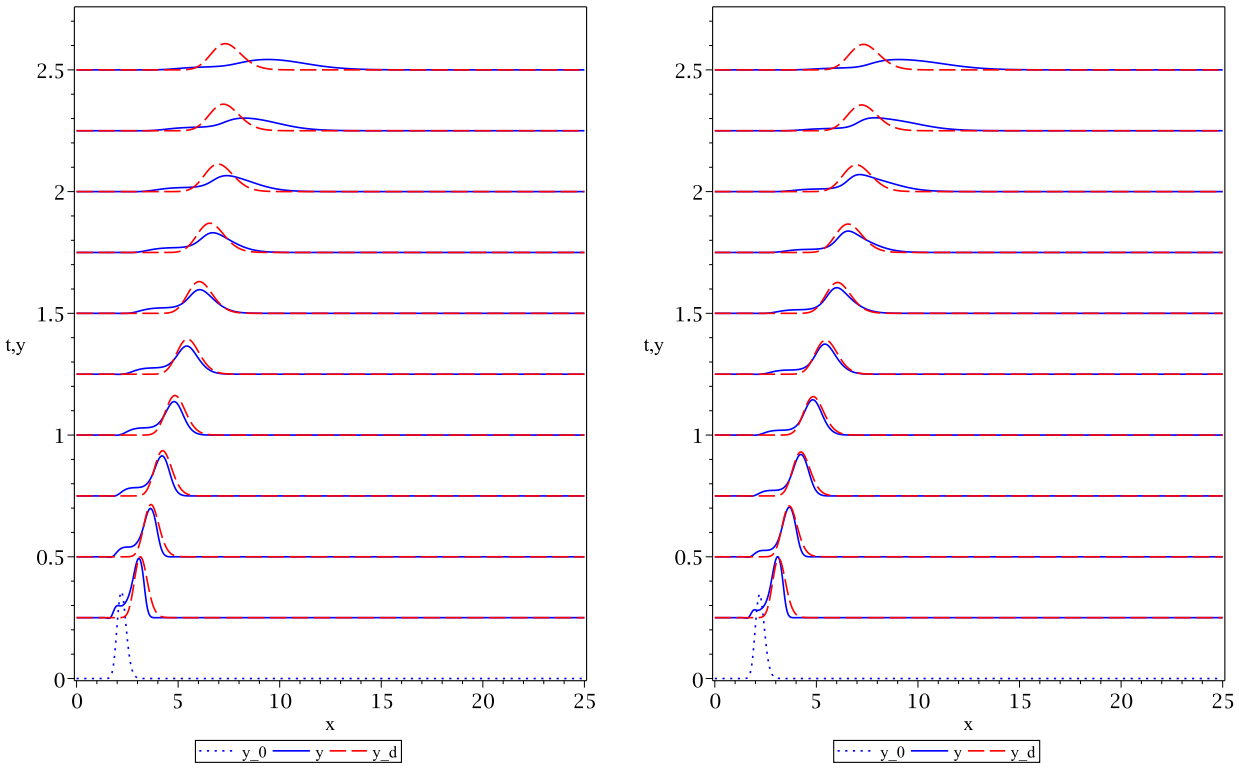


(a) $\alpha = 1, \max|p(x,t)| = 0.140537$ (b) $\alpha = 10, \max|p(x,t)| = 1.30928$

Figure 6.44: Optimal space-dependent control of the Geometric-Brownian process for $N = 10$ and varying α: The adjoint state p.

6.2.3 The Shiryaev Process

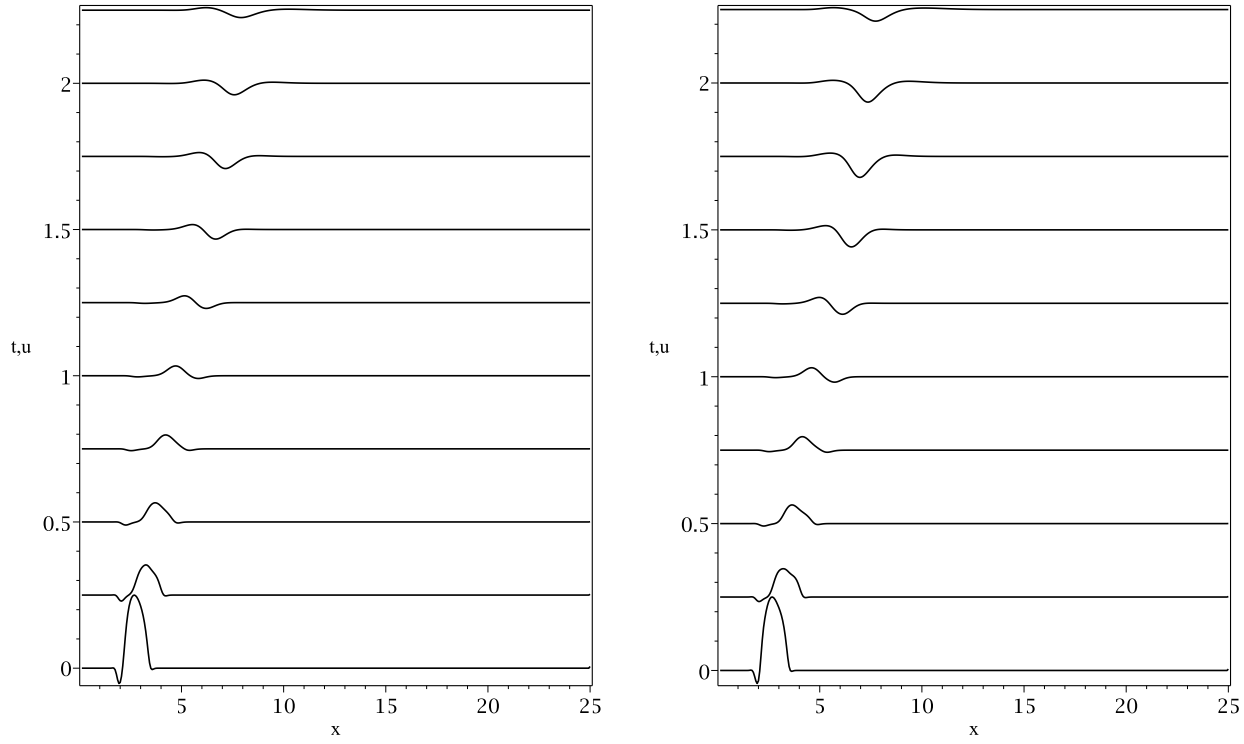
The last process considered is the Shiryaev process. The corresponding Fokker-Planck equation coincides with the one for space-independent control, cf. (6.3). Once again, the default parameter values are the same as in the space-independent case, cf. Subsection 6.1.3, unless stated otherwise.
Similar to the Geometric-Brownian process with space-dependent control, the usual approach, i.e. solving the optimality system (2.56)-(2.58) with a combination of the BFGS method and the Projected Gradient method failed; various configurations have been used. Therefore, the first results are gained by calculating the gradient of the reduced cost functional as in Remark 4.6. These are illustrated in Figures 6.45-6.46.



(a) $N = 1$ (b) $N = 2$

Figure 6.45: Optimal space-dependent control of the Shiryaev process for varying N (num.): The computed PDF y, the initial PDF y_0, and the desired PDF y_d.

Neither the control nor the computed PDF change significantly when increasing N from 1 to 2. Table 6.19 backs the figures; an increase in N leads to only slight improvements in the costs. However, the diffusion at the terminal time $t = T_E$ is not prevented. Moreover, the computed PDFs now have a tail on the left, which seems even worse than the space-independent control results, cf. e.g. Figure 6.19. It remains to be seen whether better tracking results can be achieved. To this end, as in the previous subsections, the α value is increased next. The following results have been obtained by employing only the Projected Gradient



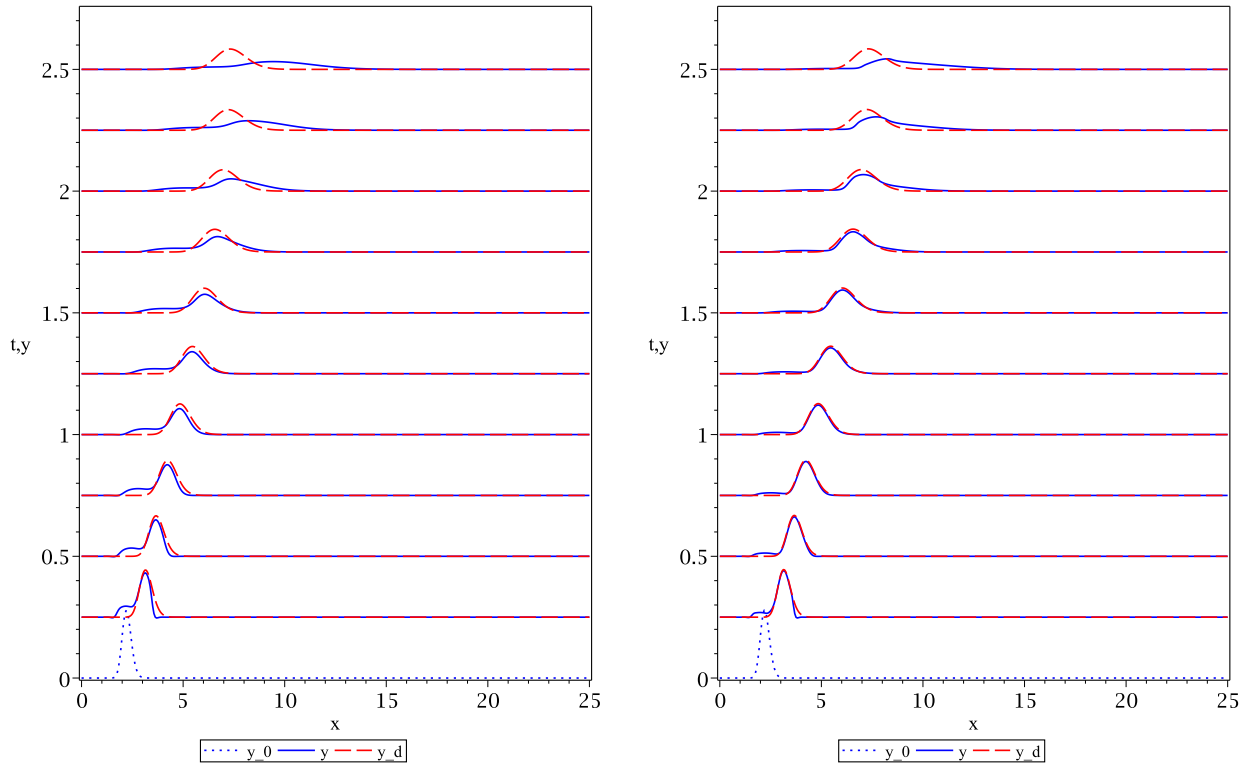
(a) $N = 1, \max|u(x,t)| = 1.86649$ (b) $N = 2, \max|u(x,t)| = 2.05304$

Figure 6.46: Optimal space-dependent control of the Shiryaev process for varying N (num.): The optimal control function.

N	Computation of the gradient	$C_{total,w}$
1	adj.	n/a
1	num.	0.193211
2	adj.	n/a
2	num.	0.1744
5	adj.	n/a
5	num.	0.165513

Table 6.19: Optimal space-dependent control of the Shiryaev process: Weighted total costs for various N.

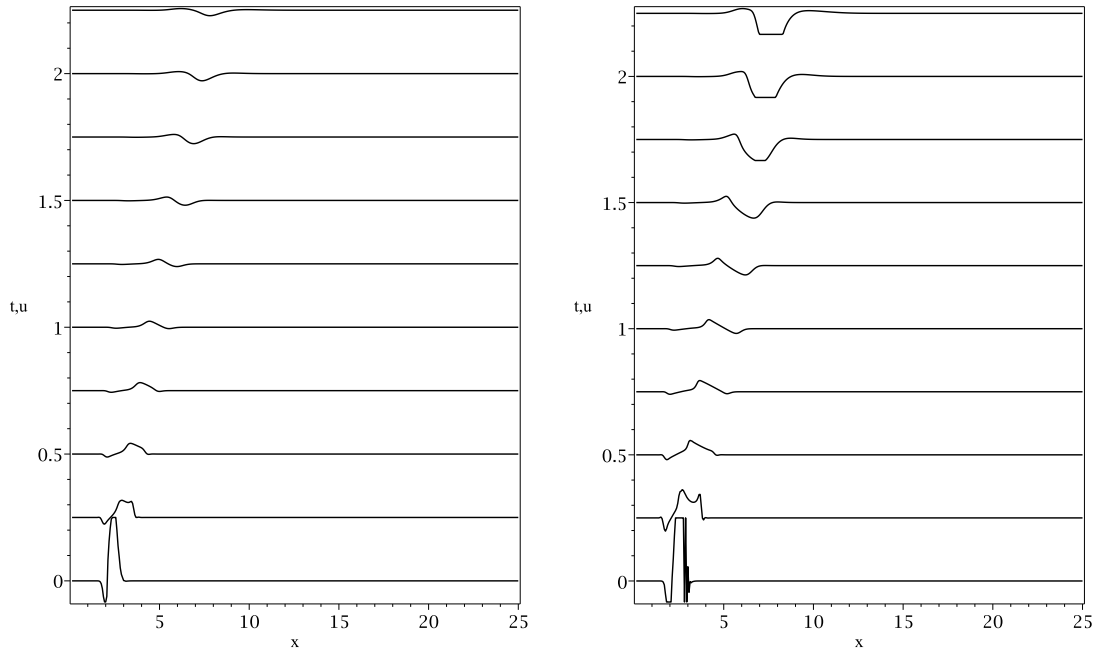
method. Also, the solution is obtained by solving the corresponding optimality system (2.60)-(2.62).
Figures 6.47-6.49 display the numerical results for $T_s = 0.025$, $N = 10$, and varying α values. First of all, the results for $\alpha = 1$ are qualitatively the same as for $T_s = 0.25$ and $N = 1$, cf. Figure 6.45(a). Like in the previous stochastic processes, this fortifies the assumption that the derived optimality system (2.60)-(2.62) characterizes (local) minima. As before, the target PDF is attained better for $\alpha = 10$. Aside from the little tails on the left, the



(a) $\alpha = 1$ (b) $\alpha = 10$

Figure 6.47: Optimal space-dependent control of the Shiryaev process for $N = 10$ and varying α: The computed PDF y, the initial PDF y_0, and the desired PDF y_d.

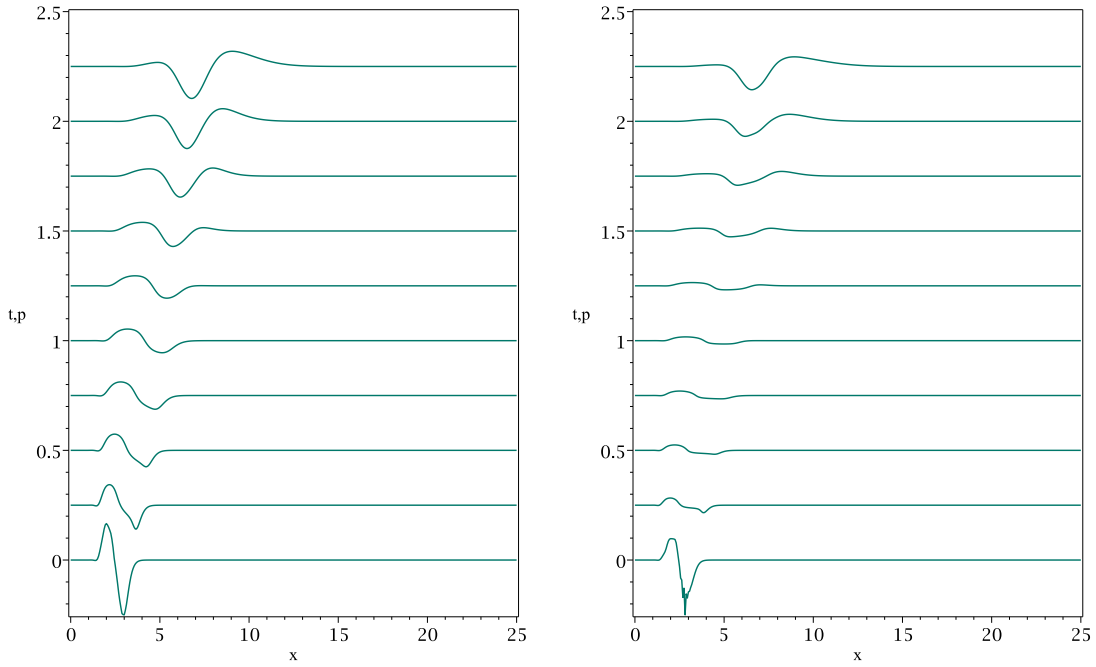
computed PDF for $\alpha = 10$ provides good tracking results that, in some respects, are better than the ones obtained with space-independent control, cf. Figure 6.25(b). In particular, the peak value of y coincides with that of y_d for a longer period of time. However, at $t = T_E$, the computed PDF strides away from the desired PDF. Looking at the controls in Figure 6.48, the limiting factor is the set of admissible controls U_{ad}: The lower bound $u_a = -1$ is maxed out. The same behavior was observed in the space-independent case. Reducing u_a to -3 helped in that matter: While the diffusion remained a problem, at least the peak values of y and y_0 were at the same position x. Consequently, the same measure should lead to better results for space-dependent control as well. Before this experiment is conducted, however, it should be noted that the control in Figure 6.48(b) is non-continuous for $t = 0$, just as in the Geometric-Brownian case. In particular, the control does not meet the requirements of the theorems concerning the existence and uniqueness of solutions of the Fokker-Planck equation, cf. Subsection 2.2.1. Yet, the solution gained by numerical simulation looks meaningful. The reason for the irregularity of the control probably lies in the irregularities appearing in the adjoint state p for $t = 0$, cf. Figure 6.49(b). If these occur due to numerical reasons, a more sophisticated or finer discretization of the adjoint state might help getting rid of the jumps in p and result in a more regular control function.



(a) $\alpha = 1, \max|u(x,t)| = 3$ (b) $\alpha = 10, \max|u(x,t)| = 3$

Figure 6.48: Optimal space-dependent control of the Shiryaev process for $N = 10$ and varying α: The optimal control function.

Eventually, the lower bound u_a is lowered to -4 to see whether the desired probability density function can be attained better, especially for $t = T_E$. As in the Ornstein-Uhlenbeck process, the last simulation is run with $\alpha = 100, \lambda = 0.01, N = 1$, and $T_s = 0.25$ to see once again whether the shortest possible horizon is sufficient to provide satisfactory results under these conditions. The gradient of the reduced cost functional is obtained by means of Remark 4.6. The preferred combination of the BFGS and the Projected Gradient method does converge in this case. Figure 6.50 shows the computed PDF y, the initial PDF y_0, and the desired PDF y_d. The corresponding control is illustrated in Figure 6.51. Just like in the Ornstein-Uhlenbeck case, the computed PDF virtually coincides with the desired one. With the lowest control value being $u(x,t) = -3.98719$, the additional freedom due to the reduced lower bound $u_a = -4$ is almost used up completely, cf. Figure 6.51.
The conclusion is analogous to those in the two previous subsections: Space-dependent control can help achieving much better tracking results, i.e. the probability density function virtually does not deviate from the target PDF. Also, solving the optimality system (2.60)-(2.62) in order to find the optimal control is a viable option. If the gradient of the reduced cost functional can be obtained by solving the corresponding adjoint problem, the computation times are acceptable.



(a) $\alpha = 1, \max|p(x,t)| = 0.167978$ (b) $\alpha = 10, \max|p(x,t)| = 1.95255$

Figure 6.49: Optimal space-dependent control of the Shiryaev process for $N = 10$ and varying α: The adjoint state p.

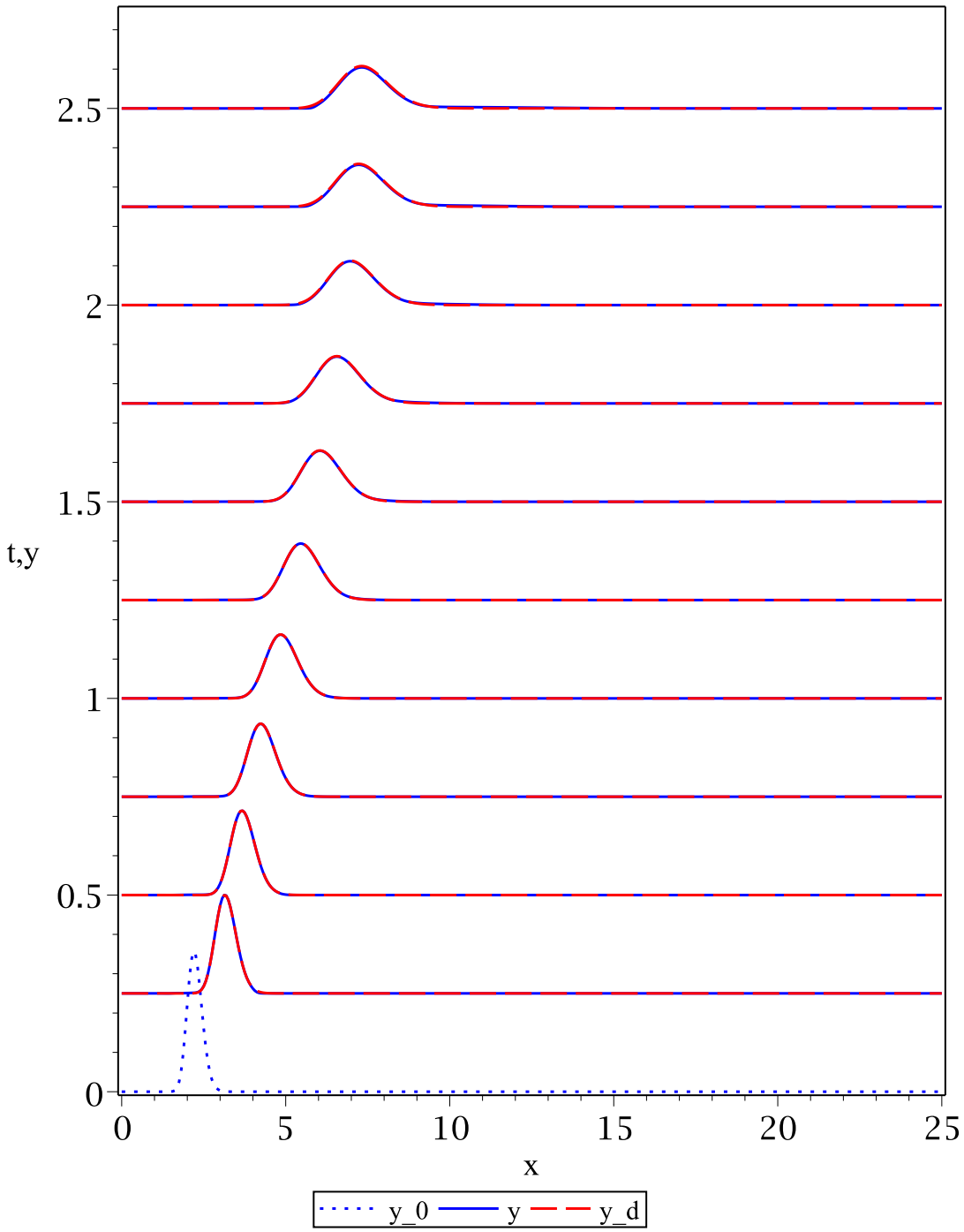


Figure 6.50: Optimal space-dependent control of the Shiryaev process for $N = 1, \alpha = 100, u_a = -4$, and $\lambda = 0.01$ (num.): The computed PDF y, the initial PDF y_0, and the desired PDF y_d.

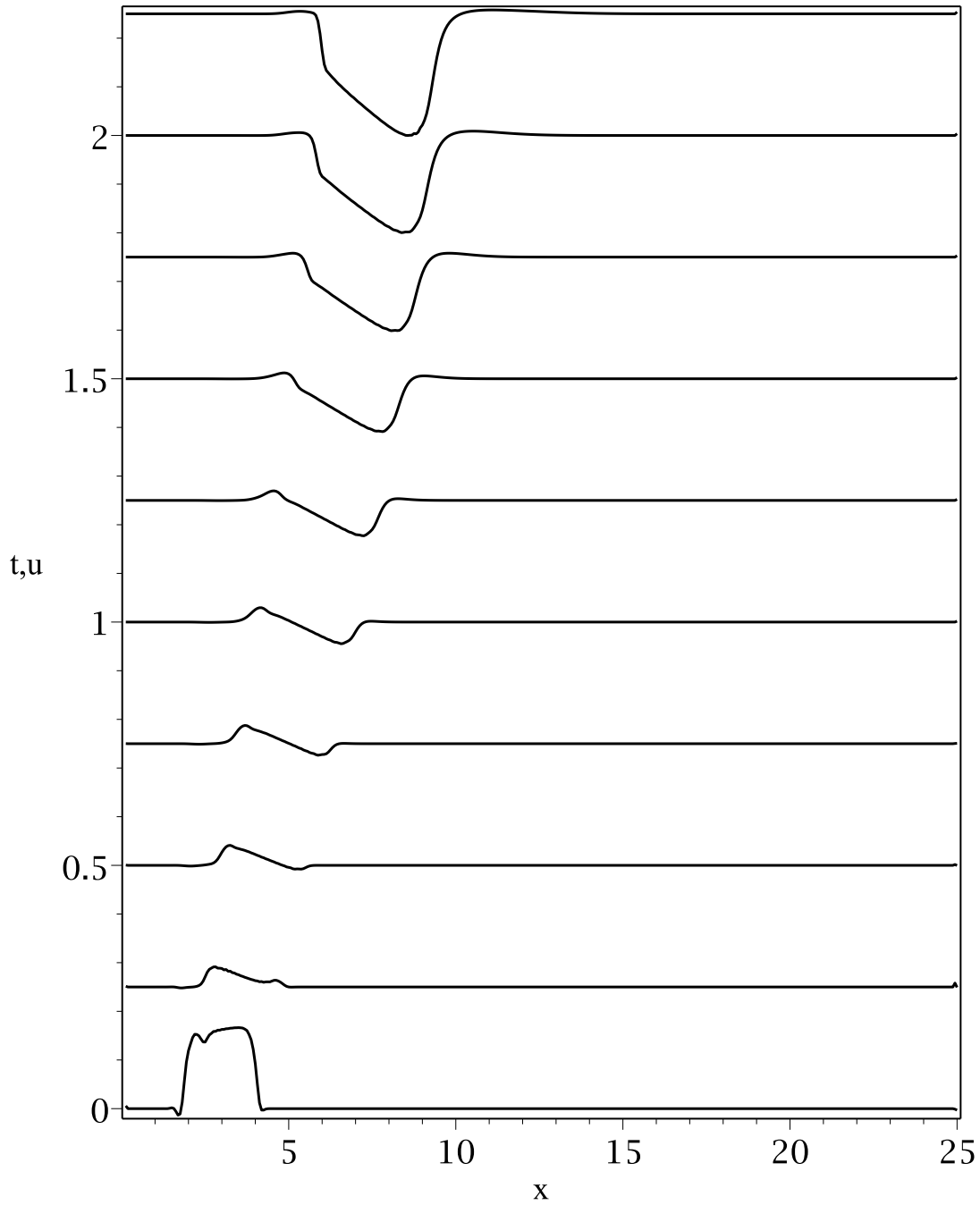


Figure 6.51: Optimal space-dependent control of the Shiryaev process for $N = 1, \alpha = 100, u_a = -4$, and $\lambda = 0.01$ (num.): The optimal control function ($\max |u(x,t)| = 3.98719$).

Chapter 7

Concluding Remarks

The aim of this thesis was to contribute to the Fokker-Planck control framework for stochastic processes introduced by Annunziato and Borzì [3], [4], by studying the impact of the horizon parameter in the MPC scheme and by examining the effects of employing a space-dependent control instead of a space-independent one. To this end, existence and uniqueness results concerning the Fokker-Planck equation for space-dependent controls have been provided. Necessary optimality conditions were derived using the formal Lagrange technique, resulting in an optimality system for each of the considered optimal control problems. These conditions were not derived rigorously, i.e. from the theoretical point of view, it remains unclear whether these conditions lead to a stationary point, not to mention a (local) minimum. However, the numerical results suggest that these optimality systems are indeed suitable to find an optimal control.

In order to be able to run the numerical simulations, discretization and optimization methods have been implemented along with the MPC scheme. The outcome is a program, which, due to its modular structure, is easily expandable, such that new problems can be tackled and more sophisticated algorithms can be added to the program in order to deal with numerical difficulties.

The three stochastic processes considered in this work are the Ornstein-Uhlenbeck Process with Additive Control, the Geometric-Brownian Process with Additive Drift Control, and the Shiryaev Process. For each of these processes, numerical simulations have been run with space-independent and space-dependent control. In all of these cases, increasing the horizon improved the results significantly only under certain circumstances. As long as a certain time interval was captured in each MPC step, reasonable results were obtained. Thus, it seems that the combination of the sampling rate T_s and the MPC horizon N is more important than the horizon parameter by itself. Even though only the shortest possible horizon is considered in [3], combining it with the high value of T_s yielded results in [3] which are hard to improve with space-independent control. Yet, the desired probability density functions are rarely attained by the computed PDFs. This is due to the too high diffusion of the computed probability density functions compared to the diffusion of the desired ones.

Significantly better results can be achieved with a space-dependent control, as the diffusion can then be handled. For every stochastic process, a setting was found in which a (near)

perfect tracking is obtained, often even with the shortest possible MPC horizon. Even so, this comes at the cost of (much) higher computation times, which highlights the importance of adjoint-based methods to compute the gradient of the (reduced) cost functional. Doing so, only two PDEs need to be solved in each iteration of the optimization algorithm. If the gradient is obtained by analyzing the effects of perturbed controls, however, a multitude of PDEs is to be solved in each iteration of the optimization algorithm. Especially for space-dependent control, computation times might quickly reach unacceptable levels. To this end, different algorithms and/or more efficient linesearch strategies might reduce the computing time significantly.

Several open questions and tasks remain regarding the Fokker-Planck optimal control framework. First of all, existence and uniqueness of optimal controls have not been proven theoretically. If this step is undertaken successfully, the next step is to rigorously derive necessary optimality conditions, and, if possible, sufficient optimality conditions. Another challenging problem is to study the relation of the MPC sampling rate and the MPC horizon analytically. The question is whether the numerical findings can be forecast by theoretical results.

Nevertheless, the results in this thesis provide a good starting point in regard of these problems.

Appendix A

Preliminaries regarding PDE theory

When trying to prove existence and uniqueness results regarding solutions of PDEs, so-called classical solutions often do not exist in practical relevant situations. These are solutions that are sufficiently smooth, e.g. twice continuously differentiable with respect to x and continuously differentiable w.r.t. t in the Fokker-Planck case (2.7)-(2.9). However, weaker solution concepts can be used to provide satisfactory results. To this end, several function spaces primarily discussed in functional analysis are utilized. This section tries to very briefly present the concept of weak solutions and the corresponding spaces, especially in light of optimal control. It is assumed that basics about real vector spaces, norms and linear operators are known to the reader. For more information and details, [30] and [16] are recommended; the information for this part was taken or adapted from these works.
First, the functional spaces that are key to the theory concerning PDEs are introduced.

Definition A.1 (Banach space)
A normed real vector space $\{X, \|\cdot\|\}$ *is called (real)* Banach space $:\Leftrightarrow$ *X is complete, i.e. any Cauchy sequence* $(u_n) \subset X$ *has a limit* $u \in X$*, i.e.:*

$$\textit{If } \lim_{m,n\to\infty} \|u_m - u_n\| = 0 \quad \Rightarrow \quad \exists u \in X : \lim_{n\to\infty} \|u_n - u\| = 0.$$

Definition A.2 (Linear functional, dual space, duality pairing)
Let X be a Banach space.

(a) A continuous linear operator $u^*: X \to \mathbb{R}$, i.e. $u^* \in \mathcal{L}(X, \mathbb{R})$, is called a continuous linear functional *on X.*

(b) Let $u \in X$. The space $X^* := \mathcal{L}(X, \mathbb{R})$ is called dual space *of X. The corresponding norm is given by*

$$\|u^*\|_{X^*} := \sup_{\|u\|_X = 1} |u^*(u)|.$$

(c) Let $u \in X, u^* \in X^*$. The notation

$$\langle u^*, u \rangle_{X^*,X} := u^*(u)$$

is called the duality pairing *of X^* and X.*

Definition A.3 (Hilbert space)
A real vector space H with inner product $(\cdot,\cdot)$, $\{H,(\cdot,\cdot)\}$, and associated norm $\|u\| := \sqrt{(u,u)}$ is called Hilbert space $:\Leftrightarrow$ *H is complete under its norm $\|u\|$.*

Theorem A.4 (Riesz representation theorem)
The dual space H^* of a Hilbert space H is isometric to H itself. More precisely, for every $v \in H$ the linear functional u^* defined by

$$\langle u^*, u\rangle_{H^*,H} := (v,u)_H, \quad u \in H$$

is in H^* with norm $\|u^*\|_{H^*} = \|v\|_H$. Vice versa, for every $u^* \in H^*$ there exists a unique $v \in H$ such that

$$\langle u^*, u\rangle_{H^*,H} = (v,u)_H, \quad u \in H$$

and $\|u^*\|_{H^*} = \|v\|_H$. In particular, H^* can be identified with H.

Proof. Cf. [1, p. 166f] □

Definition A.5 (L^p spaces)
Let $E \subset \mathbb{R}^n$ be a nonempty, bounded and Lebesgue measurable set.
(a) Let $1 \le p < \infty$. $L^p(E)$ is the function space consisting of (equivalence classes of) (Lebesgue) measurable functions $y\colon E \to \mathbb{R}$ with

$$\int_E |y(x)|^p \, dx < \infty.$$

The L^p norm is defined by

$$\|y\|_{L^p(E)} := \left(\int_E |y(x)|^p \, dx \right)^{1/p}.$$

(b) $L^\infty(E)$ is the function space consisting of (equivalence classes of) functions $y\colon E \to \mathbb{R}$ that are uniformly bounded almost everywhere, i.e. uniformly bounded up to a set of measure zero. The norm is defined as follows:

$$\|y\|_{L^\infty(E)} := \operatorname*{ess\,sup}_{x\in E} |y(x)| := \inf_{|F|=0} \left(\sup_{x \in E\setminus F} |y(x)| \right)$$

where $|F|$ is the n-dimensional Lebesgue measure of the set F.
(c) Let $1 \le p \le \infty$. $L^p_{loc}(E)$ is the function space consisting of (equivalence classes of) (Lebesgue) measurable functions $y\colon E \to \mathbb{R}$ with $y \in L^p(K)$ for all compact $K \subseteq E$.

Remark A.6
Let $1 \le p \le \infty$ and $\Omega \subset \mathbb{R}^n$. Then $L^p(\Omega) \subset L^1_{loc}(\Omega)$.

The L^p function spaces play an important role in the PDE theory: They replace $C^k(\bar{\Omega})$, the function space of $k \in \mathbb{N}_0 \cup \{\infty\}$ times continuously partially differentiable functions on open $\Omega \subset \mathbb{R}^n$ that can be extended continuously on $\bar{\Omega}$, the closure of Ω. This is done by introducing weak derivatives, which are motivated by the formula for integration by parts. To this end, the space C_0^k is needed.

Definition A.7 (C_0^k functions and support)
Let $\Omega \subset \mathbb{R}^n$ and $k \in \mathbb{N}_0 \cup \{\infty\}$. The set $C_0^k(\Omega)$ consists of functions in $C^k(\Omega)$ that have compact support in Ω, i.e.

$$C_0^k(\Omega) := \left\{ y \in C^k(\Omega) \,|\, \operatorname{supp}(y) \subset \Omega \text{ is compact} \right\}$$

where

$$\operatorname{supp}(y) := \overline{\{x \in \Omega \,|\, y(x) \neq 0\}}$$

is the support *of y.*

Definition A.8 (Weak partial derivative)
Let $\Omega \subset \mathbb{R}^n$ be open, $y \in L^1_{loc}(\Omega)$ and $\alpha \in \mathbb{N}_0^n$ be a multi index. A function $w \in L^1_{loc}(\Omega)$ is called the α-th weak partial derivative of $y :\Leftrightarrow$

$$\forall v \in C_0^\infty(\Omega) : \int_\Omega y(x) D^\alpha v(x)\, dx = (-1)^{|\alpha|} \int_\Omega w(x) v(x)\, dx.$$

In this case, $D^\alpha y := w$.

Remark A.9
If the α-th weak partial derivative of y exists, it is uniquely determined (up to a set of measure zero), see [16, Lemma 1.5].

A weak partial derivative does not necessarily exist. If it does exist, however, it can certainly belong to spaces like $L^p(\Omega)$ rather than only $L^1_{loc}(\Omega)$. This motivates the following definition.

Definition A.10 (Sobolev space)
Let $\Omega \subset \mathbb{R}^n$ be open. For $k \in \mathbb{N}_0$ and $1 \leq p \leq \infty$, the Sobolev space $W^{k,p}(\Omega)$ *is defined by*

$$W^{k,p}(\Omega) := \left\{ y \in L^p(\Omega) \,|\, y \text{ has weak derivatives } D^\alpha y \in L^p(\Omega) \text{ for all } |\alpha| \leq k \right\}.$$

The corresponding norm is given by

$$\|y\|_{W^{k,p}(\Omega)} := \left(\sum_{|\alpha| \leq k} \|D^\alpha y\|^p_{L^p(\Omega)} \right)^{1/p}, \quad 1 \leq p < \infty,$$

$$\|y\|_{W^{k,\infty}(\Omega)} := \sum_{|\alpha| \leq k} \|D^\alpha y\|_{L^\infty(\Omega)}.$$

Remark A.11
Let $\Omega \subset \mathbb{R}^n$ be open, $k \in \mathbb{N}_0$, and $1 \leq p \leq \infty$.
(a) The Sobolev spaces $W^{k,p}(\Omega)$ are Banach spaces, see [16, Theorem 1.11] and [31].
(b) The space $H^k(\Omega) := W^{k,2}(\Omega)$ ($p = 2$) becomes a Hilbert space by introducing the inner product

$$(y, v)_{H^k(\Omega)} := \sum_{|\alpha| \leq k} (D^\alpha y, D^\alpha v)_{L^2(\Omega)}.$$

In particular,

$$H^1(\Omega) = \left\{ y \in L^2(\Omega) \,\middle|\, D_i y \in L^2(\Omega), i = 1, ..., n \right\}$$

with norm

$$\|y\|_{H^1(\Omega)} = \left(\int_\Omega (y^2 + |\nabla y|^2)\, dx \right)^{1/2}$$

where $|\nabla y|^2 = \sum_{i=1}^{n} (D_i y)^2$ and inner product

$$(y, v)_{H^1(\Omega)} = \int_\Omega y\ v\, dx + \int_\Omega \nabla y \cdot \nabla v\, dx$$

is a Hilbert space, see [16, Theorem 1.11] and [30, p. 23].

In cases where no classical solutions of a PDE exist, the Sobolev spaces allow for resorting to a weaker concept of solutions with the help of weak derivatives. The idea is to take the PDE, multiply it by a so-called test function, integrate over the whole space-time cylinder and use the formula for integration by parts. In the following, this procedure is done step by step for the following problem:

$$\partial_t y(x,t) - \frac{1}{2}\partial^2_{xx} y(x,t) + \partial_x y(x,t) = 0 \qquad \text{in } Q \tag{A.1}$$

$$y(x,0) = \rho(x) \qquad \text{in } \Omega \tag{A.2}$$

$$y(x,t) = 0 \qquad \text{in } \Sigma \tag{A.3}$$

where the space-time cylinder Q is defined by $Q := \Omega \times]0, T_E[$ with $\Omega \subset \mathbb{R}$ being an open interval, i.e. $\Omega :=]x_L, x_R[$, and $T_E > 0$. Σ is given by $\Sigma := \partial\Omega \times]0, T_E[$. $\rho\colon \Omega \to \mathbb{R}$ is given and $y\colon \bar{Q} \to \mathbb{R}$ is the unknown.
First, a classical solution y is assumed, so all integrals below exist. In particular, y is assumed to be continuous on $\bar{Q}$. For simplicity, the arguments are omitted in the following.
The PDE (A.1), a special case of the Fokker-Planck equation (2.7), is multiplied by a smooth test function $v(x,t)$, i.e. $v \in C^1(\bar{Q})$, and integrated over Q:

$$\int_0^{T_E} \int_\Omega \partial_t y\ v\, dx\, dt - \int_0^{T_E} \int_\Omega \frac{1}{2}\partial^2_{xx} y\ v\, dx\, dt + \int_0^{T_E} \int_\Omega \partial_x y\ v\, dx\, dt = 0.$$

After partially integrating the second and third integral w.r.t. x, one gets

$$0 = \int_0^{T_E} \int_\Omega \partial_t y \; v \, dx \, dt - \int_0^{T_E} \left(\frac{1}{2} \partial_x y \; v \right) \Big|_{x_L}^{x_R} dt + \int_0^{T_E} \int_\Omega \frac{1}{2} \partial_x y \; \partial_x v \, dx \, dt + \int_0^{T_E} (y \; v)|_{x_L}^{x_R} \, dt - \int_0^{T_E} \int_\Omega y \; \partial_x v \, dx \, dt.$$

Next, the boundary condition (A.3) is implemented, resulting in

$$\begin{aligned} 0 &= \int_0^{T_E} \int_\Omega \partial_t y \; v \, dx \, dt - \int_0^{T_E} \left(\frac{1}{2} \partial_x y \; v \right) \Big|_{x_L}^{x_R} dt + \int_0^{T_E} \int_\Omega \frac{1}{2} \partial_x y \; \partial_x v \, dx \, dt \\ &\quad + \underbrace{\int_0^{T_E} (y \; v)|_{x_L}^{x_R} \, dt}_{=0} - \int_0^{T_E} \int_\Omega y \; \partial_x v \, dx \, dt \\ &= \int_0^{T_E} \int_\Omega \partial_t y \; v \, dx \, dt - \int_0^{T_E} \left(\frac{1}{2} \partial_x y \; v \right) \Big|_{x_L}^{x_R} dt + \int_0^{T_E} \int_\Omega \left(\frac{1}{2} \partial_x y - y \right) \partial_x v \, dx \, dt. \end{aligned} \tag{A.4}$$

Remark A.12
Using the integration by parts formula for the third integral was not necessary. Since the existence of $\partial_x y$ and $\partial_x v$ is required anyway, an additional $\partial_x y$ or $\partial_x v$ term does not influence the regularity of y or v.

The main advantage of this new formulation is that only first-order derivatives of y are required – and only in the weak sense, cf. Definition A.8. However, several problems arise:

1. In the above derivation, a classical solution y has been used, allowing for simply incorporating the boundary conditions (A.3). However, the aim is to obtain a meaningful formulation for functions in Sobolev spaces, i.e. functions which are uniquely determined only up to a set of measure zero. The question therefore is how to interpret the boundary conditions in the new setting.

2. While the derivatives w.r.t. x are symmetrically distributed between the solution y and the test function v, it is unclear whether to keep $\partial_t y$ or use the formula for integration by parts to obtain $\partial_t v$ instead. In either case, one function needs to be more regular than the other.

Both issues are tackled in the following. It is helpful to focus on the space domain Ω first. A possible way to deal with $\partial\Omega$ at least for homogeneous boundary conditions is to incorporate them into the function space.

Definition and Remark A.13 (The $W_0^{k,p}(\Omega)$ space)
Let $\Omega \subset \mathbb{R}^n$ be open, $k \in \mathbb{N}_0$, and $1 \leq p \leq \infty$.
(a) Denote by $W_0^{k,p}(\Omega)$ the closure of $C_0^\infty(\Omega)$ in $W^{k,p}(\Omega)$, i.e. for any $y \in W_0^{k,p}(\Omega)$ there exists a sequence $(v_i) \subset C_0^\infty(\Omega)$ with $\lim_{i\to\infty} \|y - v_i\|_{W^{k,p}(\Omega)} = 0$. The norm is the same as in $W^{k,p}(\Omega)$.
(b) $W_0^{k,p}(\Omega)$ is a Banach space, cf. [31], and $H_0^k(\Omega) := W_0^{k,2}(\Omega)$ is a Hilbert space, see [16, p. 21].

For general boundary values, a requirement on $\Omega \subset \mathbb{R}^n$ and on its boundary $\partial\Omega$ in particular needs to be met.

Definition A.14 ($C^{k,1}$ boundary, Lipschitz domain)
Let $n \in \mathbb{N}$ with $n \geq 2$ and $k \in \mathbb{N}_0$ Further, let $\Omega \subset \mathbb{R}^n$ be a bounded domain with boundary $\partial\Omega$. Ω has a $C^{k,1}$ boundary *:$\Leftrightarrow$ $\exists$ finitely many local coordinate systems $S_1, ..., S_M$, functions $h_1, ..., h_M$ and constants $a, b > 0$:*
(i) All h_i are k times continuously differentiable with Lipschitz continuous derivatives of order k on the closed $(n-1)$-dimensional cube

$$\bar{Q}_{n-1} := \left\{ y = (y_1, ..., y_{n-1}) \in \mathbb{R}^{n-1} \mid |y_i| \leq a, i = 1, ..., n-1 \right\}$$

(ii) $\forall P \in \partial\Omega\ \exists i \in \{1, ..., M\}$: P can be written as $P = (y, h_i(y)), y \in Q_{n-1}$ in the coordinate system S_i.
(iii) In the local coordinate system S_i, the following holds:

$$(y, y_n) \in \Omega \Leftrightarrow y \in \bar{Q}_{n-1}, h_i(y) < y_n < h_i(y) + b$$
$$(y, y_n) \notin \Omega \Leftrightarrow y \in \bar{Q}_{n-1}, h_i(y) - b < y_n < h_i(y),$$

i.e. locally, Ω lies only on one side of the boundary.
If $k = 0$, Ω has a Lipschitz boundary *and is called* Lipschitz domain.

Theorem A.15 (Trace Theorem)
Let $\Omega \subset \mathbb{R}^n$ be a bounded Lipschitz domain. Further, let $1 \leq p \leq \infty$. Then there exists a unique bounded linear operator $\tau \colon W^{1,p}(\Omega) \to L^p(\partial\Omega)$ such that

$$\forall y \in W^{1,p}(\Omega) \cap C(\bar{\Omega}) : (\tau y)(x) = y(x) \text{ for almost all } x \in \partial\Omega.$$

Proof. Cf. [1], [7], or [31]. □

Definition and Remark A.16 (Trace)
(a) τy is called the trace *of y on $\partial\Omega$.*
(b) $(\tau y)(x)$ coincides with $y_{|\partial\Omega}$ on $\partial\Omega$ for continuous functions y. For simplicity, instead of τy, often the notation $y_{|\partial\Omega}$ is used.

The Trace Theorem gives a meaning to the values on $\partial\Omega$ for functions that are uniquely determined only up to a set of measure zero. The next step is to deal with the time dimension. For parabolic PDEs, the common function spaces are the $W_2^{1,0}(Q)$ and $W_2^{1,1}(Q)$ spaces.

Definition and Remark A.17 (The $W_2^{1,0}(Q)$ space)
Let $\Omega \subset \mathbb{R}^n$ be open, $T_E > 0$, and $Q := \Omega \times]0, T_E[$.
(a) The space $W_2^{1,0}(Q)$ is defined by

$$W_2^{1,0}(Q) := \left\{ y \in L^2(Q) \,\middle|\, D_i y \in L^2(Q),\ i = 1, ..., n \right\},$$

where $D_i y$ is the first weak derivative of y w.r.t. x_i. The corresponding norm is given by

$$\|y\|_{W_2^{1,0}(Q)} := \left(\int_0^{T_E} \int_\Omega \left(y(x,t)^2 + |\nabla_x y(x,t)|^2 \right) dx\, dt \right)^{1/2},$$

where ∇_x is the gradient w.r.t. x. In this context, the subscript x often is omitted, i.e. ∇ is written instead of ∇_x.
(b) By introducing the inner product that arises from the above norm, $W_2^{1,0}(Q)$ becomes a Hilbert space, see [30, p. 111]. $W_2^{1,0}(Q)$ is therefore often denoted by $H^{1,0}(Q)$.

Definition and Remark A.18 (The $W_2^{1,1}(Q)$ space)
Let $\Omega \subset \mathbb{R}^n$ be open, $T_E > 0$, and $Q := \Omega \times]0, T_E[$.
(a) The space $W_2^{1,1}(Q)$ is defined by

$$W_2^{1,1}(Q) := \left\{ y \in L^2(Q) \,\middle|\, y_t \in L^2(Q) \text{ and } D_i y \in L^2(Q),\ i = 1, ..., n \right\},$$

where $D_i y$ is the first weak derivative of y w.r.t. x_i. The corresponding norm is given by

$$\|y\|_{W_2^{1,1}(Q)} := \left(\int_0^{T_E} \int_\Omega \left(y(x,t)^2 + |\nabla_x y(x,t)|^2 + y_t(x,t)^2 \right) dx\, dt \right)^{1/2},$$

where ∇_x is the gradient w.r.t. x. In this context, the subscript x often is omitted, i.e. ∇ is written instead of ∇_x.
(b) By introducing the inner product that arises from the above norm, $W_2^{1,1}(Q)$ becomes a Hilbert space, see [30, p. 111]. $W_2^{1,1}(Q)$ is therefore often denoted by $H^{1,1}(Q)$.

The difference between these two spaces obviously is the derivative w.r.t. time t. The usual approach in PDE theory is to use the formula for integration by parts to transfer the time derivative to the test function. Consequently, starting from (A.4) and using the initial

condition (A.2), one gets

$$
\begin{aligned}
0 &= \int_0^{T_E}\int_\Omega \partial_t y\ v\,dx\,dt - \int_0^{T_E}\left(\frac{1}{2}\partial_x y\ v\right)\Big|_{x_L}^{x_R} dt + \int_0^{T_E}\int_\Omega \left(\frac{1}{2}\partial_x y - y\right)\partial_x v\,dx\,dt \\
&= \int_\Omega (y\ v)|_0^{T_E}\ dx - \int_0^{T_E}\int_\Omega y\ \partial_t v\,dx\,dt - \int_0^{T_E}\left(\frac{1}{2}\partial_x y\ v\right)\Big|_{x_L}^{x_R} dt \\
&+ \int_0^{T_E}\int_\Omega \left(\frac{1}{2}\partial_x y - y\right)\partial_x v\,dx\,dt \\
&= \int_\Omega y(x,T_E)v(x,T_E)\,dx - \int_\Omega \underbrace{y(x,0)}_{=\rho(x)} v(x,0)\,dx - \int_0^{T_E}\left(\frac{1}{2}\partial_x y\ v\right)\Big|_{x_L}^{x_R} dt \qquad \text{(A.5)} \\
&+ \int_0^{T_E}\int_\Omega \left(\frac{1}{2}\partial_x y - y\right)\partial_x v - y\ \partial_t v\,dx\,dt \\
&= \int_\Omega y(x,T_E)v(x,T_E)\,dx - \int_\Omega \rho(x)v(x,0)\,dx - \int_0^{T_E}\left(\frac{1}{2}\partial_x y\ v\right)\Big|_{x_L}^{x_R} dt \\
&- \int_0^{T_E}\int_\Omega y\ \partial_t v - \left(\frac{1}{2}\partial_x y - y\right)\partial_x v\,dx\,dt.
\end{aligned}
$$

As before, a classical solution y is assumed so that all above integrals exist. This way, the first so-called variational formulation of (A.1)-(A.3) has been derived. However, the aim is to obtain a meaningful variational formulation even for functions $y \in W_2^{1,0}(Q)$, which are not necessarily continuous, i.e. $y(x,0)$ and $y(x,T_E)$ are not necessarily well-defined. For $y(x,0)$ the initial condition (A.2) can be plugged in.[1] However, the term $y(x,T_E)$ needs to be handled as well. A trick helps to get rid of $y(x,T_E)$: The test function is assumed to be smooth, i.e. $v \in C^1(\bar{Q})$. However, the above steps are valid also for $v \in W_2^{1,1}(Q)$. $v(x,0)$ and $v(x,T_E)$ are still well-defined – as traces in $L^2(\Omega)$, see [20]. Simply asking $v(x,T_E) = 0$ is therefore meaningful and eliminates $y(x,T_E)$.
If the focus is on PDE theory, i.e. on existence and uniqueness results, the spaces $W_2^{1,0}(Q)$ and $W_2^{1,1}(Q)$ and the above procedure are adequate. However, when it comes to optimal control, the test function will become the so-called adjoint state. Requiring the adjoint to be more regular than the solution y is implausible, as is the demand for $v(\cdot,T_E) = 0$. Therefore, a slightly different approach is needed, in which y is considered as a Banach space valued function

$$[0,T_E] \ni t \mapsto y(t) \in H_0^1(\Omega).$$

[1] It remains to show that this condition makes sense.

To this end, the notion of so-called abstract functions and their integrability and weak differentiability is necessary. Only abstract functions that are defined on a compact interval $[a,b] \subset \mathbb{R}$ are considered.

Definition A.19 (Abstract function)
A mapping from $[a,b] \subset \mathbb{R}$ to a Banach space X is called an abstract function.

Example A.20
Let $X := H^1(\Omega)$ and $y\colon [0,T_E] \to H^1(\Omega)$. Then for each $t \in [0,T_E]$, $y(t)$ is an element in $H^1(\Omega)$, i.e. a function depending on $x \in \Omega$. Therefore, for fixed $t \in [0,T_E]$, $y(t) = y(\cdot,t)$ is an element in $H^1(\Omega)$.

In order to ensure that $y(\cdot,0)$ and $y(\cdot,T_E)$ are well-defined, continuity of the abstract function y is needed.

Definition A.21 (Continuity for abstract functions)
Let $\{X, \|\cdot\|_X\}$ be a real Banach space.
(a) An abstract function $y\colon [a,b] \to X$ is continuous in $t \in [a,b]$ $:\Leftrightarrow$

$$\lim_{\tau \to t} \|y(\tau) - y(t)\|_X = 0 \text{ for } \tau \in [a,b].$$

(b) The Banach space consisting of all continuous (abstract) functions $y\colon [a,b] \to X$ is denoted by $C([a,b],X)$. The norm is given by

$$\|y\|_{C([a,b],X)} := \max_{t \in [a,b]} \|y(t)\|_X .$$

Analogously, the space $C^k([a,b],X), k \in \mathbb{N}$, consists of k times continuously differentiable functions on $[a,b]$.
The next concern is integrability. Similar to real-valued functions, L^p spaces consisting of (equivalence classes of) abstract functions can be established.

Definition A.22 (Step function)
Let $y\colon [a,b] \to X$ be an abstract function, $[a,b] = \bigcup_{i=1}^{m} M_i, m < \infty, M_i \subset [a,b]$ Lebesgue-measurable with $M_i \cap M_j = \emptyset$ for $i \neq j$, and $y_i \in X, i = 1,...,m$. y is called step function $:\Leftrightarrow \forall t \in M_i, i = 1,...,m : y(t) = y_i$.
In this case, y can be written as

$$y(t) = \sum_{i=1}^{m} \mathbb{1}_{M_i}(t)\ y_i,$$

where $\mathbb{1}_{M_i}$ is the characteristic function on M_i, i.e.

$$\mathbb{1}_{M_i}(t) := \begin{cases} 1 & \text{if } t \in M_i \\ 0 & \text{otherwise} \end{cases}.$$

Definition A.23 (Measurable function)
An abstract function $y\colon [a,b] \to X$ *is called* measurable $:\Leftrightarrow$
$\exists \{y_k\}_{k=1}^{\infty}$ *sequence of step functions* $y_k\colon [a,b] \to X : y(t) = \lim\limits_{k\to\infty} y_k(t)$ *for almost all* $t \in [a,b]$.

Definition A.24 ($L^p(a,b;X)$)
Let X *be a Banach space.*
(a) Let $1 \le p < \infty$. $L^p(a,b;X)$ *is the linear space consisting of (equivalence classes of) measurable abstract functions* $y\colon [a,b] \to X$ *with*

$$\int_a^b \|y(t)\|_X^p \, dt < \infty.$$

The corresponding norm is defined by

$$\|y\|_{L^p(a,b;X)} := \left(\int_a^b \|y(t)\|_X^p \, dt \right)^{1/p}.$$

(b) $L^\infty(a,b;X)$ *is the linear space consisting of (equivalence classes of) measurable abstract functions* $y\colon [a,b] \to X$ *with*

$$\|y\|_{L^\infty(a,b;X)} := \operatorname*{ess\,sup}_{[a,b]} \|y(t)\|_X < \infty.$$

Remark A.25
(a) Identical to the usual L^p *spaces (cf. Definition A.5), functions in* $L^p(a,b;X)$ *that differ only on sets of measure zero belong to the same equivalence class, i.e.* $y(a)$ *and* $y(b)$ *are not well-defined.*
(b) For $1 \le p \le \infty$, $L^p(a,b;X)$ *and* $C([a,b],X)$ *are Banach spaces and* $L^1(a,b;X) \supset \ldots \supset L^\infty(a,b;X) \supset C([a,b],X)$.

Example A.26
Let $\Omega \subset \mathbb{R}^n$ *be open,* $T_E > 0$ *and* $Q := \Omega \times]0,T_E[$. *Consider the space* $L^2(0,T_E;H^1(\Omega))$. *On the one hand, a function* $y \in L^2(0,T_E;H^1(\Omega))$ *is an abstract function: For every* $t \in [0,T_E]$, $y(t)$ *is an element in* $H^1(\Omega)$. *On the other hand,* y *can be interpreted as a real-valued function depending on* x *and* t, *i.e.* $y = y(x,t)$. *The norm, given by*

$$\|y\|_{L^2(0,T_E;H^1(\Omega))} = \left(\int_0^{T_E} \|y(t)\|_{H^1(\Omega)}^2 \, dt \right)^{1/2} = \left(\int_0^{T_E} \int_\Omega y(x,t)^2 + |\nabla_x y(x,t)|^2 \, dx \, dt \right)^{1/2},$$

is the same as the norm for the $W_2^{1,0}(Q)$ *space, cf. Definition A.17. Unsurprisingly, a function* $y \in W_2^{1,0}(Q)$, *after possibly changing the function values on a set of measure zero, belongs to* $L^2(0,T_E;H^1(\Omega))$ *and vice versa, i.e.*

$$W_2^{1,0}(Q) \cong L^2(0,T_E;H^1(\Omega)),$$

cf. [30, p. 115].

While the spaces $W_2^{1,0}(Q)$ and $L^2(0,T_E;H^1(\Omega))$ are isometrically isomorphic, the space of abstract functions makes way for a slightly different approach: Instead of integrating (A.1) over Q, the PDE is integrated only over Ω, while the time t is fixed. Consequently, the test function (A.1) is multiplied by is now $v \in H^1(\Omega)$. The dependence on x is omitted in the following derivation. First, one gets

$$\int_\Omega \partial_t y(t)\ v\,dx = \int_\Omega \frac{1}{2}\partial_{xx}^2 y(t)\ v\,dx - \int_\Omega \partial_x y(t)\ v\,dx.$$

Partially integrating the first integral on the right-hand side[2] results in

$$\begin{aligned}\int_\Omega \partial_t y(t)\ v\,dx &= \left(\frac{1}{2}\partial_x y(t)\ v\right)\Big|_{x_L}^{x_R} - \int_\Omega \frac{1}{2}\partial_x y(t)\ \partial_x v\,dx - \int_\Omega \partial_x y(t)\ v\,dx \\ &= \left(\frac{1}{2}\partial_x y(t)\ v\right)\Big|_{x_L}^{x_R} - \int_\Omega \frac{1}{2}\partial_x y(t)\ \partial_x v + \partial_x y(t)\ v\,dx\end{aligned} \tag{A.6}$$

for almost all fixed $t \in [0,T_E]$. However, the above derivation is only formal: the meaning of $\partial_t y(t)$ is still unclear. As above, only the existence of the above integrals is required. In particular, the time derivative of y needs to exist only in some weak sense, similar to weak partial derivatives in Definition A.8, only for abstract functions. It is no longer needed to require $v(x,T_E) = 0$. To check whether the new approach yields any advantages, the meaning of $\partial_t y(t)$ needs to be explained. In particular, the notion of weak partial derivatives for abstract functions needs to be established. Therefore, the next step is to concern oneself with integrability of abstract functions in $L^p(a,b;X)$. To this end, the so-called Bochner integral is introduced. For $X = \mathbb{R}$, this is the analogon of the Lebesgue integral.

Definition A.27 (Bochner integral)

(a) Let $y\colon [a,b] \to \mathbb{R}$ *be a step function. The* Bochner integral (for step functions) *is then defined by*

$$\int_a^b y(t)\,dt := \sum_{i=1}^m y_i\ \mu(M_i),$$

where $\mu(M_i)$ *is the Lebesgue measure of the set* M_i*. The integral is an element in* X*.*

(b) An abstract function $y\colon [a,b] \to \mathbb{R}$ *is* Bochner-integrable $:\Leftrightarrow$ $\exists \{y_k\}_{k=1}^\infty$ *sequence of step functions such that* $y_k(t) \to y(t)$ *for almost all* $t \in [a,b]$ *and*

$$\int_a^b \|y_k(t) - y(t)\|_X\ dt \to 0 \text{ as } k \to \infty.$$

[2] Again, the decision whether to use the integration by parts formula for the whole right-hand side or not was arbitrary, cf. Remark A.12.

(c) Let y be Bochner-integrable. The Bochner integral *is then defined by*

$$\int_a^b y(t)\,dt := \lim_{k\to\infty} \int_a^b y_k(t)\,dt.$$

The Bochner integral is used to define the weak derivative for abstract functions.

Definition A.28 (Weak derivative (for abstract functions))
Let $y \in L^1(a,b;X)$. $v \in L^1(a,b;X)$ is the weak derivative of y *:⇔*

$$\forall \varphi \in C_0^\infty(]a,b[) : \int_a^b \varphi'(t)\ y(t)\,dt = -\int_a^b \varphi(t)\ v(t)\,dt.$$

In this case, $\partial_t y := v$. In this context, $\partial_t y$ is often denoted by y'. The weak derivative is also called a distributional derivative *or a* derivative in the distributional sense.

Both (A.6) and Definition A.28 give clues for a suitable function space for solutions of (A.1)-(A.3). For elliptic PDEs, $H^1(\Omega)$ and $H_0^1(\Omega)$ are commonly used. As already mentioned, the $W_2^{1,0}(Q)$ space is not a practical choice for optimal control problems that are governed by parabolic PDEs. However, one could say it is pretty close: In Example A.26 it has been shown that

$$W_2^{1,0}(Q) \cong L^2(0,T_E;H^1(\Omega)).$$

It might be tempting to just require $y \in L^2(0,T_E;H^1(\Omega))$. Then from Remark A.25 and Definition A.28 it follows that $\partial_t y \in L^1(0,T_E;H^1(\Omega))$, assuming the weak derivative of the abstract function y exists. However, when looking at (A.6), one realizes that for almost all fixed $t \in [0,T_E]$, every $v \in H^1(\Omega)$ is mapped onto a real number. This mapping is linear and continuous. Therefore, for almost all fixed $t \in [0,T_E]$, the right-hand side of (A.6) is a linear and continuous functional $F = F(t)$ applied to $v \in H^1(\Omega)$, i.e.

$$\int_\Omega \partial_t y(t)\ v\,dx = F(t)v \quad \text{for almost all } t \in [0,T_E].$$

Since $F = F(t) \in H^1(\Omega)^*$, i.e. $F(t)$ is an element of the dual space of $H^1(\Omega)$, the same holds for the left-hand side of (A.6), leading to $\partial_t y(t) \in H^1(\Omega)^*$ for almost all $t \in [0,T_E]$. If one could show

$$\int_0^{T_E} \|F(t)\|_{H^1(\Omega)^*}^2\ dt < \infty,$$

i.e. $F \in L^2(0,T_E;H^1(\Omega)^*)$, then $\partial_t y \in L^2(0,T_E;H^1(\Omega)^*)$ would follow. The question then is how to handle having $\partial_t y \in L^2(0,T_E;H^1(\Omega)^*)$ on the one hand and $\partial_t y \in L^1(0,T_E;H^1(\Omega))$ on the other. This eventually leads to a new space, the so-called $W(0,T_E)$ space. To this end, a sequence of continuous and dense embeddings, called the Gelfand triple, is needed.

Definition A.29 (Gelfand triple)
Let H and V be separable Hilbert spaces with a continuous and dense embedding $V \hookrightarrow H$. According to Riesz (cf. Theorem A.4), H can be identified with its dual H^*. Doing so results in the continuous and dense embeddings

$$V \hookrightarrow H = H^* \hookrightarrow V^*,$$

which is called Gelfand triple. *The embedding $H \hookrightarrow V^*$ is given by*

$$y \in H \mapsto (y, \cdot)_H \in H^* \subset V^*.$$

Example A.30
Let $V := H_0^1(\Omega)$ and $H := L^2(\Omega)$. V and H are separable Hilbert spaces. Each $u \in H$ can be interpreted as an element in V^* by the following mapping:

$$v \mapsto (u, v)_H \in \mathbb{R}, \quad v \in V.$$

Moreover, according to Theorem A.4, the dual space H^* can be identified with H. Therefore, the following embeddings

$$H_0^1(\Omega) \hookrightarrow L^2(\Omega) \hookrightarrow H_0^1(\Omega)^*$$

hold, where each embedding is continuous and dense.

The Gelfand triple helps handling the above situation, i.e. having $\partial_t y = y' \in L^2(0, T_E; H^1(\Omega)^*)$ and $y' \in L^1(0, T_E; H^1(\Omega))$ at the same time. It has been shown that it is possible to regard an abstract function $y \in L^2(0, T_E; H^1(\Omega))$ as an abstract function in $L^2(0, T_E; H^1(\Omega)^*)$. Doing so results in $y' \in L^1(0, T_E; H^1(\Omega)^*)$, cf. Definition A.28. Since $L^1(0, T_E; H^1(\Omega)^*) \supset L^2(0, T_E; H^1(\Omega)^*)$, cf. Remark A.25, requiring $y' \in L^2(0, T_E; H^1(\Omega)^*)$ is more restrictive and is therefore sufficient to guarantee $y' \in L^1(0, T_E; H^1(\Omega)^*)$. For more general V and H, this leads to the $W(0, T_E)$ or, more precisely, the $W(0, T_E; H, V)$ space:

Definition and Remark A.31 (The $W(0, T_E; H, V)$ space)
Let $V \hookrightarrow H = H^* \hookrightarrow V^*$ be a Gelfand triple.
(a) The $W(0, T_E; H, V)$ space *is defined by*

$$W(0, T_E; H, V) := \{y \in L^2(0, T_E; V) \mid y' \in L^2(0, T_E; V^*)\}$$

with norm

$$\|y\|_{W(0,T_E;H,V)} := \left(\int_0^{T_E} \|y(t)\|_V^2 + \|y'(t)\|_{V^*}^2 \, dt \right)^{1/2}.$$

The arguments H and V are often omitted if the choice of V and H is clear. The space is then denoted by $W(0, T_E)$.

(b) The $W(0,T_E)$ space becomes a Hilbert space by introducing the inner product

$$(u,v)_{W(0,T_E)} := \int_0^{T_E} (u(t),v(t))_V \, dt + \int_0^{T_E} (u'(t),v'(t))_{V^*} \, dt.$$

The inner product $(F,G)_{V^*}$ is defined by $(J\ F, J\ G)_V$, where $J\colon V^* \to V$ is the duality mapping which maps every functional $F \in V^*$ onto the function $f \in V$ whereby F is identified, cf. Theorem A.4.

One obstacle that has not yet been overcome is that for $y \in W(0,T_E;H,V)$, it is unknown whether $y(0)$ (and $y(T_E)$) are well-defined. The following theorem takes care of that.

Theorem A.32
Let $V \hookrightarrow H \hookrightarrow V^*$ be a Gelfand triple. Then $W(0,T_E;H,V)$ is continuously embedded in $C([0,T_E],H)$, i.e. every function $y \in W(0,T_E;H,V)$ can be seen as an element of $C([0,T_E],H)$ – after possibly changing the function values on a set of measure zero.

Proof. Cf. [16, Theorem 1.32] and the references therein. □

In the $W(0,T_E)$ space, the integration by parts formula holds:

Theorem A.33
Let $V \hookrightarrow H \hookrightarrow V^*$ be a Gelfand triple and $y,v \in W(0,T_E;H,V)$. Then the integration by parts formula holds:

$$\int_0^{T_E} \langle y'(t),v(t)\rangle_{V^*,V} \, dt = (y(T_E),v(T_E))_H - (y(0),v(0))_H - \int_0^{T_E} \langle v'(t),y(t)\rangle_{V^*,V} \, dt,$$

where $\langle\cdot,\cdot\rangle_{V^*,V}$ is the duality pairing introduced in Definition A.2.

Proof. Cf. [16, Theorem 1.32] and the references therein. □

Corollary A.34
In the situation of Theorem A.33 the following holds:

$$\int_0^{T_E} \langle y'(t),y(t)\rangle_{V^*,V} \, dt = \frac{1}{2}\|y(T_E)\|_H^2 - \frac{1}{2}\|y(0)\|_H^2 .$$

Proof. Choose $v := y$ in Theorem A.33. □

In conclusion, the $W(0,T_E)$ space is a suitable space for solutions of optimal control problems governed by parabolic PDEs. The final part of this chapter is dedicated to define the notion of a weak solution of (uniformly) parabolic PDEs, e.g. of (A.1)-(A.3). This particularly involves finding a weaker, variational formulation of the PDE at hand. Up until now, only

the model problem (A.1)-(A.3) has been looked at. In the thesis, however, the focus is on a more general setting, cf. Definition 2.8. Moreover, defining weak solutions of even more general parabolic PDEs comes at virtually no additional cost. Therefore, the following setting is considered:
Let $\Omega \subset \mathbb{R}^n$ be a bounded domain, $T_E > 0$ and define the usual space-time cylinder $Q := \Omega \times]0, T_E[$. Furthermore, let $\Sigma := \partial\Omega \times]0, T_E[$. Consider the initial boundary value problem

$$\begin{aligned} \partial_t y + L\, y &= f \quad \text{in } Q, \\ y(\cdot, 0) &= y_0 \quad \text{in } \Omega, \\ y &= 0 \quad \text{in } \Sigma, \end{aligned} \tag{A.7}$$

where $f\colon Q \to \mathbb{R}$ and $y_0\colon \Omega \to \mathbb{R}$ are given and $y\colon \bar{Q} \to \mathbb{R}$ is the unknown. For each time $t \in [0, T_E]$, the operator $L = L(t)$ is a second order partial differential operator

$$L\, y := -\sum_{i,j=1}^{n} \partial_{x_j} \left(a_{ij}(x,t)\partial_{x_i} y\right) + \sum_{i=1}^{n} b_i(x,t)\partial_{x_i} y + c_0(x,t)y \tag{A.8}$$

in the so-called divergence form with $a_{ij}, b_i, c_0\colon Q \to \mathbb{R}, i, j = 1, ..., n$.

Definition A.35 (Uniformly Parabolic Operator)
The partial differential operator $\frac{\partial}{\partial t} + L$ *with* L *given in* (A.8) *is called* uniformly parabolic $:\Leftrightarrow \exists \theta > 0$:

$$\sum_{i,j=1}^{n} a_{ij}(x,t)\xi_i\xi_j \geq \theta \left\|\xi\right\|^2 \quad \textit{for almost all } (x,t) \in Q \textit{ and all } \xi \in \mathbb{R}^n.$$

As mentioned above, only uniformly parabolic PDEs are looked at. Additionally, it is assumed that

$$a_{ij}, b_i, c_0 \in L^\infty(Q),\ i, j = 1, ..., n \tag{A.9}$$

and

$$y_0 \in L^2(\Omega), \quad f \in L^2(0, T_E; H_0^1(\Omega)^*). \tag{A.10}$$

Example A.36
If $\rho \in L^2(\Omega)$, *the model problem* (A.1)-(A.3) *is a special case of* (A.7) *with*

$$n = 1, \quad a_{11} \equiv \frac{1}{2}, \quad b_1 \equiv 1, \quad c_0 \equiv 0, \quad f \equiv 0, \quad y_0 = \rho.$$

Obviously, the resulting operator $\frac{\partial}{\partial t} + L$ *is uniformly parabolic and satisfies* (A.9) *and* (A.10).

Weak solutions are searched for in the $W(0, T_E; H, V)$ space with

$$H := L^2(\Omega) \quad \text{and} \quad V := H_0^1(\Omega),$$

where the Gelfand triple is given by

$$H_0^1(\Omega) \hookrightarrow L^2(\Omega) \hookrightarrow H_0^1(\Omega)^*.$$

In (A.6), the time was fixed and instead of integrating over Q, the PDE was integrated only over Ω. This makes sense, since for almost all $t \in [0, T_E]$, (A.7) yields the following boundary value problem:

$$\begin{aligned} L(t)y(t) &= f(t) - \partial_t y(t) && \text{in } \Omega, \\ y(t) &= 0 && \text{in } \partial\Omega. \end{aligned}$$

Integrating by parts w.r.t. x resulted in a promising, yet formal expression (A.6), since the meaning of the derivative w.r.t. time t was not clear. Having solved this problem, the above ansatz is now refined and used for fairly general problems (A.7). In this setting, the integration by parts is expressed by the bilinear form

$$a(y(t), v; t) := \int_\Omega \left(\sum_{i,j=1}^n a_{ij}(t) \partial_{x_i} y(t)\ \partial_{x_j} v + \sum_{i=1}^n b_i(t) \partial_{x_i} y(t)\ v + c_0(t)\ y(t)\ v \right) dx \tag{A.11}$$

with $y(t), v \in H_0^1(\Omega)$, where the dependence on x has been omitted. For (A.1)-(A.3), the bilinear form (A.11) covers the right-hand side of (A.6). Notice that since $y(t), v \in H_0^1(\Omega)$, no additional terms appear. To subsequently incorporate the left-hand side is to require the variational equality

$$a(y(t), v; t) = \langle f(t), v \rangle_{H_0^1(\Omega)^*, H_0^1(\Omega)} - \langle \partial_t y(t), v \rangle_{H_0^1(\Omega)^*, H_0^1(\Omega)}, \quad v \in H_0^1(\Omega) \tag{A.12}$$

to hold for almost all $t \in [0, T_E]$, where $a(\cdot, \cdot; t)$ is given by (A.11). Note that, according to the Riesz representation theorem (Theorem A.4), there exists a unique element in $H_0^1(\Omega)$ such that $\langle \partial_t y(t), v \rangle_{H_0^1(\Omega)^*, H_0^1(\Omega)}$ can be identified using this element. However, this entails using the inner product in $H_0^1(\Omega)$ instead of the inner product in $H = L^2(\Omega)$, which was used in (A.6). However, a functional $F \in V^*$ can be continuously continued on H (starting from the smaller space V) if and only if it can be written as

$$F(v) = (f, v)_H,$$

where $f \in H$ is fixed. In the case at hand this results in

$$F(v) = \int_\Omega f(x)\ v(x)\, dx = (f, v)_{L^2(\Omega)}$$

for all $v \in V = H_0^1(\Omega)$. Therefore,

$$\langle \partial_t y(t), v \rangle_{H_0^1(\Omega)^*, H_0^1(\Omega)} = (\partial_t y(t), v)_{L^2(\Omega)} = \int_\Omega \partial_t y(t)\ v\, dx$$

and, consequently, (A.12) is (A.6) for $v \in H_0^1(\Omega)$ and $y \in W(0, T_E; L^2(\Omega), H_0^1(\Omega))$. This motivates the following definition.

Definition A.37 (Weak solution of a parabolic PDE)
Let $\Omega \subset \mathbb{R}^n$ be a bounded domain and let the coefficients satisfy (A.9)*. Consider with*

$$H := L^2(\Omega) \quad \textit{and} \quad V := H_0^1(\Omega)$$

the Gelfand triple

$$H_0^1(\Omega) \hookrightarrow L^2(\Omega) \hookrightarrow H_0^1(\Omega)^*.$$

Then for f, y_0 satisfying (A.10)*, a function*

$$y \in W(0, T_E; H, V)$$

is a weak solution of the initial boundary value problem (A.7)
$:\Leftrightarrow$ *For almost all $t \in [0, T_E]$, the variational equation*

$$\langle \partial_t y(t), v \rangle_{H_0^1(\Omega)^*, H_0^1(\Omega)} + a(y(t), v; t) = \langle f(t), v \rangle_{H_0^1(\Omega)^*, H_0^1(\Omega)}, \quad v \in V \tag{A.13}$$

with $a(\cdot, \cdot; t)$ defined by (A.11) *holds and the initial condition*

$$y(0) = y_0 \text{ in } \Omega$$

is satisfied.

Instead of requiring the variational equation (A.13) to hold for almost all $t \in [0, T_E]$, it may be more convenient to integrate (A.13) w.r.t. time t and require $v \in L^2(0, T_E; H_0^1(\Omega))$, arriving at the following, equivalent, definition.

Definition A.38 (Weak solution of a parabolic PDE; equivalent formulation)
Let the same assumptions hold as in Definition A.37 and use the same notation. Then $y \in W(0, T_E; H, V)$ is a weak solution of the initial boundary value problem (A.7)
$:\Leftrightarrow \forall v \in L^2(0, T_E, V)$:

$$\int_0^{T_E} \langle \partial_t y(t), v(t) \rangle_{H_0^1(\Omega)^*, H_0^1(\Omega)} \, dt + \int_0^{T_E} a(y(t), v(t); t) \, dt = \int_0^{T_E} \langle f(t), v(t) \rangle_{H_0^1(\Omega)^*, H_0^1(\Omega)} \, dt \tag{A.14}$$

$$y(0) = y_0 \text{ in } \Omega \tag{A.15}$$

with $a(\cdot, \cdot; t)$ defined by (A.11)*.*

Theorem A.39
Definitions A.37 and A.38 are equivalent.

Proof. Cf. [16, p. 42]. □

Example A.40
Consider the problem (A.1)-(A.3) *with* $\Omega \subset \mathbb{R}$ *being a bounded, open interval. Then, for all* $v \in L^2(0, T_E, H_0^1(\Omega))$*, equation* (A.14) *reads:*

$$\int_0^{T_E} \langle \partial_t y(t), v(t) \rangle_{H_0^1(\Omega)^*, H_0^1(\Omega)} \, dt + \int_0^{T_E} \int_\Omega \frac{1}{2} \partial_x y(t) \; \partial_x v(t) + \partial_x y(t) \; v(t) \, dx \, dt = 0,$$

or, following the above argument,

$$\int_0^{T_E} \int_\Omega \partial_t y(t) \; v(t) \, dx \, dt + \int_0^{T_E} \int_\Omega \frac{1}{2} \partial_x y(t) \; \partial_x v(t) + \partial_x y(t) \; v(t) \, dx \, dt = 0,$$

cf. Example A.36. Again, the dependence on x has been omitted.

The notion of weak solutions has eventually been established. The reason why weak solutions were introduced in the first place was the non-existence of classical solutions in practical relevant situations. However, it is yet unclear whether the above definitions really help: Not a single result regarding existence and uniqueness of weak solutions has been quoted or derived yet. For the Fokker-Planck equation discussed in this work, this is covered in Section 2.2.1. For more general problems, [30], [16], [7], and the references therein are recommended.

Appendix B

Contents of the CD-ROM

The enclosed CD-ROM contains four main folders. Aside from the PDF file of the Master's thesis and the source code and the documentation of PDE-MPC, all numerical results brought up in this work are included, along with the Maple worksheets that were used to generate the respective plots. In addition, animated versions of these plots are stored on the CD-ROM. An overview of the contents is provided in Table B.1.

Folder	Description
`thesis`	It contains the PDF file of the Master's thesis.
`pde-mpc`	This folder contains two sub-folders, `src` and `docu`. The source code of PDE-MPC is found in the `src` folder. The corresponding documentation is provided in the `docu` folder. Furthermore, the CodeBlocks project file is included.
`results`	Here, the output files of every numerical simulation brought up in this work are included. A description of the individual files can be found in Table 5.1. The results are sub-divided into three sub-folders `ex1-ou`, `ex2-gb`, and `ex3-shiryaev`, each indicating a stochastic process considered in this thesis. The outcomes for the Ornstein-Uhlenbeck process are found in `ex1-ou`. Analogously, the results in the case of the Geometric-Brownian process and the Shiryaev process are found in `ex2-gb` and `ex3-shiryaev`, respectively. Each of the three sub-folders contains more sub-folders. The structure of these is described below. Furthermore, two files, `ut-all-images.html` and `uxt-all-images.html`, can be found in each of the folders `ex1-ou`, `ex2-gb`, and `ex3-shiryaev`. They show all animated plots for the corresponding stochastic process on one page, either for space-independent (`ut`) or space-dependent control (`uxt`).
`maple`	This folder contains the Maple worksheets that were employed to generate the plots used in this work and animated versions thereof from the output files of PDE-MPC.

Table B.1: Contents of the enclosed CD-ROM.

The structure of `ex1-ou`, `ex2-gb`, and `ex3-shiryaev` is as follows. Each of these folders consists of several sub-folders of the form *Ctrl*`_mpcSteps-`*Steps*`_N-`*Horizon*, where

- *Ctrl* $\in$ {`ut`, `uxt`} indicates whether a space-independent control (`ut`) or a space-dependent control (`uxt`) is used;
- *Steps* is the number of MPC steps performed, i.e. for $T_E = 5$ and $T_s = 0.5$ the number of MPC steps is 10;
- *Horizon* is the MPC horizon $N \in \mathbb{N}$, where the notation introduced in Remark 3.13 is used, i.e. $N = 1$ is the shortest possible horizon.

The suffix `_num` in the folder name indicates that the gradient of the reduced cost functional is obtained by means of Remark 4.6. Deviations of standard parameters are also signaled by suffixes. For example, if α is set to 10, it is denoted by `_alpha-10` at the end of the folder name.

List of Figures

List of Tables

Bibliography

[1] H. W. Alt. *Lineare Funktionalanalysis*. Springer, 6th edition, 2012.

[2] N. Altmüller and L. Grüne. Distributed and boundary model predictive control for the heat equation. *GAMM-Mitteilungen*, 35(2):131–145, 2012.

[3] M. Annunziato and A. Borzì. Optimal Control of Probability Density Functions of Stochastic Processes. *Mathematical Modelling and Analysis*, 15:393–407, 2010.

[4] M. Annunziato and A. Borzì. A Fokker-Planck control framefork for multidimensional stochastic processes. *Journal of Computational and Applied Mathematics*, 237:487–507, 2013.

[5] D. G. Aronson. Non-negative solutions of linear parabolic equations. *Annali della Scuola Normale Superiore di Pisa, Classe di Scienze* 3^e *série*, 22(4):607–694, 1968.

[6] A. Borzì and V. Schulz. *Computational Optimization of Systems Governed by Partial Differential Equations*. SIAM, Philadelphia, 2011.

[7] L. C. Evans. *Partial Differential Equations*, volume 19 of *Graduate Studies in Mathematics*. American Mathematical Society, 2nd edition, 2010.

[8] M. G. Forbes, M. Guay, and J. F. Forbes. Control design for first-order processes: shaping the probability density of the process state. *Journal of process control*, 14(4):399–410, 2004.

[9] I. I. Gihman and A. V. Skorohod. *Stochastic Differential Equations*. Springer-Verlag, 1972.

[10] L. Grüne. Numerische Methoden für gewöhnliche Differentialgleichungen (lecture notes). Retrieved June 8, 2013 from `http://num.math.uni-bayreuth.de/de/team/Gruene_Lars/lecture_notes/num2/num2_4.pdf`, 2010.

[11] L. Grüne and J. Pannek. *Nonlinear Model Predictive Control*. Springer, 2011.

[12] W. Hackbusch. Theorie und Numerik elliptischer Differentialgleichungen (lecture notes). Retrieved September 22, 2011 from `http://www.mis.mpg.de/publications/other-series/ln/lecturenote-2805.html`, 2005.

[13] R. Herzog and K. Kunisch. Algorithms for PDE-constrained optimization. *GAMM-Mitteilungen*, 33(2):163–176, 2010.

[14] M. R. Hestenes and E. Stiefel. Methods of Conjugate Gradients for Solving Linear Systems. *Journal of Research of the National Bureau of Standards*, 49(6):409–436, December 1952.

[15] M. Hintermüller. Numerische Verfahren in der Optimierung (lecture notes). Retrieved September 30, 2013 from `http://www.uni-graz.at/imawww/hintermueller/optimierung.pdf`.

[16] M. Hinze, R. Pinnau, M. Ulbrich, and S. Ulbrich. *Optimization with PDE Constraints*. Springer, 2009.

[17] K. Ito and K. Kunisch. Augmented Lagrangian Methods for Nonsmooth, Convex Optimization in Hilbert Spaces. *Nonlinear Analysis: Theory, Methods and Applications*, 41(5-6):591–616, August 2000.

[18] G. Jumarie. Tracking control of non-linear stochastic systems by using path cross-entropy and Fokker-Planck equation. *International Journal of Systems Science*, 23(7):1101–1114, 1992.

[19] I. Karatzas and S. E. Shreve. *Brownian Motion and Stochastic Calculus*. Springer-Verlag, 1991.

[20] O. A. Ladyzhenskaya, V.A. Solonnikov, and N. N. Ural'ceva. *Linear and Quasi-linear Equations of Parabolic Type*, volume 23 of *Translations of Mathematical Monographs*. American Mathematical Society, 1968.

[21] J. L. Lions. *Optimal Control of Systems Governed by Partial Differential Equations*. Springer-Verlag, Berlin, 1971.

[22] J. Nocedal and S. J. Wright. *Numerical Optimization*. Springer, New York, 2nd edition, 2006.

[23] H. J. Oberle. Variationsrechnung und optimale Steuerung (lecture notes). Retrieved October 21, 2013 from `http://www.math.uni-hamburg.de/home/oberle/skripte/varopt.html`.

[24] M. Plail. *Die Entwicklung der optimalen Steuerungen. Von den Anfängen bis zur eigenständigen Disziplin in der Mathematik*. Vandenhoeck & Ruprecht, 1998.

[25] S. J. Qin and T. A. Badgwell. A survey of industrial model predictive control technology. *Control Engineering Practice*, 11:733–764, 2003.

[26] J. B. Rawlings and D. Q. Mayne. *Model Predictive Control: Theory and Design*. Nob Hill Pub, 2009.

[27] H.-J. Reinhardt. Differenzenapproximationen partieller Differentialgleichungen (lecture notes). Retrieved September 22, 2011 from `http://www.uni-siegen.de/fb6/numerik/skripts/skripte/` (File: DPDscript.zip), 1997.

[28] R. Risken. *The Fokker-Planck Equation: Methods of Solution and Applications*. Springer, Berlin, 1996.

[29] C.-W. Shu. Essentially Non-Oscillatory and Weighted Essentially Non-Oscillatory Schemes for Hyperbolic Conservation Laws. In A. Quarteroni, editor, *Advanced Numerical Approximation of Nonlienar Hyperbolic Equations*, volume 1697 of *Lecture Notes in Mathematics*, pages 325–432. Springer, 1998.

[30] F. Tröltzsch. *Optimale Steuerung partieller Differentialgleichungen – Theorie, Verfahren und Anwendungen*. Vieweg+Teubner, 2nd edition, 2009.

[31] J. Wloka. *Partielle Differentialgleichungen*. Teubner, 1982.

Erklärung

Hiermit erkläre ich, dass ich die vorliegende Arbeit selbstständig und nur unter Verwendung der angegebenen Quellen und Hilfsmittel angefertigt habe.

Diese Arbeit hat in gleicher oder ähnlicher Form noch keiner anderen Prüfungsbehörde vorgelegen.

Bayreuth, den 31. Januar 2014

(Arthur Fleig)