Model Predictive Control for the Fokker–Planck Equation

Von der Universität Bayreuth
zur Erlangung des Grades eines
Doktors der Naturwissenschaften (Dr. rer. nat.)
genehmigte Abhandlung

von

Arthur Fleig

aus Semipalatinsk

1. Gutachter:	Prof. Dr. Lars Grüne
2. Gutachter:	Prof. Dr. Alfio Borzì

Tag der Einreichung:	27.11.2020
Tag des Kolloquiums:	23.03.2021

Acknowledgments

"Imaginary mountains build themselves from our efforts to climb them, and it is our repeated attempts to reach the summit that turns those mountains into something real."

— Bennett Foddy

Firstly, I wish to express my gratitude to my supervisor Prof. Dr. Lars Grüne not only for the continuous guidance, constructive discussions, and impressively quick responses throughout the research project, but also for the excellent working conditions and the many opportunities for international scientific exchange.

Let me thank Prof. Dr. Alfio Borzì for sparking this research topic and for agreeing to review this thesis. I also thank Prof. Dr. Jörg Rambau and Prof. Dr. Anton Schiela for being members of the examination committee.

Likewise, I thank the many (former and current) members and colleagues at the Chair of Applied Mathematics and the Chair of Serious Games for a pleasant and stimulating working environment: Dr. Miroslav Bachinski, Dr. Robert Baier, Dr. Michael Baumann, Dr. Philipp Braun, Florian Fischer, Matthias Höger, Dr. Thomas Jahn, Markus Klar, Dr. Huijuan Li, Dr. Georg Müller, Dr. Julián Ortiz Lopez, Dr. Vryan Gil S. Palma, Viktorija Paneva, Dr. Simon Pirkelmann, Manuel Schaller, Dr. Sofia Seinfeld, Dr. Manuela Sigurani, Tobias Sproll, Dr. Marleen Stieler, and Dr. Matthias Stöcklein. I am particularly grateful to Prof. Dr. Jörg Müller for giving me the opportunity to finish my research project while allowing me to work on exciting new challenges. Special thanks to Dr. Roberto Guglielmi for helping me jump-start my research project and to Prof. Dr. Karl Worthmann for his guidance, support, and invaluable advice from my time as an undergraduate student until today. My great respect and many thanks for Sigrid Kinder's and Nadine Rexfort's impeccable organization skills, keeping me and others as safe as possible from administrative issues.

Der größte Dank gilt meiner Frau und meinen Kindern (♡), die mich in all der Zeit ermutigt, er- und getragen haben, und meinen Eltern und Schwiegereltern für ihre geduldige Unterstützung und Kinderbetreuung. Ohne sie wäre dieses Unterfangen nicht möglich gewesen.

Abstract (english / german)

Abstract

This work contributes to a better understanding of Model Predictive Control (MPC) in the context of the Fokker–Planck equation.

The Fokker–Planck equation is a partial differential equation (PDE) that describes the evolution of a probability density function in time. One possible application is the (optimal) control of stochastic processes described by stochastic differential equations (SDEs). Here, a macroscopic perspective is taken and instead of, e.g., individual particles (described by the SDE), all particles are controlled in terms of their density (described by the Fokker–Planck equation). This results in a PDE-constrained optimal control problem.

Model Predictive Control is an established and widely used technique in industry and academia to (approximately) solve optimal control problems. The idea of the "receding horizon" is easy to understand, the implementation is simple, and above all: MPC works very often in practice. The challenge, however, is to specify conditions under which this can be guaranteed or to verify these conditions for concrete systems.

In this thesis it is analyzed in detail under which conditions the MPC closed loop is provably (practically) asymptotically stable, i.e., under which conditions it converges to the desired target or to a neighborhood thereof. For this purpose we first introduce the Fokker–Planck framework and show the existence of optimal space- and time-dependent controls under (weak) regularity assumptions. Subsequently, we consider both the case of stabilizing MPC and economic MPC and include both space-independent and space-dependent control functions in our analysis.

In the case of stabilizing MPC, we show asymptotic stability of the MPC closed loop for a class of linear stochastic processes if the prediction horizon N is long enough. Moreover, we specify the minimal stabilizing horizon for specific stochastic processes. In the course of the analysis difficulties of the used L^2 cost function come to light and the question arises whether other cost functions allow an easier analysis.

In the case of the economic MPC, we thus fix a specific stochastic process but consider different cost functions instead. Here, the crucial system property for the effective use of the MPC controller is strict dissipativity. This property is verified for different cost functions, where the main challenge is to find a suitable storage function. It turns out that for the commonly used L^2 cost it is much more difficult to find such a storage function than for another cost function we propose.

Details of the numerical implementation with additional simulations and further research questions conclude the work.

Kurzfassung

Diese Arbeit trägt dazu bei, Modellprädiktive Regelung (MPC) im Zusammenhang mit der Fokker–Planck Gleichung besser zu verstehen.

Die Fokker–Planck Gleichung ist eine partielle Differentialgleichung (PDE), die die zeitliche Entwicklung einer Wahrscheinlichkeitsdichtefunktion beschreibt. Eine mögliche Anwendung ist die (optimale) Steuerung stochastischer Prozesse, die durch stochastische Differentialgleichungen (SDEs) beschrieben werden. Hierbei wird eine makroskopische Perspektive eingenommen und anstelle von z.B. einzelnen Partikeln (beschrieben durch die SDE) die Gesamtheit aller Partikel in Form ihrer Dichte (beschrieben durch die Fokker–Planck Gleichung) gesteuert. Dadurch erhält man ein Optimalsteuerungsproblem mit einer PDE als Nebenbedingung.

Modellprädiktive Regelung ist eine etablierte und in Industrie und Wissenschaft weit verbreitete Technik, mit der Optimalsteuerungsprobleme (approximativ) gelöst werden. Die Idee des "receding horizon" ist einfach zu verstehen, die Implementierung ist simpel und vor allem: MPC funktioniert in der Praxis sehr oft. Die Herausforderung ist es hingegen, Bedingungen, unter denen man dies garantieren kann, anzugeben bzw. diese Bedingungen für konkrete Systeme zu verifizieren.

In dieser Arbeit wird genauer untersucht, unter welchen Bedingungen der geschlossene MPC-Regelkreis garantiert (praktisch) asymptotisch stabil ist, d.h. zum gewünschten Ziel bzw. in eine Umgebung des Ziels konvergiert. Hierzu stellen wir zunächst das Fokker–Planck Framework vor und zeigen die Existenz von optimalen orts- und zeitabhängigen Kontrollen unter (schwachen) Regularitätsannahmen. Anschließend betrachten wir sowohl den Fall des stabilisierenden MPC als auch den des ökonomischen MPC und berücksichtigen sowohl ortsunabhängige als auch ortsabhängige Kontrollfunktionen.

Im Falle des stabilisierenden MPC zeigen wir die asymptotische Stabilität des geschlossenen MPC-Regelkreises für eine Klasse von linearen stochastischen Prozessen, sofern der Prädiktionshorizont N lang genug ist und spezifizieren den minimal nötigen Horizont für spezifische stochastische Prozesse. Im Laufe der Analyse kristallisieren sich Schwierigkeiten der verwendeten L^2-Kostenfunktion heraus und es stellt sich die Frage, ob andere Kostenfunktionen eine einfachere Analyse ermöglichen.

Im Falle des ökonomischen MPC halten wir daher einen spezifischen stochastischen Prozess fest und betrachten dafür verschiedene Kostenfunktionen. Die zentrale Systemeigenschaft für die effektive Nutzung des MPC-Reglers hier ist strikte Dissipativität. Für verschiedene Kostenfunktionen wird diese Eigenschaft nachgewiesen, wobei hier die Herausforderung darin besteht, eine passende Speicherfunktion zu finden. Hierbei stellt sich heraus, dass es für die üblich verwendeten L^2-Kosten erheblich schwieriger ist, eine solche Speicherfunktion zu finden, als für eine andere Kostenfunktion, die wir vorstellen.

Details zur numerischen Implementierung mit zusätzlichen Simulationen und weiteren Forschungsfragen schließen die Arbeit ab.

Contents

Introduction

1

Initiated by Kolmogorov's work [63], the study of the Fokker–Planck (FP) equation, also known as Kolmogorov forward equation, has received great and increasing attention, since, for a large class of stochastic processes, it describes the evolution of the associated probability density function (PDF). The FP equation is a parabolic partial differential equation (PDE). Using the FP equation has proven to be a viable approach in several physical, chemical, and biological applications that involve noise. A large amount of literature has been developed on the FP equation in connection with transition PDFs that are associated to stochastic processes; see, for example, [41, 58]. In recent years, the well-posedness of the FP equation under low regularity assumptions on the coefficients has been studied in connection with existence, uniqueness and stability of martingale solutions to the related stochastic differential equation [65, 31]. Furthermore, control properties of the FP equation have become of major interest in mean-field game theory; see [77].

Our focus is on the optimal control of the FP equation. It stems from a statistical approach, which allows to recast an optimal control problem (OCP) subject to an Itô stochastic differential equation into a deterministic optimization problem subject to a FP equation. The idea behind this approach is that the state of a stochastic process can be characterized by the associated PDF. The approach has similarities to solving stochastic OCPs via the Hamilton-Jacobi-Bellman (HJB) approach, see [7], the difference being that the optimal control is derived by optimizing the solution of a PDE (the FP equation) rather than deriving the optimal control from the solution of a PDE (the HJB equation).

Controlling the PDF is an interesting alternative to classical approaches in stochastic optimization that optimize the mean or higher moments. It yields an accurate and flexible control strategy, which can accommodate a wide class of objectives; see also [17, Sect. 4]. In this direction, in [19, 40, 60, 61, 99], PDF-control schemes were proposed, where the cost functional depends on the PDF of the stochastic state variable. In this way, a deterministic objective results. In particular, no average over all possible states of the stochastic process appears in the objective functional, which is usually the case in stochastic OCPs; see, e.g., [39]. Still, in [40, 60, 61, 99], stochastic methods were adopted in order to approximate the state variable of the random process. In contrast to this, in [4, 5] the authors approach the problem of tracking the PDF associated with the stochastic process directly. Since then, this approach was used in different contexts, e.g., in [16, 85, 86]. In the numerical simulations in [4, 5, 36], Model Predictive Control (MPC) has proven to be an efficient method for the control of PDFs of controlled stochastic processes. In this approach, the distance of the actual PDF to the desired reference PDF, integrated or summed over several time steps into the future, is minimized using the FP equation as a model for

predicting the actual PDF. The first piece of the resulting optimal control function is then applied to the stochastic system and the whole process is repeated iteratively.

The optimal control problem to be solved in each step of the MPC scheme belongs to the class of tracking type OCPs governed by PDEs and the usual norm for measuring the distance to a reference in PDE-based optimal tracking control is the L^2-norm [95]. The L^2 norm is advantageous because L^2 is a Hilbert space, which significantly simplifies, e.g., the computation of gradients, which in turn is crucial for the implementation of numerical optimization algorithms. In a large part of this thesis, we thus follow the existing literature and use the L^2 norm as distance measure in our MPC optimal control problem.

So far, the efficiency of MPC for the Fokker–Planck equation was only verified by means of numerical simulations. Particularly, it is not clear whether the process controlled by MPC—the so-called *MPC closed loop*—will converge to the desired reference PDF. This is the question about the stability of the closed loop at the reference PDF. Moreover, it is not clear how large the time span into the future over which the distance is optimized—the so-called *optimization horizon*—must be in order to obtain stability. For smaller time spans the MPC closed loop might not converge to the desired reference PDF. On the other hand, the shorter the optimization horizon, the less computationally demanding the numerical solution of the OCP in each MPC step. Thus, one main goal of the thesis is to establish rigorous mathematical results that guarantee stability and in some cases also an upper bound on the necessary optimization horizon.

Unfortunately, although the Fokker–Planck MPC framework is in principle applicable to arbitrary nonlinear stochastic control systems and arbitrary initial and reference PDFs, a rigorous analysis of such a general setting appears out of reach to the moment. Therefore, the analysis of the MPC closed loop will be carried out in a more limited setting, e.g., for linear stochastic dynamics and Gaussian PDFs. This class of systems often appears in engineering problems and has the advantage that its controllability properties are well understood due to the recent paper [22]. Yet, even with the availability of the results from [22] the analysis of the MPC scheme is not straightforward, because the implications of these controllability properties for the PDFs on the controllability of the L^2 stage cost are indirect and difficult to analyze. This is the point where the use of the otherwise very convenient L^2 stage cost turns out to be disadvantageous and a substantial part of this thesis, particularly Chapter 5, is thus devoted to an in-depth analysis of this cost.

All the more justified is the question of alternative costs. Luckily, the Gaussian setting, although limited, allows us to use the Wasserstein distance W^2—a metric that is well-suited for measuring the distance between two PDFs [44]—much more comfortably due to its simplified structure in this particular case. Moreover, we are able to suggest a third stage cost that is suitable for the Gaussian setting and, although very similar to the W^2 cost, is much easier to analyze. Furthermore, we believe that the insights from this restricted setting are very valuable for the general nonlinear setting: Clearly, if certain approaches do (provably) not work in the linear Gaussian setting, they will inevitably also fail in more general settings.

In the remainder of this chapter we introduce the Fokker–Planck optimal control framework in Section 1.1 and present the outline of the thesis and list the contributions in Section 1.2.

1.1 The Fokker–Planck Optimal Control Framework

Given a final time $T > 0$, let us consider a controlled continuous-time stochastic process described by the (Itô) stochastic differential equation (SDE)

$$dX_t = b(X_t, t; u(X_t, t))dt + \tilde{a}(X_t, t)dW_t, \quad t \in]0, T[, \tag{1.1}$$

with an initial condition $\mathring{X} \in \mathbb{R}^d$, $d \geq 1$, where $\mathring{X}$ is a random variable that is distributed according to some probability density function $\mathring{\rho}$. Here, $W_t \in \mathbb{R}^m$ is an m-dimensional Wiener process, $b = (b_1, ..., b_d)$ is the vector valued drift function, and the diffusion matrix $\tilde{a}(X_t, t) \in \mathbb{R}^{d\times m}$ is assumed to have full rank. The control $u(X_t, t)$, acting on (1.1) through the drift term b, has to be chosen from a suitable space of control functions $\mathcal{U}$.

Under appropriate assumptions on the coefficients $\tilde{a}$ and b, cf. [79, p. 227] and [80, p. 297], and given the initial probability density function $\mathring{\rho}$, the evolution of probability density functions ρ associated with (1.1) is modeled by the Fokker–Planck equation, also called forward Kolmogorov equation:

$$\partial_t \rho(x,t) - \sum_{i,j=1}^{d} \partial_{ij}^2 \left(a_{ij}(x,t)\rho(x,t)\right) + \sum_{i=1}^{d} \partial_i \left(b_i(x,t;u(x,t))\rho(x,t)\right) = 0 \quad \text{in } Q, \tag{1.2a}$$

$$\rho(\cdot, 0) = \mathring{\rho}(\cdot) \text{ in } \Omega. \tag{1.2b}$$

In this parabolic PDE and throughout the work, we will denote by ∂_i and ∂_t the partial derivative with respect to space x_i and time t, respectively, where $i = 1, ..., d$. The domain of interest is given by $Q := \Omega \times]0, T[$, where, in this work, either $\Omega = \mathbb{R}^d$ or $\Omega \subset \mathbb{R}^d$ is a bounded domain with C^1 boundary. The diffusion coefficients $a_{ij} \colon Q \to \mathbb{R}$ are related to $\tilde{a}$ from (1.1) via $a_{ij} = \sum_k \tilde{a}_{ik}\tilde{a}_{jk}/2$ for $i, j = 1, ..., d$. The drift coefficients $b_i \colon Q \times U \to \mathbb{R}$ are the respective components of the vector valued drift function b from (1.1). The control u acting on the drift term may depend on time and/or space. The function $\mathring{\rho} \colon \Omega \to \mathbb{R}_{\geq 0}$ is a given initial PDF and $\rho \colon Q \to \mathbb{R}_{\geq 0}$ is the unknown PDF. For an exhaustive theory and more details on the connection between stochastic processes and the FP equation, including several applications regarding the description of transitions of a system from a macroscopic point of view, we refer to [84].

Since ρ is required to be a probability density function, it shall furthermore satisfy the standard properties of a PDF, i.e., non-negativity and conservation of mass:

$$\rho(x,t) \geq 0 \quad \forall (x,t) \in Q \qquad \text{and} \qquad \int_\Omega \rho(x,t)\, dx = 1 \quad \forall t \in]0, T[. \tag{1.3}$$

If the FP equation evolves on a bounded domain $\Omega \subset \mathbb{R}^d$, e.g., in case of localized SDEs [92], suitable boundary conditions on $\partial\Omega \times]0, T[$ have to be employed. For a complete characterization of possible boundary conditions in one space dimension, see the work of Feller [30]. In the multidimensional case, one possible choice is the zero-flux boundary condition $n(x) \cdot j(x,t) = 0$ on $\partial\Omega \times]0, T[$, where j denotes the probability flux[1] and $n(x)$ is the unit normal vector to the surface $\partial\Omega$, see [5, 16]. With this so-called reflecting boundary condition, the conservation of mass property in (1.3) holds. Another

[1] The probability flux describes the flow of probability in terms of probability per unit time per unit area.

possibility is to use absorbing boundary conditions [79, p. 231] as in [4, 5, 36], also known as homogeneous Dirichlet boundary conditions:

$$\rho(x,t) = 0 \quad \text{on} \quad \partial\Omega \times]0,T[. \tag{1.4}$$

Absorbing boundary conditions are appropriate in some scenarios. For instance, when considering the Shiryaev stochastic diffusion [74] on a bounded domain rather than on $[0,\infty[$, a particle hitting the boundary shall leave the domain (by being absorbed) instead of being reflected back. Thus, for absorbing boundary conditions, conservation of mass in space is not an appropriate requirement. Yet, if the objective is to keep the PDF within a given compact subset of Ω and the probability to find X_t outside of Ω is negligible, then this issue is mitigated for a large enough Ω, as numerical simulations show [4, 36]. See also [58, Ch. 5] for a comparison between the Gihman–Skorohod [43] and the Feller classification of boundary conditions.

In the case $\Omega = \mathbb{R}^d$ we want to focus on Gaussian distributions. As such, we consider natural boundary conditions, i.e.,

$$\rho(x,t) \to 0 \text{ as } \|x\| \to \infty \text{ for all } t \in]0,T[. \tag{1.5}$$

Since Gaussian PDFs can be fully characterized by their mean and their covariance matrix, we look at solutions of (1.2) of the form

$$\rho(x,t;u) := |2\pi\Sigma(t;u)|^{-1/2} \exp\left(-\frac{1}{2}(x-\mu(t;u))^\top \Sigma(t;u)^{-1}(x-\mu(t;u))\right), \tag{1.6}$$

where $\mu(t;u) \in \mathbb{R}^d$ is the (controlled) mean and $\Sigma(t;u) \in \mathbb{R}^{d\times d}$ is the (controlled) covariance matrix, which is symmetric and positive definite. For a matrix $A \in \mathbb{R}^{d\times d}$, throughout this work, we write $|A| := \det(A)$.

One specific process that will often appear in the analysis is the so-called Ornstein–Uhlenbeck process. Besides the geometric Brownian motion, it is one of the simplest and most widely used processes defined by a stochastic differential equation. It originally comes from physics and models the velocity of a massive Brownian particle under friction [96]. The multidimensional extension presented below is a special case of modeling dispersion of particles in shallow water [56]. Moreover, it can be used to obtain an n-factor Vasicek model [98, 69, 88] describing the evolution of interest rates.

We start with the one-dimensional case. For $d = 1$ and given parameters $\theta, \varsigma > 0$ and $\nu \in \mathbb{R}$, the uncontrolled Ornstein–Uhlenbeck process is defined by

$$dX_t = \theta\,(\nu - X_t)\,dt + \varsigma dW_t, \quad t \in]0,T[,$$

with an initial condition $\mathring{X} \in \mathbb{R}^d$. The parameter θ is called *mean reversion rate*; it models the "attraction level" to the so-called *mean reversion level* ν to which the process tends to drift. Lastly, ς represents the impact of randomness.

Next, we add a control u that, as in (1.1), acts on the drift term.

$$dX_t = [\theta\,(\nu - X_t) + u(X_t,t)]\,dt + \varsigma dW_t, \quad t \in]0,T[.$$

The control u will not always depend on X_t, but we can always translate the control by subtracting $\theta\nu$. Hence, we set $\nu = 0$ without loss of generality and arrive at

$$dX_t = [-\theta X_t + u(X_t,t)]\,dt + \varsigma dW_t, \quad t \in]0,T[. \tag{1.7}$$

An extension to the multi-dimensional setting is made by considering d equations of type (1.7). In this case the parameters become vectors, i.e., $\theta = (\theta_1, ..., \theta_d)$ with $\theta_i > 0$, $i = 1, ..., d$, and so on. In the Fokker–Planck equation (1.2) we thus have

$$\text{a drift term} \qquad b_i(x, t; u(x, t)) = -\theta_i x_i + u_i(x_i, t), \tag{1.8a}$$

$$\text{and a diffusion matrix} \qquad a(x, t) = \operatorname{diag}(\varsigma_1, ..., \varsigma_d). \tag{1.8b}$$

For given dynamics (1.2) and suitable boundary conditions, we then consider optimal control problems in which we want to minimize some state- and control-dependent cost functional $\tilde{J}$ over some set of admissible controls $\mathcal{U}_{ad} \subset \mathcal{U}$, i.e.,

$$\min_{u \in \mathcal{U}_{ad}, \rho} \tilde{J}(\rho, u) \text{ s.t. (1.2) and either (1.4) or (1.5).} \tag{1.9}$$

Note that, although the uncontrolled FP equation, i.e., (1.2) with $u \equiv 0$, is linear, due to the control u appearing in the drift term we have to deal with a bilinear OCP, considerably complicating the analysis, see, e.g., Chapter 2.

One particular objective is to steer to (and remain at) a given desired PDF $\bar{\rho}$. In continuous time, this can be formulated as an infinite-horizon OCP by setting

$$\tilde{J}(\rho, u) = \int_0^\infty \ell(\rho(x, t), u(t))\, dt,$$

where ℓ is the so-called *stage cost* or *running cost* yet to be defined. It typically penalizes the distance between the current and the desired PDF as well as the control. These optimization problems are addressed using Model Predictive Control (MPC), by now a standard method for controlling linear and nonlinear systems if constraints and/or optimal behavior of the closed loop are important. It is introduced in Chapter 3.

In the OCP (1.9) we do not demand non-negativity and conservation of mass (1.3) explicitly, for the following reasons. As will be shown in the subsequent chapters, the former holds automatically if the initial state is non-negative. Regarding the latter, on the one hand, in the above Shiryaev example, the loss of the conservation of mass property is pertinent to the model. On the other hand, numerical results in [4, 5, 36] indicate that requiring this property can, at least in practice, often be circumvented by choosing a large enough domain Ω. However, under these conditions, the state is not necessarily a PDF. For better differentiation, a solution to the FP equation that is not a PDF will be denoted by y instead of ρ throughout the work.

1.2 Outline and Contribution

Chapter 2 – Optimal Control of the FP Equation with Space-Dependent Controls This chapter is dedicated to the analysis of the bilinear optimal control problem introduced in Section 1.1 from the perspective of PDE-constrained optimization. We prove the well-posedness of the controlled Fokker–Planck equation and show that its unique solution is non-negative provided the initial state is non-negative. The existence of optimal controls is shown for a general class of objective functionals. Moreover, for common quadratic cost functionals of tracking and terminal type, first order necessary optimality conditions are derived using the adjoint state. Furthermore, we provide pointwise

conditions for the variational inequality occuring in the first order necessary optimality conditions.

The bilinear structure of the OCP and the fact that the control function depends on both time and space and moreover acts as a coefficient of the advection term greatly restricts the use of many classical results found in, e.g., [95]. Even so, we are able to avoid any differentiability requirements of the control function and only require suitable integrability properties instead.

The results of this chapter have been presented in [37, 38].

Chapter 3 – Model Predictive Control In a series of papers [4, 5, 36], Model Predictive Control of the Fokker–Planck equation has been established as a numerically feasible way for controlling stochastic processes via their probability density functions. To prove the effectiveness of MPC in this setting, we provide an introduction to MPC and list existing results regarding the stability and performance of the MPC closed loop in this chapter. These results are used subsequently.

Chapter 4 – Stabilizing MPC – Space-independent control This chapter marks the beginning of the analysis of the MPC closed loop. We start with the case of stabilizing MPC. Numerical simulations [4, 5] suggest that (in many cases) the MPC controller yields an asymptotically stable closed-loop system for optimization horizons looking only one time step into the future.

In this chapter a formal proof of this fact is provided for the Fokker–Planck equation corresponding to the controlled Ornstein–Uhlenbeck process using an L^2 stage cost and control functions that are constant in space. The key step of the proof consists in the verification of an exponential controllability property with respect to the stage cost. One difficulty to overcome in this context is the increasing optimal value function at time $t = 0$ for some parameters, which prohibits to conclude stability of the closed-loop system for the shortest possible horizon. An equivalent cost function that yields the same optimal control sequence provides a remedy.

The results of this chapter have been presented in [33]. However, compared to [33], a different and more general equivalent cost function is used in the case $\alpha > 1$. Moreover, the exponential controllability property in this case is verified more rigorously. Furthermore, more exact numerical simulations were performed, yielding new and updated plots.

Chapter 5 – Stabilizing MPC – Linear Control The setting of Chapter 4 is extended to encompass a large class of (controllable) linear processes. Moreover, the control is space-dependent (but limited to being linear in space). For this class of linear processes, we show that asymptotic stability of the MPC closed-loop system can be guaranteed for large enough horizon lengths N, proving rigorously that the MPC controller is a viable choice for steering PDFs. Moreover, in case of the Ornstein–Uhlenbeck process we prove asymptotic stability of the MPC closed-loop system for the shortest possible horizon, extending the results of Chapter 4 to linear control functions. As in the previous chapter, an L^2 stage cost is used.

The results of this chapter have been presented in [34]. Compared to [34], some proofs and statements were added and/or updated.

Chapter 6 – Economic MPC – Linear Control We extend our analysis of the MPC closed loop to the case of economic MPC, in which the stage cost does not have to be positive definite with respect to the desired equilibrium state. The pivotal property in order to conclude (practical) stability of the MPC closed-loop system and to make statements about its performance is *strict dissipativity* of the corresponding optimal control problems. This fact was revealed in a series of recent papers, see, e.g., [25, 3, 46] or the monographs and survey papers [81, 49, 29], and has triggered a renewed interest in this classical systems theoretic property that goes back to [101]. Thus, the focus is on verifying the strict dissipativity property. We focus on the Ornstein–Uhlenbeck process. In addition to the L^2 stage cost, we consider the quadratic Wasserstein cost, W^2, and another quadratic stage cost, called $2F$, which is specifically tailored to the linear Gaussian setting and resembles commonly used cost functions in optimal control.

The main difficulty in proving strict dissipativity is to find a suitable storage function, if it exists. Our results show that linear storage functions, which are easiest to find, can only be used reliably for the $2F$ stage cost. For the L^2 and the W^2 stage cost, we show that for many model parameters no suitable linear storage function exists. Exemplarily, we provide nonlinear storage functions that allow to conclude strict dissipativity in these cases. We observe that the OCPs have to be looked at individually, depending on the model parameters, in order to find a suitable storage function.

The results of this chapter have been presented in [32, 35].

Chapter 7 – Numerical Implementation and Simulations This chapter is dedicated to the numerical implementation and to numerical examples that might be of interest, but were not discussed in the previous chapters.

Here we explain the main program, `PDE-MPC`, which is used to numerically solve optimal control problems subject to the (d-dimensional) Fokker–Planck equation (1.2) using MPC. We provide details about the used algorithms and explain the structure of the program.

Moreover, we introduce `OU-MPC`, a program that is used to solve optimal control problems in the case of the Ornstein–Uhlenbeck process with Gaussian PDFs. Numerical errors in the discretization are eliminated by using the closed form solution that exists in this case, which also speeds up the computation considerably compared to using `PDE-MPC`.

Furthermore, to return from the macroscopic perspective to the underlying stochastic process at hand, we present `SDEControl`, a small program that numerically solves stochastic ODEs with a given control using the Euler–Maruyama method. We use it to verify the results obtained by the Fokker–Planck approach on the microscopic level.

We end this chapter with some numerical simulations that further demonstrate the power of the Fokker–Planck optimal control framework.

Chapter 8 – Future Research In this chapter we present open questions and topics that are particularly interesting for future research. This chapter concludes the thesis.

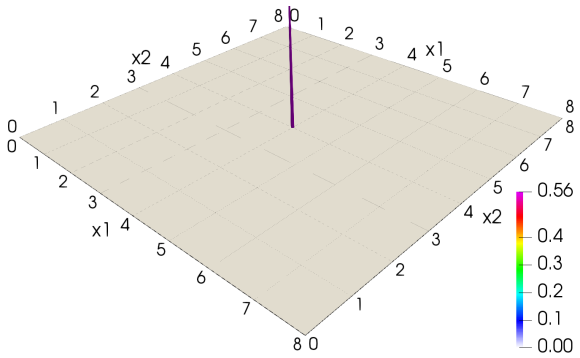
Optimal Control of the Fokker–Planck Equation with Space-Dependent Controls

2

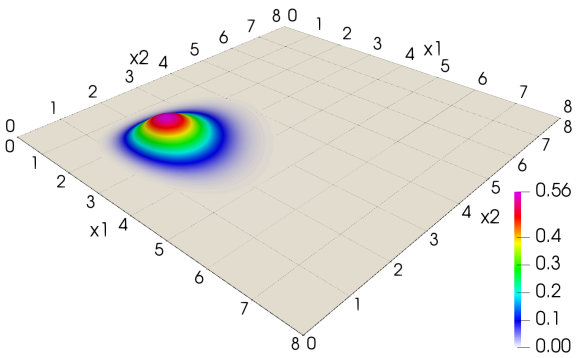
In the optimal control problems introduced in Section 1.1, the control acts through the drift term. Hence, the evolution of the PDF is controlled through the advection term of the FP equation. This is a rather weak action of the controller on the system, usually called of bilinear type, since the control appears as a coefficient in the state equation. Indeed, only few controllability results are known for such kind of control systems, for instance in connection with quantum control systems and stochastic control [13] or in relation to the planning problem for the mean-field game system [76]. Concerning the existence of bilinear optimal controls for a parabolic system of fourth order, a first result was given in [1], with a control function that only depends on time. This has been used in [5] in order to show existence of optimal controls for a FP equation with constant or time-dependent control functions. In this setting, however, due to the absence of space-dependent controls, there is no mechanism to cope with the diffusion term in the FP equation. Hence, unsurprisingly, acting on the space variable substantially improves tracking performance, as demonstrated in the numerical simulations in [36] and illustrated in Figure 2.1.

The aim of this chapter is to extend the theoretical study on the existence of bilinear optimal controls of the FP equation by [5] to the case of more general control functions, which depend on both time and space. We do not require any differentiability property of the control, which is in accordance with the simulations in [36]. For this reason, a careful analysis of the well-posedness of the FP equation is required. Indeed, suitable integrability assumptions are needed on the coefficient of the advection term in order to give meaning to the weak formulation of the equation. For this purpose, we use the functional framework proposed in the works of Aronson [8] and Aronson-Serrin [9]. In this setting, the advection coefficient belongs to a Bochner space that prevents us from choosing the set of square-integrable functions as the space of controls. As a result, the optimization problem is defined on a Banach space, a setting often considered whenever the state variable is subject to a nonlinear PDE; see, for example, [20, 83]. In recent works [65, 77], the well-posedness of the FP equation has been established even for drift coefficients that are square-integrable in time and space, in the context of renormalized solutions. These papers could describe the right framework for studying the optimal control problem of the FP equation in a Hilbert setting.

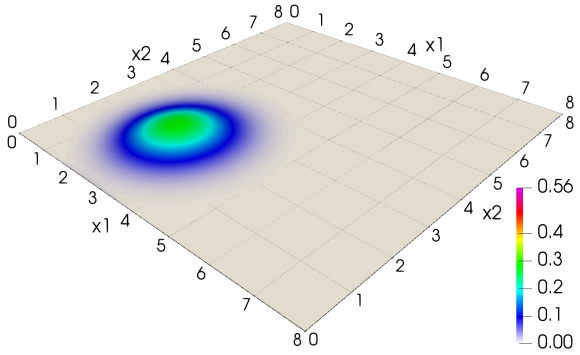
The remainder of this chapter is organized as follows. In Section 2.1, we formulate our optimal control problem and state general assumptions. In Section 2.2, we ensure the existence and uniqueness of (non-negative) solutions to the state equation. Section 2.3 is devoted to recast the FP equation in an abstract setting and to deduce a-priori estimates of its solution. These are used to prove our main result (Theorem 2.7 and Corollary 2.9)



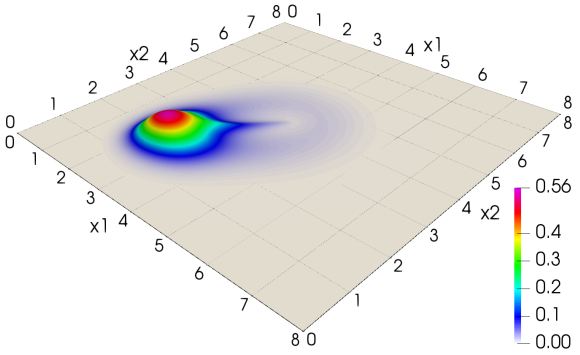
(a) Initial PDF: (Smoothed) Dirac delta.



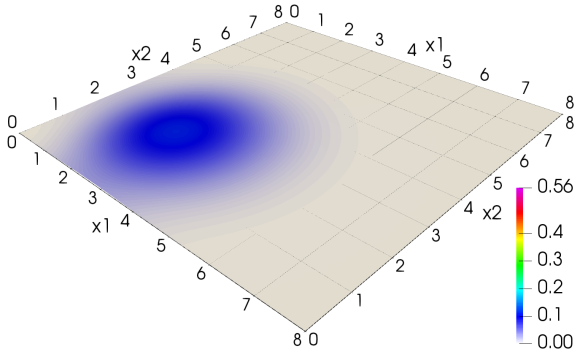
(b) Desired state.



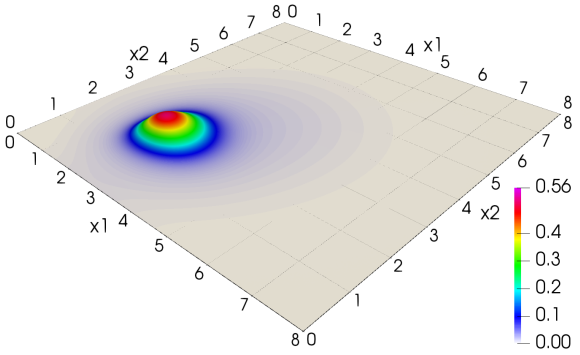
(c) State at $t = 0.5$ (space-independent control).



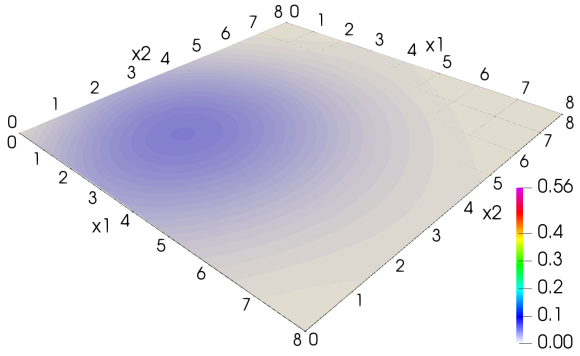
(d) State at $t = 0.5$ (space-dependent control).



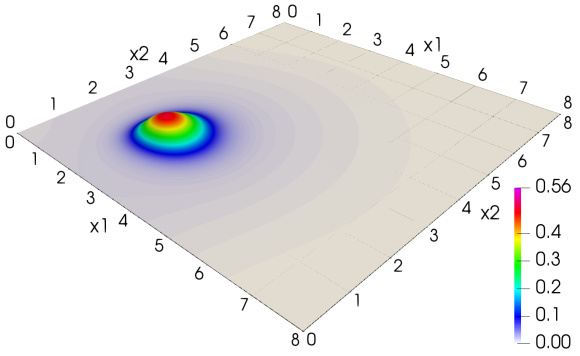
(e) State at $t = 1.5$ (space-independent control).



(f) State at $t = 1.5$ (space-dependent control).



(g) State at $t = 5$ (space-independent control).



(h) State at $t = 5$ (space-dependent control).

Figure 2.1: Comparison of space-independent ($u(t)$) and space-dependent ($u(x,t)$) control of a PDF associated to a stochastic process modeling the dispersion of substance in shallow water, cf. Example 7.1.

on existence of solutions to the considered optimal control problem for a general class of cost functionals. In Section 2.5, we deduce the system of first order necessary optimality conditions for common quadratic cost functionals. Section 2.6 concludes this chapter.

2.1 Problem Formulation and Assumptions

As outlined in Section 1.1 the aim is to control the stochastic process (1.1) via the FP equation (1.2) in an optimal way, i.e., by minimizing some state- and control-dependent cost functional $\tilde{J}$. More precisely, we consider the following optimal control problem:

$$\begin{aligned} \min_{u\in\mathcal{U}_{ad},y} \tilde{J}(y,u) \text{ s.t.: } \partial_t y - \sum_{i,j=1}^{d} \partial_{ij}^2 \left(a_{ij} y\right) + \sum_{i=1}^{d} \partial_i \left(b_i(u) y\right) &= 0 \quad \text{in } Q, \\ y(\cdot,0) &= \mathring{y}(\cdot) \text{ in } \Omega, \\ y &= 0 \quad \text{on } \partial\Omega \times]0,T[, \end{aligned} \tag{P}$$

where $\Omega \subset \mathbb{R}^d$ is a bounded domain with C^1 boundary and

$$\mathcal{U}_{ad} := \{u \in \mathcal{U} : u_a \leq u(x,t) \leq u_b \quad \text{for almost all } (x,t) \in Q\}, \tag{2.1}$$

with $u_a, u_b \in \mathbb{R}^d$ and $u_a \leq u_b$ component-wise. The space of controls

$$\mathcal{U} := L^q(0,T;L^\infty(\Omega;\mathbb{R}^d)) \subset L^2(0,T;L^\infty(\Omega;\mathbb{R}^d)) \tag{2.2}$$

with $2 < q \leq \infty$ is motivated by the integrability requirements in [8] to ensure well-posedness of the state equation; see Section 2.2.

Recall that we denote the state by y instead of ρ since, in general, we cannot guarantee the conservation of mass property in (1.3) due to the absorbing boundary conditions. Likewise, the initial state is denoted by $\mathring{y}$ instead of $\mathring{\rho}$. The arguments (x,t) are omitted here and throughout this chapter, whenever clear from the context. Similarly, we use the notation $b_i(u)$ and $b_i(t;u(t))$ in order to stress the action of the control u through the coefficient b_i and to underline the time dependence, respectively, omitting the other arguments.

Unless stated otherwise, we will use the above spaces $\mathcal{U}_{ad}$ and $\mathcal{U}$ throughout the chapter. Moreover, we impose the following requirements.

Assumption 2.1. *1.* $\forall i,j = 1,...,d : a_{ij} \in C^1(\overline{\Omega})$.

2. $\exists \theta > 0$ *such that* $\forall \xi \in \mathbb{R}^d$ *and for almost all* $x \in \Omega : \sum_{i,j=1}^{d} a_{ij}(x)\xi_i\xi_j \geq \theta|\xi|^2$.

3. The function $b\colon \mathbb{R}^{d+1} \times \mathcal{U} \to \mathbb{R}^d, (x,t;u) \mapsto b(x,t;u)$ *satisfies the growth condition*

$$\sum_{i=1}^{d} |b_i(x,t;u)|^2 \leq M(1+|u(x,t)|^2) \quad \forall x \in \mathbb{R}^d, \tag{2.3}$$

for every $i = 1,...,d$, $t \in [0,T]$, $u \in \mathcal{U}$, *and some constant* $M > 0$.

For simplicity, we assume the coefficients a_{ij} to be independent of time, which results in an autonomous operator. In Sections 2.4 and 2.5, Assumption 2.1(3) is replaced by the following, stronger requirement:

Assumption 2.2. $\exists r_i \in L^\infty(\Omega) : b_i(x,t;u) = r_i(x) + u_i(x,t),\ i = 1,...,d.$

The fact that b is affine in u is exploited in particular in the proofs of Theorem 2.7 and Lemma 2.11, in order to prove the existence of optimal solutions and the differentiability of the control-to-state operator, which will be introduced in Section 2.4.

2.2 Well-Posedness of the Fokker–Planck Equation

In this section, we establish the well-posedness of the FP equation in (P), where we add a source term $f\colon Q \to \mathbb{R}$ on the right-hand side, which will be of use for the well-posedness of the adjoint equation in Section 2.5.

Setting $\tilde{b}_j(u) := \sum_{i=1}^d \partial_i a_{ij} - b_j(u)$, we can recast the FP equation in flux formulation

$$\partial_t y - \sum_{j=1}^d \partial_j \Big(\sum_{i=1}^d a_{ij}\partial_i y + \tilde{b}_j(u) y \Big) = f \quad \text{in } Q\,.$$

Together with the initial and boundary conditions in (P), we have the associated weak formulation

$$\begin{aligned}\iint_Q fv \,\mathrm{d}x\mathrm{d}t &= \iint_Q \partial_t y v \,\mathrm{d}x\mathrm{d}t - \iint_Q \Big(\sum_{j=1}^d \partial_j \Big(\sum_{i=1}^d a_{ij}\partial_i y + \tilde{b}_j(u) y \Big)\Big) v \,\mathrm{d}x\mathrm{d}t \\ &= -\iint_Q y\partial_t v \,\mathrm{d}x\mathrm{d}t - \int_\Omega y(\cdot,0)v(\cdot,0)\,\mathrm{d}x + \iint_Q \sum_{j=1}^d \Big(\sum_{i=1}^d a_{ij}\partial_i y + \tilde{b}_j(u) y \Big)\partial_j v \,\mathrm{d}x\mathrm{d}t\end{aligned}$$

for test functions $v \in W_2^{1,1}(Q)$ with $v_{|\partial\Omega} = 0$ and $v(\cdot,T) = 0$.

We make use of this in the following theorem, which is a special case of [8, Thm. 1, p. 634] and guarantees the existence and uniqueness of (non-negative) solutions.

Theorem 2.3. *Let $\mathring{y} \in L^2(\Omega)$. Additionally, let $f \in L^q(0,T;L^\infty(\Omega))$ or $f = \operatorname{div}(\tilde{f})$ for some $\tilde{f}\colon Q \to \mathbb{R}^d$ with $\tilde{f}_j \in L^2(Q)$, $j = 1,...,d$. Then, there exists a unique $y \in L^2(0,T;H_0^1(\Omega)) \cap L^\infty(0,T;L^2(\Omega))$ satisfying*

$$\iint_Q -y\partial_t v + \sum_{j=1}^d \Big(\sum_{i=1}^d a_{ij}\partial_i y + \tilde{b}_j(u) y \Big)\partial_j v - fv \,\mathrm{d}x\mathrm{d}t = \int_\Omega \mathring{y} v(\cdot,0)\,\mathrm{d}x \tag{2.4}$$

for every $v \in W_2^{1,1}(Q)$ with $v_{|\partial\Omega} = 0$ and $v(\cdot,T) = 0$, i.e., y is the unique weak solution of the Fokker–Planck initial boundary value problem defined in (P)*, including a right-hand side f in the FP equation. Moreover, if $f \equiv 0$ and $0 \le \mathring{y} \le m$ almost everywhere in Ω for some $m > 0$, then y is bounded and $y(x,t) \ge 0$ almost everywhere in Q.*

Note that, due to the choice of $\mathcal{U}$ in (2.2) and Assumption 2.1(3), $\tilde{b}_j(u)$ belongs to $L^q(0,T;L^\infty(\Omega))$ for $j=1,...,d$, as required in [8].

The solution obtained by Theorem 2.3 is more regular; indeed, it belongs to the $W(0,T)$ space. We recall that

$$W(0,T) := \{y \in L^2(0,T;V) : \dot{y} \in L^2(0,T;V')\} \subset C([0,T];H)\,,$$

where $\dot{y}$ denotes the weak time derivative of y and $H := L^2(\Omega)$, $V := H_0^1(\Omega)$, and $V' := H^{-1}(\Omega)$, the dual space of V, endowed with norms

$$\|y\|_H^2 := \int_\Omega y^2 \,\mathrm{d}x\,, \quad \|y\|_V^2 := \int_\Omega |\nabla y|^2 \,\mathrm{d}x\,, \quad \|L\|_{V'} := \sup_{y\in V, \|y\|_V=1} |\langle L, y\rangle_{V',V}|\ ,$$

respectively, form the Gelfand triple $V \hookrightarrow H \hookrightarrow V'$. We denote by $|\cdot|$ the Euclidean norm and by $\langle\cdot,\cdot\rangle_{V',V}$ the duality map between V and V'. This notation and these spaces are used throughout the chapter.

Proposition 2.4. *Under the assumptions of Theorem* 2.3*, the solution y in Theorem* 2.3 *belongs to $W(0,T)$, possibly after a modification on a set of measure zero.*

Proof. The proof is analogous to the one of [95, Thm. 3.12], the only change being a different functional F. The idea is to show that F belongs to $L^2(0,T;V')$ and to rewrite the variational formulation of the PDE in terms of F to show that $\dot{y} = F$ in the sense of vector-valued distributions. In our case, for any fixed t, the linear functional is given by $F(t)\colon V \to \mathbb{R}$,

$$v \mapsto -\sum_{j=1}^d \Big(\sum_{i=1}^d a_{ij}\partial_i y(t) + \tilde{b}_j(t;u(t))y(t), \partial_j v\Big)_H + (f(t),v)_H\,.$$

We first assume $f \in L^q(0,T;L^\infty(\Omega))$ with $2 < q \le \infty$. $F(t)$ is bounded and thus continuous for all $t \in]0,T[$:

$$\begin{aligned}
|F(t)v| &= \left| -\sum_{j=1}^d \int_\Omega \Big(\sum_{i=1}^d a_{ij}\partial_i y(t) + \tilde{b}_j(t;u(t))y(t)\Big)\partial_j v \,\mathrm{d}x + \int_\Omega f(t)v\,\mathrm{d}x \right| \\
&\le \sum_{j=1}^d \int_\Omega \sum_{i=1}^d |a_{ij}|\,|\partial_i y(t)|\,|\partial_j v|\,\mathrm{d}x + \int_\Omega |f(t)||v|\,\mathrm{d}x \\
&\quad + \sum_{j=1}^d \int_\Omega |\tilde{b}_j(t;u(t))|\,|y(t)|\,|\partial_j v|\,\mathrm{d}x \\
&\le \underbrace{\sum_{i,j=1}^d \|a_{ij}\|_{L^\infty(\Omega)}}_{=:C} \|y(t)\|_V \|v\|_V + c_\Omega \|f(t)\|_H \|v\|_V \\
&\quad + \sum_{j=1}^d \|\tilde{b}_j(t;u(t))\|_{L^\infty(\Omega)} \|y(t)\|_H \|v\|_V\ ,
\end{aligned}$$

where c_Ω is such that $\|v\|_H \le c_\Omega \|v\|_V$ for any $v \in V = H_0^1(\Omega)$. Therefore,

$$\|F(t)\|_{V'} \le C\,\|y(t)\|_V + \sum_{j=1}^{d} \|\tilde{b}_j(t;u(t))\|_{L^\infty(\Omega)} \|y(t)\|_H + c_\Omega\,\|f(t)\|_H\,. \tag{2.5}$$

Since $\|y(t)\|_V \in L^2(0,T)$, $\|\tilde{b}_j(t;u(t))\|_{L^\infty(\Omega)} \in L^q(0,T)$, $\|y(t)\|_H \in L^\infty(0,T)$, and $\|f(t)\|_H \in L^q(0,T)$ with $q > 2$, the right-hand side of (2.5) belongs to $L^2(0,T)$, i.e., $F \in L^2(0,T;V')$. The remaining steps are the same as in the proof of [95, Thm. 3.12].

If $f = \operatorname{div}(\tilde{f})$, the spatial derivatives are transferred to v, which results in a very similar calculation and, in particular, also in $F \in L^2(0,T;V')$. □

Furthermore, note that we have $\int_\Omega \mathring{y} v \,\mathrm{d}x = \lim_{t\to 0} \int_\Omega y(t) v \,\mathrm{d}x = \int_\Omega y(0) v \,\mathrm{d}x$ for all $v \in V$, where the first equality follows from (2.4) and the second holds because $W(0,T) \subset C([0,T];H)$. Consequently, $y(0) = \mathring{y}$ in Ω.

2.3 A-priori Estimates

The purpose of this section is to deduce a-priori estimates of solutions to the Fokker–Planck initial boundary value problem given in (P), including a right-hand side $f \in L^2(0,T;V')$ in the FP equation. For the sake of clarity, we recast it in abstract form

$$\begin{cases} \dot{y}(t) + Ay(t) + B(u(t),y(t)) = f(t) \quad \text{in } V',\ t \in]0,T[, \\ y(0) = \mathring{y}\,, \end{cases} \tag{2.6}$$

where $\mathring{y} \in H$, $A\colon V \to V'$ is a linear and continuous operator such that

$$\langle Az, \varphi\rangle_{V',V} := \int_\Omega \sum_{i,j=1}^{d} \partial_i(a_{ij} z)\, \partial_j \varphi \,\mathrm{d}x \qquad \forall \varphi \in V\,,$$

and the operator $B\colon L^\infty(\Omega;\mathbb{R}^d) \times H \to V'$ is defined by

$$\langle B(u,y), \varphi\rangle_{V',V} := -\int_\Omega \sum_{i=1}^{d} b_i(u) y\, \partial_i \varphi \,\mathrm{d}x = -\int_\Omega y b(u).\nabla\varphi \,\mathrm{d}x \qquad \forall \varphi \in V\,.$$

In the following, $\mathcal{E}(\mathring{y},u,f)$ refers to (2.6) whenever we want to point out the data $(\mathring{y},u,f)$.

To ease the notation, we will still denote by A and B the two operators $A\colon L^2(0,T;V) \to L^2(0,T;V')$ and $B\colon \mathcal{U} \times L^\infty(0,T;H) \to L^q(0,T;V')$ such that for all $\varphi \in L^2(0,T;V)$, we have $Az = -\sum_{i,j=1}^d \partial_{ij}^2 (a_{ij} z)$ and

$$\int_0^T \langle Az(t), \varphi(t)\rangle_{V',V} \,\mathrm{d}t = \iint_Q \sum_{i,j=1}^{d} \partial_i(a_{ij} z)\, \partial_j \varphi \,\mathrm{d}x\mathrm{d}t\,, \tag{2.7}$$

and $B(u,y) = \sum_{i=1}^d \partial_i (b_i(u) y) = \operatorname{div}(b(u)y)$ such that

$$\int_0^T \langle B(u(t),y(t)), \varphi(t)\rangle_{V',V} \,\mathrm{d}t = -\iint_Q \sum_{i=1}^{d} b_i(u) y\, \partial_i \varphi \,\mathrm{d}x\mathrm{d}t \tag{2.8}$$

for all $\varphi \in L^{q'}(0,T;V)$ with $1/q + 1/q' = 1$. Indeed, thanks to Assumption 2.1(3), we have $\operatorname{div}(b(u)y) \in L^q(0,T;V')$ and

$$\|B(u,y)\|_{L^q(0,T;V')} = \|\operatorname{div}(b(u)y)\|_{L^q(0,T;V')} \le M(1+\|u\|_{\mathcal{U}})\,\|y\|_{L^\infty(0,T;H)}\,.$$

Note that the integral on the r.h.s. in (2.7) is not symmetric in z and φ, owing to the fact that A is not self-adjoint. The bilinear form $a\colon]0,T[\times V \times V \to \mathbb{R}$ associated with the FP equation is defined by

$$\begin{aligned} a(t,\psi,\varphi) &:= \int_\Omega \left(\sum_{i,j=1}^d \partial_i(a_{ij}\psi)\,\partial_j\varphi - \sum_{i=1}^d b_i(t;u(t))\psi\partial_i\varphi \right) \mathrm{d}x \\ &= \int_\Omega \left(\sum_{i,j=1}^d a_{ij}\partial_i\psi\,\partial_j\varphi + \sum_{j=1}^d \tilde{b}_j(t,u(t))\psi\partial_j\varphi \right) \mathrm{d}x. \end{aligned}$$

Thanks to the uniform ellipticity of A and Young's inequality, for every $\varepsilon > 0$, $t \in]0,T[$, and every $\varphi \in V$, we have that

$$\begin{aligned} \theta \int_\Omega |\nabla\varphi|^2\,\mathrm{d}x \le \int_\Omega \sum_{i,j=1}^d a_{ij}\partial_i\varphi\,\partial_j\varphi\,\mathrm{d}x &= a(t,\varphi,\varphi) - \int_\Omega \sum_{j=1}^d \tilde{b}_j(t;u(t))\varphi\partial_j\varphi\,\mathrm{d}x \\ &\le a(t,\varphi,\varphi) + \|\tilde{b}(t;u(t))\|_{L^\infty(\Omega;\mathbb{R}^d)} \left(\varepsilon \int_\Omega |\nabla\varphi|^2\,\mathrm{d}x + \frac{1}{4\varepsilon}\int_\Omega |\varphi|^2\,\mathrm{d}x \right). \end{aligned}$$

Thus, with $\varepsilon = 3\theta/(4\|\tilde{b}(t;u(t))\|_{L^\infty(\Omega;\mathbb{R}^d)})$, we conclude

$$\frac{\theta}{4}\,\|\varphi\|_V^2 \le a(t,\varphi,\varphi) + C_1(t)\,\|\varphi\|_H^2\,, \tag{2.9}$$

where

$$C_1(t) := \|\tilde{b}(t;u(t))\|^2_{L^\infty(\Omega;\mathbb{R}^d)}/(3\theta). \tag{2.10}$$

We now derive some a-priori estimates on the solution of (2.6). We will need them in the following sections. In this chapter, from this section on, we denote by M and C generic, positive constants that might change from line to line.

Lemma 2.5. *Let $\mathring{y} \in H$, $f \in L^2(0,T;V')$ and $u \in \mathcal{U}$. Then a solution y of (2.6) satisfies the estimates*

$$\|y\|^2_{L^\infty(0,T;H)} \le M(u)\left(\|y(0)\|_H^2 + \|f\|^2_{L^2(0,T;V')} \right), \tag{2.11}$$

$$\|y\|^2_{L^2(0,T;V)} \le (1+\|u\|^2_{\mathcal{U}})M(u)\left(\|y(0)\|_H^2 + \|f\|^2_{L^2(0,T;V')} \right), \tag{2.12}$$

$$\|\dot{y}\|^2_{L^2(0,T;V')} \le (1+\|u\|^2_{\mathcal{U}})M(u)\left(\|y(0)\|_H^2 + \|f\|^2_{L^2(0,T;V')} \right) + 2\,\|f\|^2_{L^2(0,T;V')}\,, \tag{2.13}$$

where $M(u) := Ce^{c(1+\|u\|^2_{\mathcal{U}})}$ for some positive constants c, C.

Proof. Let y be a solution of (2.6) and $t \in]0,T[$. Multiplying (2.6) by $y(t)$, we get

$$\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\left(\|y(t)\|_H^2\right) + a(t,y(t),y(t)) = \langle f(t), y(t)\rangle_{V',V}\,, \quad t \in]0,T[,$$

and thus

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\big(\,\|y(t)\|_H^2\,\big)+\frac{\theta}{2}\,\|y(t)\|_V^2 &\le \frac{\mathrm{d}}{\mathrm{d}t}\big(\,\|y(t)\|_H^2\,\big)+2a(t,y(t),y(t))+2C_1(t)\,\|y(t)\|_H^2\\ &=2\langle f(t),y(t)\rangle_{V',V}+2C_1(t)\,\|y(t)\|_H^2\\ &\le 2\varepsilon\,\|y(t)\|_V^2+\frac{1}{2\varepsilon}\,\|f(t)\|_{V'}^2+2C_1(t)\,\|y(t)\|_H^2\,.\end{aligned}$$

Fixing $\varepsilon=\theta/8$, we deduce the relation

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\|y(t)\|_H^2\right)+\frac{\theta}{4}\,\|y(t)\|_V^2\le\frac{4}{\theta}\,\|f(t)\|_{V'}^2+2C_1(t)\,\|y(t)\|_H^2\,. \tag{2.14}$$

Applying Gronwall's inequality, we have that

$$\|y(t)\|_H^2\le e^{\int_0^t 2C_1(\tau)\mathrm{d}\tau}\left(\|y(0)\|_H^2+\frac{4}{\theta}\int_0^t\|f(\tau)\|_{V'}^2\,\mathrm{d}\tau\right).$$

For $u\in\mathcal{U}$, the inequality

$$\|u\|_{L^2(0,T;L^\infty(\Omega;\mathbb{R}^d))}\le T^{\frac{q-2}{2q}}\,\|u\|_{\mathcal{U}} \tag{2.15}$$

holds. With $C_1(t)$ from (2.10) and due to Assumption 2.1(3) and (2.15), we deduce that $\int_0^T 2C_1(t)\mathrm{d}t\le M(1+\|u\|_{\mathcal{U}}^2)$, and thus

$$\|y\|_{L^\infty(0,T;H)}^2\le Ce^{c(1+\|u\|_{\mathcal{U}}^2)}\left(\|y(0)\|_H^2+\|f\|_{L^2(0,T;V')}^2\right).$$

Moreover, integrating (2.14) over $]0,T[$, we conclude that

$$\begin{aligned}\|y\|_{L^2(0,T;V)}^2&\le C\left(\|y(0)\|_H^2+\|f\|_{L^2(0,T;V')}^2\right)+C(1+\|u\|_{\mathcal{U}}^2)\,\|y\|_{L^\infty(0,T;H)}^2\\ &\le C(1+\|u\|_{\mathcal{U}}^2)e^{c(1+\|u\|_{\mathcal{U}}^2)}\left(\|y(0)\|_H^2+\|f\|_{L^2(0,T;V')}^2\right).\end{aligned}$$

We recall that C might change from line to line. Finally, multiplying (2.6) by $\varphi\in L^2(0,T;V)$ and integrating over $]0,T[$ yields

$$\begin{aligned}\left|\int_0^T\langle\dot y(t),\varphi(t)\rangle_{V',V}\,\mathrm{d}t\right|&\le\|y\|_{L^\infty(0,T;H)}\,\|u\|_{L^2(0,T;L^\infty(\Omega;\mathbb{R}^d))}\,\|\varphi\|_{L^2(0,T;V)}\\ &\quad+C_\alpha\,\|y\|_{L^2(0,T;V)}\,\|\varphi\|_{L^2(0,T;V)}+\|f\|_{L^2(0,T;V')}\,\|\varphi\|_{L^2(0,T;V)}\,,\end{aligned}$$

where $C_\alpha>0$ is such that $\|A\xi\|_{V'}\le C_\alpha\,\|\xi\|_V$ for all $\xi\in V$. Thanks to (2.15),

$$\|\dot y\|_{L^2(0,T;V')}\le C_\alpha\,\|y\|_{L^2(0,T;V)}+C\,\|y\|_{L^\infty(0,T;H)}\,\|u\|_{\mathcal{U}}+\|f\|_{L^2(0,T;V')}\,.$$

Using twice the estimate $(a+b)^2\le 2a^2+2b^2$, we derive (2.13) by the estimates on $\|y\|_{L^\infty(0,T;H)}$ and $\|y\|_{L^2(0,T;V)}$. □

2.4 Existence of Optimal Controls

This section contains our main result of this chapter: the existence of optimal controls for (P), with $\mathcal{U}_{ad}$ and $\mathcal{U}$ as in (2.1) and (2.2). Fixing $\mathring{y} \in H$, we introduce the control-to-state operator $\Theta\colon \mathcal{U} \to C([0,T];H)$ such that $u \mapsto y \in C([0,T];H)$ is a solution of $\mathcal{E}(\mathring{y}, u, 0)$. Thus, the optimization problem turns into minimizing the so-called reduced cost functional $J\colon \mathcal{U} \to \mathbb{R}$ such that $J(u) := \tilde{J}(\Theta(u), u)$, which we assume to be bounded from below, over the non-empty subset of admissible controls $\mathcal{U}_{ad} \subset \mathcal{U}$. We recall that Assumption 2.2 is used in this section.

In order to prove the main theorem, we will need the following compactness result (see [10], [67, Thm. 5.1, p. 58] or [89]).

Theorem 2.6. *Let I be an open and bounded interval of $\mathbb{R}$, and let X, Y, Z be three Banach spaces, with dense and continuous embeddings $Y \hookrightarrow X \hookrightarrow Z$, the first one being compact. Then, for every $p \in [0, \infty[$ and $r > 1$, we have the compact embeddings*

$$L^p(I;Y) \cap W^{1,1}(I;Z) \hookrightarrow L^p(I;X)$$

and

$$L^\infty(I;Y) \cap W^{1,r}(I;Z) \hookrightarrow C(\overline{I};X).$$

Theorem 2.7. *Let $\mathring{y} \in H$. Consider the minimization of the reduced cost functional $J(u) = \tilde{J}(\Theta(u), u)$ over $\mathcal{U}_{ad}$. Assume that J is bounded from below and (sequentially) weakly-star lower semicontinuous. Then there exists a pair $(\bar{y}, \bar{u}) \in C([0,T];H) \times \mathcal{U}_{ad}$ such that $\bar{y}$ solves $\mathcal{E}(\mathring{y}, \bar{u}, 0)$ and $\bar{u}$ minimizes J in $\mathcal{U}_{ad}$.*

Proof. Let $(u_n)_{n\geq 1}$ be a minimizing sequence, i.e., $J(u_n) \to I := \inf_{u\in\mathcal{U}_{ad}} J(u)$ as $n \to \infty$. Since $(u_n)_{n\geq 1} \subset \mathcal{U}_{ad}$, we have $\|u_n\|_{\mathcal{U}} \leq c\|u_n\|_{L^\infty(Q)} \leq C$ for some constants $c, C > 0$ and any $n \geq 1$. Moreover, the pair (u_n, y_n) satisfies the state equation

$$\dot{y}_n(t) + Ay_n(t) + B(u_n(t), y_n(t)) = 0\,, \quad y_n(0) = \mathring{y}\,. \tag{2.16}$$

The a-priori estimates of Lemma 2.5 ensure that there exists a positive constant, still denoted by C, such that, for all $n \in \mathbb{N}$,

$$\|y_n\|_{L^\infty(0,T;H)}\,, \quad \|y_n\|_{L^2(0,T;V)}\,, \quad \|\dot{y}_n\|_{L^2(0,T;V')} \leq C\,,$$

and so we deduce that

$$\begin{aligned} \|Ay_n\|_{L^2(0,T;V')} &\leq C_\alpha \|y_n\|_{L^2(0,T;V)} \leq C\,, \\ \|B(u_n, y_n)\|_{L^2(0,T;V')} &\leq c\,\|B(u_n, y_n)\|_{L^q(0,T;V')} \\ &\leq M(1 + \|u_n\|_{\mathcal{U}})\,\|y_n\|_{L^\infty(0,T;H)} \leq C\,, \end{aligned}$$

where we recall that the constant $C_\alpha > 0$, which appears in the proof of Lemma 2.5, is such that $\|A\xi\|_{V'} \leq C_\alpha \|\xi\|_V$ for all $\xi \in V$. Thus, there exist subsequences (still indexed with the subscript n) such that

$$\begin{aligned} &u_n \overset{*}{\rightharpoonup} \bar{u} && \text{weakly-star in } \mathcal{U}\,, \\ &y_n \overset{*}{\rightharpoonup} \bar{y} && \text{weakly-star in } L^\infty(0,T;H)\,, \\ &y_n \rightharpoonup \bar{y} && \text{weakly in } L^2(0,T;V)\,, \\ &\dot{y}_n \rightharpoonup \psi && \text{weakly in } L^2(0,T;V')\,, \\ &Ay_n \rightharpoonup \chi && \text{weakly in } L^2(0,T;V')\,, \\ &B(u_n, y_n) \rightharpoonup \Lambda && \text{weakly in } L^2(0,T;V')\,. \end{aligned}$$

Since the Banach-Alaoglu theorem ensures that $\mathcal{U}_{ad}$ is weakly-star closed [23], we deduce that $\bar{u} \in \mathcal{U}_{ad}$. We now want to pass to the limit in the state equation (2.16). First of all, we observe that $\psi = \dot{\bar{y}}$, thanks to the convergence in the $\sigma(\mathcal{D}(0,T;V), \mathcal{D}'(0,T;V'))$ topology. Thus, $\bar{y} \in W(0,T) \subset C([0,T];H)$. Moreover, since the operator $A\colon L^2(0,T;V) \to L^2(0,T;V')$ is strongly continuous, and therefore weakly continuous, too, we deduce that $A\bar{y} = \chi$. Finally, we claim that $B(\bar{u},\bar{y}) = \Lambda$, which, because of the bilinear action of the control, is the most difficult part of the proof. Note that, thanks to the first relation in Theorem 2.6 with $Y := V$, $X := H$, and $Z := V'$, the embedding $W(0,T) \subset L^2(0,T;H)$ is compact. Thus, $(y_n)_n$ admits a subsequence strongly convergent to $\bar{y}$ in $L^2(0,T;H)$. Therefore, for every $\varphi \in L^2(0,T;V)$,

$$
\begin{aligned}
&\int_0^T \langle B(\bar{u}(t),\bar{y}(t)) - \Lambda(t), \varphi(t)\rangle_{V',V}\,\mathrm{d}t \\
&= -\iint_Q \bar{y} b(\bar{u}).\nabla\varphi\,\mathrm{d}x\mathrm{d}t - \lim_{n\to\infty}\int_0^T \langle B(u_n(t),y_n(t)),\varphi(t)\rangle_{V',V}\,\mathrm{d}t \\
&= -\iint_Q \bar{y} b(\bar{u}).\nabla\varphi\,\mathrm{d}x\mathrm{d}t + \lim_{n\to\infty}\iint_Q y_n b(u_n).\nabla\varphi\,\mathrm{d}x\mathrm{d}t \\
&= -\lim_{n\to\infty}\iint_Q (\bar{y} b(\bar{u}) - y_n b(u_n)).\nabla\varphi\,\mathrm{d}x\mathrm{d}t \\
&= -\lim_{n\to\infty}\iint_Q \bar{y}(b(\bar{u}) - b(u_n)).\nabla\varphi\,\mathrm{d}x\mathrm{d}t - \lim_{n\to\infty}\iint_Q (\bar{y} - y_n)\, b(u_n).\nabla\varphi\,\mathrm{d}x\mathrm{d}t\,,
\end{aligned}
$$

where ∇ denotes the gradient with respect to the space variable $x \in \mathbb{R}^d$. We observe that $\bar{y} \in L^\infty(0,T;H)$ and $\partial_i\varphi \in L^2(0,T;H)$ for all $i = 1,...,d$, thus $\bar{y}\partial_i\varphi \in L^2(0,T;L^1(\Omega)) \subset L^{q'}(0,T;L^1(\Omega))$ with q' such that $1/q + 1/q' = 1$ and $L^q(0,T;L^\infty(\Omega)) = [L^{q'}(0,T;L^1(\Omega))]^*$, since the Lebesgue measure is $\sigma-$finite. We recall that b is affine on u; see Assumption 2.2. Therefore, $b(\bar{u}) - b(u_n) = (\bar{u}_i - u_{n,i})_{i=1,\dots,d}$. Now $u_n \overset{*}{\rightharpoonup} \bar{u}$ weakly-star in $\mathcal{U}$ ensures that the first integral goes to 0 as $n \to +\infty$. Furthermore, since the sequence $(b(u_n))_n$ is uniformly bounded and $y_n \to \bar{y}$ strongly in $L^2(0,T;H)$,

$$
\Big|\iint_Q (\bar{y} - y_n)\, b(u_n).\nabla\varphi\,\mathrm{d}x\mathrm{d}t\Big| \le C\,\|\bar{y} - y_n\|_{L^2(0,T;H)}\,\|\varphi\|_{L^2(0,T;V)} \to 0
$$

as $n \to +\infty$. Additionally, we observe that $\bar{y}(0) = \mathring{y}$, hence

$$
\dot{\bar{y}}(t) + A\bar{y}(t) + B(\bar{u}(t),\bar{y}(t)) = 0\,, \quad \bar{y}(0) = \mathring{y}\,.
$$

Finally, owing to the weakly-star lower semicontinuity of J, we conclude that

$$
J(\bar{u}) \le \liminf_{n\to\infty} J(u_n) = I.
$$

Thus, $(\bar{y},\bar{u})$ is an optimal pair for the considered optimal control problem. □

Theorem 2.7 clearly also holds for any $\mathcal{U}_{ad} \subset \mathcal{U}$ bounded and weakly-star closed. However, observe that in the unconstrained case $\mathcal{U}_{ad} \equiv \mathcal{U}$, asking only $J(u) \ge \gamma\|u\|_{\mathcal{U}}$

for some $\gamma > 0$ is not enough. Instead, one can modify the proof straightforwardly by requiring $J(u) \geq \gamma \|u\|_{L^\infty(Q;\mathbb{R}^d)}$, which is not very practical. An alternative might be to require more regularity on the state y and on the control u, in order to gain the same level of compactness required to deduce that $B(\bar{u}, \bar{y}) = \Lambda$. Indeed, further regularity of y can be ensured by standard improved regularity results; see, for example, [102, Thms. 27.2 and 27.5] and [64, Thm. 6.1 and Rem. 6.3]. However, these results require more regularity of the coefficients in the PDE and hence, on the control. In particular, one would need differentiability of u both in time and space. In comparison, requiring box constraints as in (2.1) seems to be a less restrictive choice. Note that, in case of bilinear action of the control into the system, even box constraints might not suffice to ensure the existence of optimal controls in general; see, for example, [68, Sect. 15.3, p. 237].

Remark 2.8. *We have shown in the previous proof that the control-to-state map $\Theta \colon \mathcal{U}_{ad} \subset \mathcal{U} \to C([0,T];H) \subset L^2(0,T;H), u \mapsto \Theta(u) = y \in L^2(0,T;H)$, where y solves $\mathcal{E}(\mathring{y}, u, 0)$, is sequentially continuous from $\mathcal{U}_{ad}$ (with the weak-star topology induced by $\mathcal{U}$) to $L^2(0,T;H)$ (with the strong topology).*

Corollary 2.9. *Let $y_d \in L^2(0,T;H)$, $y_\Omega \in H$, and $\alpha, \beta, \gamma \geq 0$. Consider the final time observation operator $S_T \colon W(0,T) \to H$ such that $y \mapsto y(T)$. Then an optimal pair $(\bar{y}, \bar{u}) \in C([0,T];H) \times \mathcal{U}_{ad}$ exists for the reduced cost functional*

$$J(u) := \frac{\alpha}{2} \|\Theta(u) - y_d\|^2_{L^2(Q)} + \frac{\beta}{2} \|S_T\Theta(u) - y_\Omega\|^2_{L^2(\Omega)} + \frac{\gamma}{2} \|u\|^2_{L^2(Q;\mathbb{R}^d)} . \tag{2.17}$$

Proof. The cost functional (2.17) is bounded from below by zero. Moreover, it is weakly lower semicontinuous in $L^2(Q;\mathbb{R}^d)$. This is due to Remark 2.8 and the fact that the embedding $W(0,T) \subset C([0,T];H)$ is continuous, that the operator S_T is linear and continuous, and that the norm functionals $\|\cdot\|^2_H$ and $\|\cdot\|^2_{L^2(Q;\mathbb{R}^d)}$ are weakly lower semicontinuous on H and $L^2(Q;\mathbb{R}^d)$, respectively. Moreover, a minimizing sequence $(u_n)_{n\geq 1}$ in $\mathcal{U}_{ad}$ converging to I is uniformly bounded both in $\mathcal{U}$ and in $L^2(Q;\mathbb{R}^d)$. Since the weak-star convergence in $\mathcal{U}$ implies the weak convergence in $L^2(Q;\mathbb{R}^d)$, we do not need to require weakly-star lower semicontinuity of J. Therefore, we can conclude the existence of an optimal pair $(\bar{y}, \bar{u}) \in C([0,T];H) \times \mathcal{U}_{ad}$. □

Remark 2.10. *Corollary* 2.9 *applies analogously to the case of time-independent controls u in the admissible space*

$$\tilde{\mathcal{U}}_{ad} := \{u \in L^\infty(\Omega;\mathbb{R}^d) : u_a \leq u(x) \leq u_b \quad \textit{for almost every } x \in \Omega\} \tag{2.18}$$

for the reduced cost functional

$$J_2(u) := \frac{\alpha}{2} \|\Theta(u) - y_d\|^2_{L^2(Q)} + \frac{\beta}{2} \|\Theta(u)(T) - y_\Omega\|^2_{L^2(\Omega)} + \frac{\gamma}{2} \|u\|^2_{L^2(\Omega;\mathbb{R}^d)} .$$

2.5 Adjoint State and Optimality Conditions

For the optimal control problem (P), modified by using the reduced cost functionals considered in Corollary 2.9 and Remark 2.10, we derive the first order necessary optimality conditions in this section. Incidentally, let us point out that these quadratic objective

functionals are commonly used in theory and practice; see, for example, [5, 36, 95]. As in the previous section, we require Assumption 2.2. We start by denoting the operator

$$D(u,y) := B(u-r,y) = \operatorname{div}(uy) \quad \forall u \in \mathcal{U}\,,\ y \in L^\infty(0,T;H)\,,$$

and proving the differentiability of the control-to-state operator.

Lemma 2.11. *Let $\mathring{y} \in H$. The control-to-state operator Θ from Section 2.4 is Fréchet-differentiable and, for every $\bar{u}$, $h \in \mathcal{U}$, the function $\Theta'(\bar{u})h$ satisfies*

$$\begin{cases} \dot{z}(t) + Az(t) + B(\bar{u}(t), z(t)) = -D(h(t), \bar{y}(t)) \quad \text{in } V'\,,\ t \in]0,T[, \\ z(0) = 0\,, \end{cases} \tag{2.19}$$

where $\bar{y} = \Theta(\bar{u})$.

Note that Theorem 2.3 ensures the existence of a unique weak solution of (2.19).

Proof. Thanks to Assumption 2.2, the map $L\colon \mathcal{U} \to C([0,T];H)$, such that $h \mapsto z \in C([0,T];H)$ is a solution of (2.19), is linear. Moreover, L is continuous; indeed, the estimate (2.11) yields

$$\begin{aligned} \|z\|^2_{L^\infty(0,T;H)} &\le Ce^{c(1+\|\bar{u}\|^2_{\mathcal{U}})} \|D(h,\bar{y})\|^2_{L^2(0,T;V')} \\ &\le Ce^{c(1+\|\bar{u}\|^2_{\mathcal{U}})} \|\bar{y}\|^2_{L^\infty(0,T;H)} \|h\|^2_{\mathcal{U}} \le C\|h\|^2_{\mathcal{U}}\,. \end{aligned}$$

Let us now introduce $y_h := \Theta(\bar{u}+h)$, the solution of $\mathcal{E}(\mathring{y}, \bar{u}+h, 0)$, and set $y := y_h - \bar{y}$. Thus, y satisfies

$$\begin{cases} \dot{y}(t) + Ay(t) + B(\bar{u}(t), y(t)) = -D(h(t), y_h(t)) \quad \text{in } V'\,,\ t \in]0,T[, \\ y(0) = 0\,. \end{cases}$$

Moreover, $D(h, y_h) \in L^q(0,T;V') \subset L^2(0,T;V')$, and (2.11) ensures

$$\begin{aligned} \|y\|^2_{L^\infty(0,T;H)} &\le Ce^{c(1+\|\bar{u}\|^2_{\mathcal{U}})} \|D(h,y_h)\|^2_{L^2(0,T;V')} \\ &\le Ce^{c(1+\|\bar{u}\|^2_{\mathcal{U}})} \|y_h\|^2_{L^\infty(0,T;H)} \|h\|^2_{\mathcal{U}}\,, \end{aligned}$$

with $\|y_h\|^2_{L^\infty(0,T;H)} \le Ce^{c(1+\|\bar{u}+h\|^2_{\mathcal{U}})} \|\mathring{y}\|^2_H$, which is locally bounded in h. Finally, $w := y - z$ is a solution of $\mathcal{E}(0, \bar{u}, -D(h(t), y(t)))$ and satisfies

$$\begin{aligned} \|w\|^2_{L^\infty(0,T;H)} &\le Ce^{c(1+\|\bar{u}\|^2_{\mathcal{U}})} \|D(h,y)\|^2_{L^2(0,T;V')} \\ &\le Ce^{c(1+\|\bar{u}\|^2_{\mathcal{U}})} \|y\|^2_{L^\infty(0,T;H)} \|h\|^2_{\mathcal{U}}\,, \end{aligned}$$

that is,

$$\|\Theta(\bar{u}+h) - \Theta(\bar{u}) - z\|^2_{L^\infty(0,T;H)} \le C\|\mathring{y}\|^2_H\, e^{c(1+\|\bar{u}+h\|^2_{\mathcal{U}})} \|h\|^4_{\mathcal{U}}\,.$$

Therefore, Θ is Fréchet differentiable and, for all $\bar{u}, h \in \mathcal{U}$, the operator $\Theta'(\bar{u})\colon \mathcal{U} \to C([0,T];H)$ is defined by $\Theta'(\bar{u})h := z$, where z solves (2.19). □

Next, we introduce the two operators

$$A^*\colon L^2(0,T;V) \to L^2(0,T;V') \quad \text{such that} \quad A^*z := -\sum_{i,j=1}^{d} a_{ij}\partial_{ij}^2 z,$$
$$\tilde{B}\colon L^2(0,T;V) \to L^2(0,T;L^2(\Omega;\mathbb{R}^d)) \quad \text{such that} \quad \tilde{B}(z) := \nabla z,$$

where ∇ denotes the gradient with respect to $x \in \mathbb{R}^d$. Observe that, for every $v, \varphi \in L^2(0,T;V)$,

$$\int_0^T \langle A^*v(t), \varphi(t)\rangle_{V',V}\,\mathrm{d}t = \int_0^T \langle A\varphi(t), v(t)\rangle_{V',V}\,\mathrm{d}t$$

and for every $u \in \mathcal{U}, v \in L^2(0,T;V)$ and $w \in L^\infty(0,T;H)$,

$$\int_0^T (b(u).\,\tilde{B}(v), w)_H\,\mathrm{d}t = \iint_Q \sum_{i=1}^{d} b_i(u) w\,\partial_i v\,\mathrm{d}x\mathrm{d}t$$
$$= -\int_0^T \langle B(u(t), w(t)), v(t)\rangle_{V',V}\,\mathrm{d}t, \quad (2.20)$$

and the above integrals are well-defined.

With this in mind, we can provide an explicit representation formula for the derivative of J as in Corollary 2.9.

Proposition 2.12. *Consider J of the form (2.17) with $y_d \in L^q(0,T;L^\infty(\Omega))$ and $y_\Omega \in H$. Let $\mathring{y} \in L^\infty(\Omega)$. Then, J is differentiable in $\mathcal{U}$ and, for all $u, h \in \mathcal{U}$,*

$$\mathrm{d}J(u)h = \sum_{i=1}^{d} \iint_Q h_i\,[y\partial_i p + \gamma u_i]\,\mathrm{d}x\mathrm{d}t \quad (2.21)$$

holds, where $y \in W(0,T) \cap L^\infty(Q)$ is the solution of $\mathcal{E}(\mathring{y}, u, 0)$ and $p \in W(0,T)$ is the solution of the adjoint equation

$$\begin{cases} -\dot{p}(t) + A^*p(t) - b(u(t)).\,\tilde{B}p(t) = \alpha\,[y(t) - y_d(t)] \quad \text{in } V',\ t \in]0,T[, \\ p(T) = \beta\,[y(T) - y_\Omega]\ . \end{cases} \quad (2.22)$$

Observe that, for all $i = 1, ..., d$, the function $h_i\langle \partial_i p, y\rangle_{V',V}\colon]0,T[\to \mathbb{R}$ belongs to $L^1(0,T)$, owing to $h_i \in L^q(0,T;L^\infty(\Omega))$ with $q > 2$, $y \in L^2(0,T;V)$ and $\partial_i p \in L^\infty(0,T;V')$. Existence and uniqueness of solutions for (2.22) are ensured by Aronson [8]. Indeed, $\mathring{y} \in L^\infty(\Omega)$ implies $y \in L^\infty(Q)$, and therefore $y - y_d \in L^q(0,T;L^\infty(\Omega))$, as required by Theorem 2.3. Moreover, we have $y(T) - y_\Omega \in L^2(\Omega)$. By the change of variable $q(t) = p(T-t)$, $v(t) = u(T-t)$ and $f(t) = \alpha[y(T-t) - y_d(T-t)]$, (2.22) is recast in a form such that the same results as in Theorem 2.3 and Proposition 2.4 can be applied, following the remark in [8, p. 621] concerning the adjoint operator.

Proof. Thanks to Lemma 2.11, the functional J is differentiable in $\mathcal{U}$. Let z solve (2.19). We set $z = \Theta'(u)h \in C([0,T];H)$ and derive

$$\mathrm{d}J(u)h = \langle z, \alpha[y - y_d]\rangle_{L^2(0,T;H)} + \langle z(T), \beta[y(T) - y_\Omega]\rangle_H + \gamma\langle h, u\rangle_{L^2(0,T;L^2(\Omega;\mathbb{R}^d))}$$

for all $u, h \in \mathcal{U}$, where y is the solution of the state equation $\mathcal{E}(\mathring{y}, u, 0)$. We now exploit the adjoint state p in order to figure out the dependence of $\mathrm{d}J(u)h$ on h. Indeed, owing to relations (2.20) and (2.22), we have that

$$\begin{aligned}
&\int_0^T \langle z(t), \alpha[y(t) - y_d(t)]\rangle_H \,\mathrm{d}t \\
&= \int_0^T \langle -\dot{p}(t) + A^*p(t) - b(u(t)).\,\tilde{B}p(t), z(t)\rangle_{V',V} \,\mathrm{d}t \\
&= \int_0^T \langle \dot{z}(t) + Az(t) + B(u(t), z(t)), p(t)\rangle_{V',V} \,\mathrm{d}t - \langle z(T), p(T)\rangle_H + \langle z(0), p(0)\rangle_H \\
&= -\langle z(T), p(T)\rangle_H - \int_0^T \langle D(h(t), y(t)), p(t)\rangle_{V',V} \,\mathrm{d}t \\
&= -\langle z(T), p(T)\rangle_H + \iint_Q y(t)h(t).\,\nabla p(t) \,\mathrm{d}x\mathrm{d}t\,.
\end{aligned}$$

Since $\langle z(T), \beta[y(T) - y_\Omega]\rangle_H = \langle z(T), p(T)\rangle_H$, we conclude

$$\mathrm{d}J(u)h = \iint_Q yh.\,\nabla p \,\mathrm{d}x\mathrm{d}t + \gamma \sum_{i=1}^d \langle h_i, u_i\rangle_{L^2(Q)} = \sum_{i=1}^d \iint_Q h_i \left[y\partial_i p + \gamma u_i\right] \mathrm{d}x\mathrm{d}t,$$

which is the assertion. □

A priori, $\mathrm{d}J(u)$ is defined only in $\mathcal{U}$, for every $u \in \mathcal{U}$. However, thanks to the representation formula (2.21), it admits an extension operator that is well-defined on $L^2(0,T;L^2(\Omega;\mathbb{R}^d))$.

With this, Proposition 2.12, and the variational inequality $dJ(\bar{u})(u - \bar{u}) \geq 0$, which holds for any $u \in \mathcal{U}_{ad}$ and locally optimal solution $\bar{u}$, we deduce the system of first order necessary optimality conditions. Note that, since the control-to-state operator is nonlinear, the reduced cost functional is non-convex even for standard quadratic costs like (2.17). In particular, there may be controls that are not optimal, not even locally, and nevertheless satisfy the necessary optimality conditions. Yet, this system plays an important role in the development of efficient numerical methods for the FP optimal control; see [4, 5]. Moreover, the simulations in [4, 5, 36] suggest that these conditions are viable to be used in practice.

Corollary 2.13. *Let $\mathring{y} \in L^\infty(\Omega)$, $y_d \in L^q(0,T;L^\infty(\Omega))$, and $y_\Omega \in H$. Consider the cost functional J defined by (2.17) with $\alpha, \beta, \gamma \geq 0$. An optimal pair $(\bar{y}, \bar{u}) \in C([0,T];H) \times \mathcal{U}_{ad}$ for J with corresponding adjoint state $\bar{p}$ satisfies the following necessary conditions:*

$$\begin{aligned}
&\partial_t \bar{y} - \sum_{i,j=1}^d \partial_{ij}^2(a_{ij}\bar{y}) + \sum_{i=1}^d \partial_i\big((r_i + \bar{u}_i)\bar{y}\big) = 0 && \text{in } Q\,,\\
&-\partial_t \bar{p} - \sum_{i,j=1}^d a_{ij}\partial_{ij}^2\bar{p} - \sum_{i=1}^d (r_i + \bar{u}_i)\partial_i\bar{p} = \alpha[\bar{y} - y_d] && \text{in } Q\,,\\
&\bar{y} = \bar{p} = 0 && \text{on } \partial\Omega \times]0,T[,\\
&\bar{y}(\cdot,0) = \mathring{y}(\cdot)\,, \quad \bar{p}(\cdot,T) = \beta[\bar{y}(\cdot,T) - y_\Omega(\cdot)] && \text{in } \Omega\,,\\
&\iint_Q \left[\bar{y}\partial_i\bar{p} + \gamma\bar{u}_i\right](u_i - \bar{u}_i)\,\mathrm{d}x\mathrm{d}t \geq 0 && \forall u \in \mathcal{U}_{ad}, i = 1, ..., d.
\end{aligned} \tag{2.23}$$

Proof. The necessary optimality conditions (2.23) are derived by combining the state equation $\mathcal{E}(\mathring{y}, \bar{u}, 0)$ for $\bar{y}$, (2.22) for the adjoint $\bar{p}$, and the variational inequality $\mathrm{d}J(\bar{u})(u - \bar{u}) \geq 0$ for all $u \in \mathcal{U}_{ad}$ and locally optimal $\bar{u}$. Thanks to (2.7), (2.8), and (2.20), which define the operators A, B, and $\tilde{B}$, respectively, we deduce the desired system. □

Following [95, Sect. 2.8], we can derive pointwise conditions for the variational inequality in (2.23). Indeed, if $\gamma = 0$, it follows for all $i = 1, ..., d$ and almost all $(x, t) \in Q$ that,

$$\bar{u}_i(x,t) = \begin{cases} u_{a_i}\,, & \text{if} \quad \bar{y}(x,t)\partial_i\bar{p}(x,t) > 0\,, \\ u_{b_i}\,, & \text{if} \quad \bar{y}(x,t)\partial_i\bar{p}(x,t) < 0\,, \end{cases}$$

and no value can be assigned if $\bar{y}(x,t)\partial_i\bar{p}(x,t) = 0$. If $\gamma > 0$, then we get the standard projection formula for almost all $(x, t) \in Q$:

$$\bar{u}_i(x,t) = \mathbb{P}_{[u_{a_i}, u_{b_i}]}\left\{ -\frac{1}{\gamma}\bar{y}(x,t)\partial_i\bar{p}(x,t) \right\}.$$

In case of time-independent controls considered in Remark 2.10, the only modification needed in the optimality system (2.23) is the variational inequality, which, for $\tilde{\mathcal{U}}_{ad}$ given by (2.18), changes to

$$\int_\Omega \left[\int_0^T \bar{y}\partial_i\bar{p}\,\mathrm{d}t + \gamma\bar{u}_i \right](u_i - \bar{u}_i)\,\mathrm{d}x \geq 0 \quad \forall u \in \tilde{\mathcal{U}}_{ad}, i = 1, ..., d.$$

2.6 Conclusion

In this chapter, we have considered a bilinear optimal control problem subject to the Fokker–Planck equation with homogeneous Dirichlet boundary conditions and a time- and space-dependent control. Without any differentiability requirements on the control we have proved the existence of optimal controls associated with a non-negative state solution and have derived the first order necessary optimality conditions rigorously, thereby extending the results of [5]. Very recently, similar results have been established for zero-flux boundary conditions in conjunction with a space-dependent control of specific structure in [16]. Thus, although finding sufficient conditions and proving uniqueness of the optimal control are still open questions—the main difficulty being the non-convexity of the problem due to the nonlinear control-to-state operator—the basis for solving the OCPs introduced in Section 1.1 has been established. As such, we switch to solving these OCPs. For this we use Model Predictive Control, which is introduced next.

Model Predictive Control

3

Model predictive control has developed into a standard method for controlling linear and nonlinear systems if constraints and/or optimal behavior of the closed loop are important. In this chapter we briefly present the concept of (nonlinear) MPC, a technique to solve optimal control problems of the type introduced in Section 1.1. A more detailed introduction can be found in the monographs [49] and [81].

In this approach, the so-called *running cost*—usually the distance of the actual state to the desired reference state—is integrated or summed over several time steps into the future. The resulting objective function is then minimized using a given model for predicting the actual state. In our case, the states are PDFs and the model for predicting the actual PDF is the Fokker–Planck equation. The first piece of the resulting optimal control function is then applied to the stochastic system and the whole process is repeated iteratively. This results in a closed-loop system—the so-called *MPC closed loop*.

To prove that MPC is an effective control method in our setting, we need to analyze the qualitative (and quantitative) behavior of the MPC closed loop. Depending on the structure of the running cost, the considered optimal control problem falls either into the category of so-called *stabilizing MPC* or *economic MPC*. The tools to analyze the behavior of the MPC closed loop are presented for both these frameworks in their respective sections.

3.1 Preliminaries

As we will describe below, in MPC the control input is synthesized by iteratively solving optimal control problems at discrete points in time. It is therefore convenient to consider the dynamics in discrete time. Hence, suppose we have a process whose state $z(k)$ is measured at discrete times t_k, $k \in \mathbb{N}_0$. Furthermore, suppose we can control it on the time interval $[t_k, t_{k+1}[$ via a control signal $u(k)$. Then we can consider nonlinear discrete-time control systems

$$z(k+1) = f(z(k), u(k)), \quad z(0) = \mathring{z}, \tag{3.1}$$

with state $z(k) \in \mathbb{X} \subset Z$ and control $u(k) \in \mathbb{U} \subset U$, where Z and U are metric spaces. State and control constraints are incorporated in $\mathbb{X}$ and $\mathbb{U}$, respectively. Whenever clear from the context, we might abbreviate the definition of the control system in (3.1) by

$$z^+ = f(z, u).$$

Continuous-time models such as the one presented in Section 1.1 can be considered in the discrete-time setting by sampling with a (constant) sampling time $T_s > 0$, i.e.,

$t_k = t_0 + kT_s$, or by replacing it with a numerical discretization. Given an initial state $\mathring{z}$ and an admissible *control sequence* $\mathbf{u}$, either finite, i.e., $\mathbf{u} = (u(k))_{k=0,\dots,N-1} \in \mathbb{U}^N$, or infinite, i.e., $\mathbf{u} = (u(k))_{k\in\mathbb{N}_0} \in \mathbb{U}^\infty$, the solution trajectory is denoted by $z_{\mathbf{u}}(\cdot;\mathring{z})$. Note that we do not require the control $u(k)$ to be constant on $[t_k, t_{k+1}[$—in general, $u(k)$ can be a time-dependent function on $[t_k, t_{k+1}[$.

As mentioned in Section 1.1, stabilization and tracking problems such as steering to a desired state and remaining there can be recast as infinite-horizon OCPs. However, solving OCPs governed by PDEs on large or even infinite horizons is, in general, computationally hard. The idea behind MPC is to circumvent this issue by iteratively solving optimal control problems on a shorter, finite time horizon and use the resulting (open-loop) optimal control values to construct a feedback law $\mathcal{F}\colon \mathbb{X} \to \mathbb{U}$ for the *MPC closed-loop system*

$$z_{\mathcal{F}}(k+1) = f(z_{\mathcal{F}}(k), \mathcal{F}(z_{\mathcal{F}}(k))). \tag{3.2}$$

Given a *stage cost* $\ell\colon Z \times U \to \mathbb{R}$, instead of solving the infinite-horizon OCP

$$\begin{aligned} J_\infty(\mathring{z}, \mathbf{u}) &:= \sum_{k=0}^{\infty} \ell(z_{\mathbf{u}}(k;\mathring{z}), u(k)) \to \min_{\mathbf{u}\in\mathbb{U}^\infty}! \\ \text{s.t. } & z_{\mathbf{u}}(k+1;\mathring{z}) = f(z_{\mathbf{u}}(k;\mathring{z}), u(k)),\ z_{\mathbf{u}}(0;\mathring{z}) = \mathring{z}, \\ & z_{\mathbf{u}}(k;\mathring{z}) \in \mathbb{X} \text{ for all } k \in \mathbb{N}_0, \end{aligned} \tag{OCP_∞}$$

the feedback law $\mathcal{F}$ is constructed through the following MPC scheme:

Algorithm 3.1 (MPC scheme)**.** *0. Given an initial value $z_{\mathcal{F}}(0) \in \mathbb{X}$, fix the length of the receding horizon $N \geq 2$ and set $n = 0$.*

1. Initialize the state $\mathring{z} = z_{\mathcal{F}}(n)$ and solve the following finite-horizon OCP:

$$\begin{aligned} J_N(\mathring{z}, \mathbf{u}) &:= \sum_{k=0}^{N-1} \ell(z_{\mathbf{u}}(k;\mathring{z}), u(k)) \to \min_{\mathbf{u}\in\mathbb{U}^N}! \\ \text{s.t. } & z_{\mathbf{u}}(k+1;\mathring{z}) = f(z_{\mathbf{u}}(k;\mathring{z}), u(k)),\ z_{\mathbf{u}}(0;\mathring{z}) = \mathring{z}, \\ & z_{\mathbf{u}}(k;\mathring{z}) \in \mathbb{X} \text{ for all } k \in \{0,\dots,N\}. \end{aligned} \tag{OCP_N}$$

Apply the first value of the resulting optimal control sequence denoted by $\mathbf{u}^* \in \mathbb{U}^N$, i.e., set $\mathcal{F}(z_{\mathcal{F}}(n)) := u^*(0)$.

2. Evaluate $z_{\mathcal{F}}(n+1)$ according to relation (3.2), set $n := n+1$ and go to step 1.

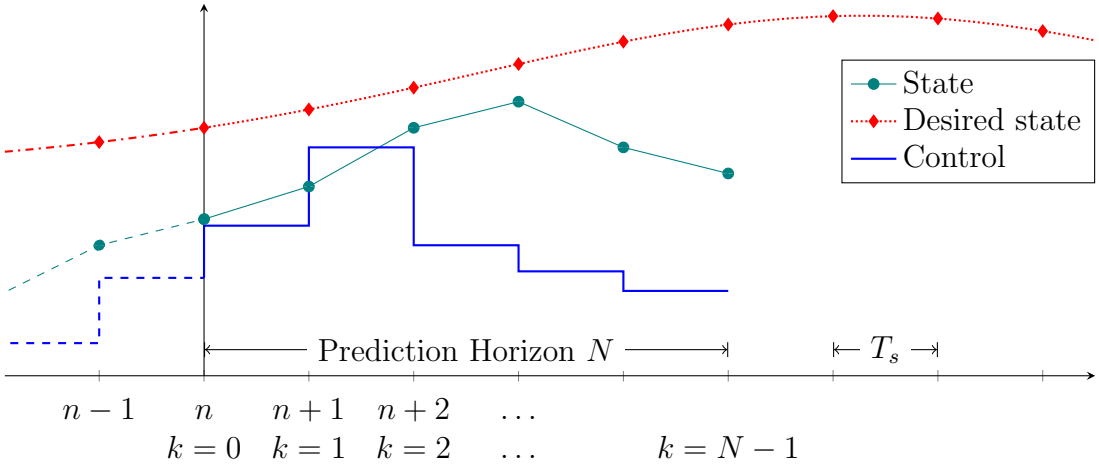
This scheme is illustrated in Figure 3.1. In connection with the above scheme, the index n denotes the "global" time index, while k denotes the index in the open-loop optimal control problem (OCP_N), as illustrated in the figure. Whenever we want to point out the importance of the *horizon length* N, we will denote the feedback by $\mathcal{F}_N$ instead of $\mathcal{F}$.

For both the infinite and the finite-horizon OCP we introduce the optimal value function.

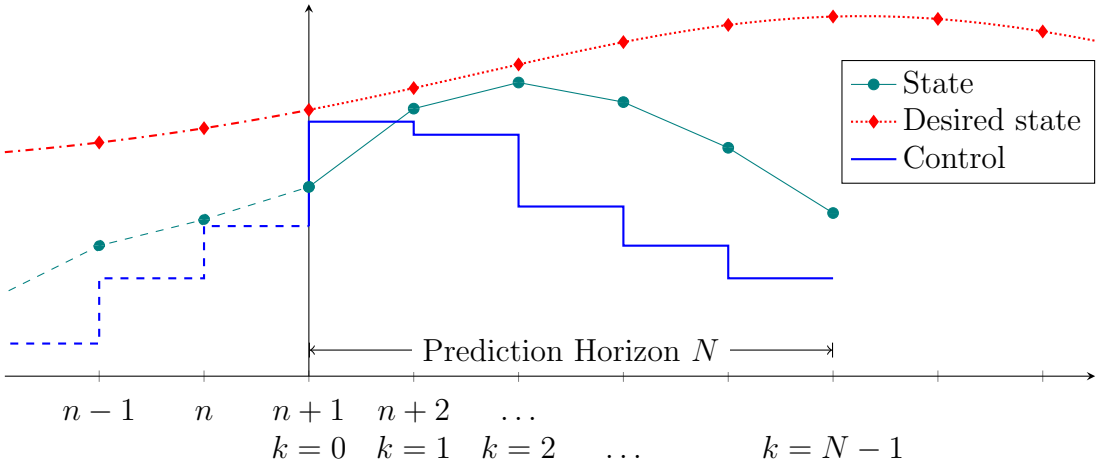
Definition 3.2 (Optimal value function)**.** *The functions*

$$V_\infty(\mathring{z}) := \inf_{\mathbf{u}} J_\infty(\mathring{z}, \mathbf{u}) \quad \textit{and} \quad V_N(\mathring{z}) := \inf_{\mathbf{u}} J_N(\mathring{z}, \mathbf{u}) \tag{3.3}$$

are referred to as optimal value functions.



(a) Discrete time n.



(b) Discrete time $n+1$.

Figure 3.1: Illustration of the discrete-time MPC scheme for a tracking problem with piecewise constant controls in time. The first part of the open-loop optimal control sequence is applied, then the horizon is shifted and the procedure is repeated. Past values are represented by dashes.

When passing from the infinite-horizon formulation to the MPC scheme, a priori it is not clear, at all, whether we will obtain approximately optimal trajectories. In fact, it is not even clear whether the closed-loop system is asymptotically stable.

One way to enforce stability is to add terminal conditions to ($\mathrm{OCP_N}$). In the PDE setting, this approach has been investigated, e.g., in [59, 28, 27]. Terminal constraints are added to the state constraints $\mathbb{X}$, while terminal costs influence the cost functional J_N. However, constructing a suitable terminal region or finding an appropriate terminal cost is a challenging task, cf. [49]. MPC schemes that do not rely on these methods are much easier to set up and implement and are therefore often preferred in practice. In this case, the choice of the horizon length N in step 0 of the MPC algorithm is crucial: Longer horizons make the problem computationally harder; shorter horizon lengths may lead to instability of the MPC closed loop. Therefore, the smallest horizon that yields a stabilizing feedback is of particular interest, both from the theoretical and practical point of view.

A key difference for the analysis of MPC schemes lies in the stage cost ℓ: Given some *equilibrium pair* $(\bar{z}, \bar{u})$ of (3.1), i.e., $f(\bar{z}, \bar{u}) = \bar{z}$, the question is whether ℓ is positive definite with respect to $(\bar{z}, \bar{u})$ or not. In the former case, we want to stabilize that desired equilibrium, hence the name *stabilizing MPC*. A prime example is the stage cost

$$\ell(z(k), u(k)) = \frac{1}{2} \|z(k) - \bar{z}\|^2 + \frac{\gamma}{2} \|u(k) - \bar{u}\|^2, \tag{3.4}$$

for some norm $\|\cdot\|$ and some weight $\gamma > 0$. This case is considered in Section 3.2.

The above stage cost, however, has a notable disadvantage: one needs to know the corresponding $\bar{u}$ for a desired $\bar{z}$ beforehand, which may be cumbersome to compute. A stage cost that is less complicated to design and thus easier to implement is

$$\ell(z(k), u(k)) = \frac{1}{2} \|z(k) - \bar{z}\|^2 + \frac{\gamma}{2} \|u(k)\|^2. \tag{3.5}$$

This function is also more common in optimal-control literature and structurally similar to the cost functional (2.9). Moreover, from a performance point of view it may be more desirable to penalize the control effort, anyway. For $\bar{u} \neq 0$, the new stage cost ℓ is not positive definite with respect to $(\bar{z}, \bar{u})$ since $\ell(\bar{z}, \bar{u}) \neq 0$.[1] The specific stage cost (3.5) models a so-called *unreachable setpoint problem* [82], which is a particular instance of an *economic MPC* problem. This setting is considered in Section 3.3.

The conceptual difference between stabilizing and economic MPC is that, instead of stabilizing a prescribed equilibrium pair $(\bar{z}, \bar{u})$ via a stage cost that is positive definite with respect to that pair, in economic MPC the interplay of the stage cost and dynamics determines the optimal (long-term) behavior.

3.2 Stabilizing MPC

In this section we consider stage costs ℓ that are positive definite with respect to an equilibrium pair $(\bar{z}, \bar{u})$ that we want to attain. More specifically, we assume ℓ to be of type (3.4). The goal of this section is to list known results and tools to prove asymptotic stability of the MPC closed loop.

Similar to [2], in the case of stabilizing MPC we rely on a stability condition proposed in [49] that, together with the *exponential controllability* property below, ensures a relaxed Lyapunov inequality to hold, cf. [49, Thm. 6.15 and Prop. 6.18]. This inequality has been introduced in [66] to guarantee stability of the MPC closed-loop solution.

Definition 3.3. *The system* (3.1) *is called exponentially controllable with respect to the stage cost* $\ell :\Leftrightarrow \exists C \geq 1, \delta \in]0,1[$ *such that for each state* $\tilde{z} \in Z$ *there exists a control* $u_{\tilde{z}} \in U$ *satisfying*

$$\ell(z_{u_{\tilde{z}}}(k; \tilde{z}), u_{\tilde{z}}(k)) \leq C\delta^k \min_{u \in U} \ell(\tilde{z}, u) \tag{3.6}$$

for all $k \in \mathbb{N}_0$.

Using the stability condition from [49], an upper bound for the minimal stabilizing horizon can be deduced from the values of the overshoot bound C and the decay rate δ.

[1] Redefining $\ell_2(z, u) := \ell(z, u) - \ell(\bar{z}, \bar{u})$ usually does not help as it may lead to $\ell_2(z, u) < 0$ for some (z, u).

For more details, see [2]. The most important difference in the influence of C and δ for our study is that for fixed C, it is generally impossible to arbitrarily reduce the horizon N by reducing δ. However, for $C = 1$, stability can be ensured even for the shortest meaningful horizon $N = 2$.

The condition (3.6) depends on the stage cost ℓ. In particular, ℓ being positive definite with respect to an equilibrium pair $(\bar{z}, \bar{u})$ is a necessary condition for the following theorem resulting from [49, Thm. 6.20 and Sect. 6.6] to hold.

Theorem 3.4. *Consider the MPC scheme with stage cost* (3.4) *satisfying the exponential controllability property from Definition 3.3 with* $C \geq 1$ *and* $\delta \in]0, 1[$. *Then the following holds:*

(a) There exists some optimization horizon $\bar{N} \geq 2$ *such that the equilibrium* $\bar{z}$ *is globally asymptotically stable for the MPC closed loop for each optimization horizon* $N \geq \bar{N}$.

(b) If $C = 1$ *then* $\bar{N} = 2$.

In both cases, the optimal value function V_N *to the optimization problem* ($\mathrm{OCP_N}$) *is a Lyapunov function for the MPC closed loop, which in particular satisfies* $V_N(z_{\mathcal{F}}(n+1)) < V_N(z_{\mathcal{F}}(n))$ *whenever* $V_N(z_{\mathcal{F}}(n)) \neq 0$.

This result states that the MPC closed loop (3.2) has the same qualitative stability property as the solution of the infinite-horizon optimal control problem (OCP_∞). Note that the control $u_{\tilde{z}}$ in Definition 3.3 does not have to be optimal. Thus, in order to apply Theorem 3.4, the main task is to find a (suboptimal) control $u_{\tilde{z}}$ that satisfies condition (3.6), preferably with $C = 1$.

Theorem 3.4 requires the exponential controllability property to hold globally. However, even if it only holds in a neighborhood of the equilibrium $\bar{z}$, the MPC algorithm yields an asymptotically stable closed loop on suitable recursively feasible sets (outside areas of "bad" behavior), provided the horizon is large enough [14]. Here, recursive feasibility is defined in the sense of forward invariance with respect to the MPC feedback law $\mathcal{F}$: A set $\mathcal{C} \subseteq \mathbb{X}$ is *recursively feasible* if for all $z \in \mathcal{C}$ we have $\mathcal{F}(z) \in \mathbb{U}$ and $f(z, \mathcal{F}(z)) \in \mathcal{C}$. The following result is a special case of [14, Thm. 6].

Theorem 3.5. *Consider the MPC scheme with stage cost* (3.4) *satisfying the exponential controllability property from Definition 3.3 with* $C \geq 1$ *and* $\delta \in]0, 1[$ *on a neighborhood* $\mathcal{M}$ *of* $\bar{z}$. *Let* $\mathcal{C} \subset V_\infty^{-1}[0, +\infty[\setminus \mathcal{O}$ *be a compact set, where*

$$\mathcal{O} := \lim_{n\to\infty} V_\infty^{-1}[n, +\infty[= \bigcap_{n\in\mathbb{N}} \overline{V_\infty^{-1}[n, +\infty[} \tag{3.7}$$

and $V_\infty^{-1}[n, +\infty[= \{z \in \mathbb{X} : n \leq V_\infty(z) < \infty\}$. *Then there exists some optimization horizon* $\bar{N}_{\mathcal{C}} \geq 2$ *such that for every* $N \geq \bar{N}_{\mathcal{C}}$ *the MPC closed loop is asymptotically stable with basin of attraction* $\mathcal{S} \supseteq \mathcal{C}$, *and* $\mathcal{S}$ *is recursively feasible.*

In addition to this qualitative stability property, the results from [49] also yield that the MPC closed loop is approximately optimal for (OCP_∞), i.e., that the MPC closed loop is quantitatively similar to the solution of the infinite-horizon problem. A selection of these *performance results* are introduced in the next section in the more general framework of economic MPC. These results can also be applied in the stabilizing MPC case, see Remark 3.9. Performance results tailored to the setting of stabilizing MPC, which yield better quantitative results than the ones presented in the next section, can be found in [49, Sect. 6].

3.3 Economic MPC

In this section, we weaken the requirement on the stage cost ℓ from the previous section. More specifically, we consider stage costs ℓ of type (3.5), which model an unreachable setpoint problem. We concern ourselves with both stability and performance of the MPC closed loop in this new setting.

Throughout this section, (z^e, u^e) denotes an equilibrium pair, i.e., $f(z^e, u^e) = z^e$. Whenever clear from the context, we might omit the word "pair". Although we do not stabilize a prescribed equilibrium, equilibria stay equally important. However, the definition of the decisive optimal equilibrium changes.

Definition 3.6 (Optimal Equilibrium)**.** *An equilibrium $(z^e, u^e) \in \mathbb{X} \times \mathbb{U}$ is called* optimal $:\Leftrightarrow \forall (z,u) \in \mathbb{X} \times \mathbb{U}$ *with* $f(z,u) = z : \ell(z^e, u^e) \leq \ell(z,u)$.

Assuming an equilibrium (z^e, u^e) exists and if f and ℓ are continuous and $\mathbb{X} \times \mathbb{U}$ is compact, then an optimal equilibrium exists, see, e.g., [49, Lemma 8.4]. It can be computed by solving the optimization problem

$$\min_{(z,u)\in\mathbb{X}\times\mathbb{U}} \ell(z,u) \quad \text{s.t. } z - f(z,u) = 0. \tag{3.8}$$

The next question is under which circumstances—if at all—the optimal equilibrium is asymptotically stable for the MPC closed loop. In [3, 53] it was shown that *strict dissipativity* is the decisive property. In order to define it, we use the notation

$$|z_1|_{z_2} := d_Z(z_1, z_2) \tag{3.9}$$

for the distance from $z_1 \in Z$ to $z_2 \in Z$ and recall the notion of comparison functions, which were introduced by Hahn in [55] and became increasingly popular since Sontag's work on input-to-state stability [90].

Definition 3.7 (Comparison functions)**.** *(a) Let $\alpha\colon \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ be a continuous function. Then*

- *$\alpha \in \mathcal{K}$ $:\Leftrightarrow$ α is strictly increasing and $\alpha(0) = 0$,*
- *$\alpha \in \mathcal{K}_\infty :\Leftrightarrow \alpha \in \mathcal{K}$ and α is unbounded,*
- *$\alpha \in \mathcal{L}$ $:\Leftrightarrow$ α is strictly decreasing and $\lim_{t\to\infty} \alpha(t) = 0$.*

(b) A continuous function $\beta\colon \mathbb{R}_{\geq 0} \times \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ is called a $\mathcal{KL}$ function $:\Leftrightarrow \forall t \geq 0 : \beta(\cdot, t) \in \mathcal{K}$ *and* $\forall r > 0 : \beta(r, \cdot) \in \mathcal{L}$.

Similar to how the exponential controllability property from the stabilizing MPC case was tied to the stage cost ℓ, strict dissipativity depends on ℓ:

Definition 3.8 ((Strict) Dissipativity, Storage Function, Modified Cost)**.** *(a) The optimal control problem* ($\mathrm{OCP_N}$) *with stage cost ℓ is called* strictly dissipative *at an equilibrium pair $(z^e, u^e) \in \mathbb{X} \times \mathbb{U}$ if there exist a function $\lambda\colon \mathbb{X} \to \mathbb{R}$ that is bounded from below and satisfies $\lambda(z^e) = 0$ and a function $\varrho \in \mathcal{K}_\infty$ such that for all $(z,u) \in \mathbb{X} \times \mathbb{U}$:*

$$\ell(z,u) - \ell(z^e, u^e) + \lambda(z) - \lambda(f(z,u)) \geq \varrho(|z|_{z^e}). \tag{3.10}$$

(b) If (a) holds with $\varrho \equiv 0$ then the optimal control problem is called dissipative.

(c) *The function λ in (a) is called* storage function.

(d) *The left-hand side of* (3.10)*, i.e.,*

$$\tilde{\ell}(z,u) := \ell(z,u) - \ell(z^e,u^e) + \lambda(z) - \lambda(f(z,u)), \tag{3.11}$$

is called modified cost *or* rotated cost.

Remark 3.9. *(a) Note that the requirement $\lambda(z^e) = 0$ in Definition 3.8(a) can always be satisfied by a constant translation of λ without influencing the inequality* (3.10)*.*

(b) The OCPs considered in the stabilizing MPC case in Section 3.2 are strictly dissipative at the desired equilibrium $(\bar{z},\bar{u})$ with a storage function $\lambda \equiv 0$:

$$\ell(z,u) - \ell(\bar{z},\bar{u}) = \ell(z,u) = \frac{1}{2}\|z-\bar{z}\|^2 + \frac{\gamma}{2}\|u-\bar{u}\|^2 \geq \frac{1}{2}\|z-\bar{z}\|^2 =: \varrho(|z|_{\bar{z}}).$$

(c) Although Definition 3.8 is formulated for general equilibria, if an OCP is strictly dissipative at a particular equilibrium (z^e,u^e), then this equilibrium is optimal, cf. [49, Prop. 8.9]. Hence, we only need to check strict dissipativity at optimal equilibria. From the same proposition we get the so-called optimal operation at steady-state*, i.e., that for all $z \in \mathbb{X}$ and for all admissible $\mathbf{u} \in \mathbb{U}^\infty$,*

$$\limsup_{M\to\infty} \frac{1}{M} \sum_{k=0}^{M-1} \ell(z_{\mathbf{u}}(k;z),u(k)) \geq \ell(z^e,u^e). \tag{3.12}$$

Under additional controllability assumptions, this property implies (non-strict) dissipativity, cf. [71].

In a classical interpretation of (3.10), $\lambda(z)$ serves as a quantifier for the amount of energy stored at state z, $\ell(z,u) - \ell(z^e,u^e)$ can be viewed as a *supply rate* that tracks the amount of energy supplied to or withdrawn from the system via the control u, and $\varrho(|z|_{z^e})$ is the amount of energy the system releases (or *dissipates*) to the environment in each step. Note, however, that in the optimal control problems we discuss here there is not necessarily a notion of "energy" in a physical sense.

Strict dissipativity is the main required property in the subsequent stability and performance results. As such, the focus will be on that property. However, in addition, we require appropriate continuity properties. For the sake of completeness, these more technical requirements will be introduced next. To this end, analogous to the optimal value functions from Definition 3.2, we define $\tilde{V}_N(\mathring{z}) := \inf_{\mathbf{u}} \tilde{J}_N(\mathring{z},\mathbf{u})$ where, similar to the modified cost $\tilde{\ell}$, $\tilde{J}_N$ is given by

$$\tilde{J}_N(z,\mathbf{u}) := J_N(z,\mathbf{u}) - N\ell(z^e,u^e) + \lambda(z) - \lambda(z_{\mathbf{u}}(N;z)). \tag{3.13}$$

Assumption 3.10 (Continuity of $\lambda, V_N, \tilde{V}_N$, and V_∞ at z^e)**.**

(a) $\exists \gamma_\lambda \in \mathcal{K}_\infty \quad \forall z \in \mathbb{X}: \quad |\lambda(z) - \lambda(z^e)| \leq \gamma_\lambda(|z|_{z^e})$

(b) $\exists \gamma_V \in \mathcal{K}_\infty, \omega \in \mathcal{L} \quad \forall z \in \mathbb{X}, N \in \mathbb{N}: |V_N(z) - V_N(z^e)| \leq \gamma_V(|z|_{z^e}) + \omega(N)$

(c) $\exists \gamma_{\tilde{V}} \in \mathcal{K}_\infty \quad \forall z \in \mathbb{X}, N \in \mathbb{N}: |\tilde{V}_N(z) - \tilde{V}_N(z^e)| \leq \gamma_{\tilde{V}}(|z|_{z^e})$

(d) $\exists \gamma_{V_\infty} \in \mathcal{K}_\infty \quad \forall z \in \mathbb{X}: \quad |V_\infty(z) - V_\infty(z^e)| \leq \gamma_{V_\infty}(|z|_{z^e})$

Since, in general, neither V_N, nor $\tilde{V}_N$ nor V_∞ are known, the above continuity assumptions are difficult to verify. This problem can be circumvented by sufficient conditions for Assumption 3.10 that may be easier to show.

Definition 3.11 (Local controllability)**.** *The system* (3.1) *is called* locally controllable *at z^e if there exist a neighborhood E of z^e, a time $s \in \mathbb{N}$ and functions $\gamma_z, \gamma_u, \gamma_c \in \mathcal{K}_\infty$ such that for any $z_0, z_1 \in E$ there exists a control $u \in \mathbb{U}^s$ satisfying*

$$\begin{aligned} z_{\mathbf{u}}(s; z_0) &= z_1, \\ \|z_{\mathbf{u}}(k; z_0) - z^e\| &\leq \gamma_z(\delta), \\ \|u(k) - u^e\| &\leq \gamma_u(\delta), \\ \|\ell(z_{\mathbf{u}}(k; z_0), u(k)) - \ell(z^e, u^e)\| &\leq \gamma_c(\delta), \end{aligned} \tag{3.14}$$

for $\delta := \max\{\|z_0 - z^e\|, \|z_1 - z^e\|\}$ and all $k = 0, \dots, s-1$.

The following proposition is taken from [47, Prop. 5.6] and is extended to $\tilde{V}_N$.

Proposition 3.12. *Assume* ($\mathrm{OCP_N}$) *is strictly dissipative at (z^e, u^e) with a bounded storage function λ.*

(a) If the system (3.1) *is locally controllable at z^e, then Assumptions 3.10(b) and (d) hold.*

(b) Let Assumption 3.10(a) hold. If the system (3.1) *is locally controllable at z^e with $\tilde{\ell}$ instead of ℓ in* (3.14)*, then Assumption 3.10(c) holds.*

The optimal value functions in Assumption 3.10 are used as Lyapunov functions in order to conclude stability of the MPC closed loop. In the stabilizing MPC case, the proof of Theorem 3.4 relies on using V_N as a Lyapunov function. The argument can be adapted to the economic MPC case by using $\tilde{V}_N$ as a *practical* Lyapunov function for the modified cost $\tilde{\ell}$, cf. [49, Sect. 8.6].[2] The drawback is that we only get *semiglobal practical stability*.

Theorem 3.13 (Stability result)**.** *Consider the MPC scheme with an optimal control problem* ($\mathrm{OCP_N}$) *that is strictly dissipative at (z^e, u^e) with a bounded storage function λ. Moreover, let Assumption 3.10(a)-(c) hold. Then the equilibrium z^e is semiglobally practically asymptotically stable on $\mathbb{X}$ with respect to the optimization horizon N, i.e., there exists $\beta \in \mathcal{KL}$ such that the following holds: for each $\delta, \Delta > 0$ there exists $N_{\delta,\Delta} \in \mathbb{N}$ such that for all $N \geq N_{\delta,\Delta}$ and all $\mathring{z} \in \mathbb{X}$ with $|\mathring{z}|_{z^e} \leq \Delta$ the inequality*

$$|z_{\mathcal{F}_N}(k; \mathring{z})|_{z^e} \leq \max\{\beta(|\mathring{z}|_{z^e}, k), \delta\} \tag{3.15}$$

holds for all $k \in \mathbb{N}_0$.

Semiglobal practical asymptotic stability is a relaxation of global asymptotic stability in two ways: "Semiglobal", because we are limiting the initial values to all $\mathring{z} \in \mathbb{X}$ with $|\mathring{z}|_{z^e} \leq \Delta$. "Practical", because in (3.15) we only require asymptotic stability until the trajectory reaches a δ-neighborhood of z^e, see Figure 3.2. Both δ and Δ can be made arbitrarily small and large, respectively, but not for a fixed optimization horizon N.

[2] V_N cannot be used since the optimal trajectories for ℓ and $\tilde{\ell}$ do not have to coincide due to the last, $\mathbf{u}$-dependent term in (3.13).

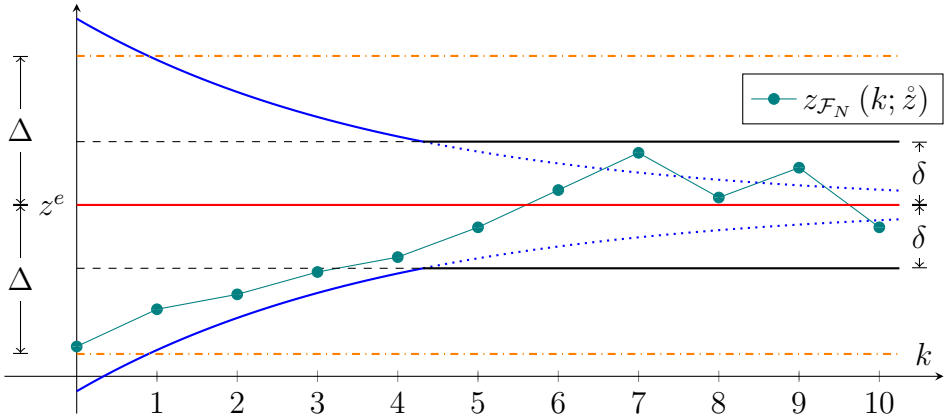


Figure 3.2: Illustration of semiglobal practical asymptotic stability. The blue tube (first solid, then dotted) is defined by $\beta(|\mathring{z}|_{z^e}, k)$. The blue and black solid lines represent $\max\{\beta(|\mathring{z}|_{z^e}, k), \delta\}$.

Under assumptions similar to Theorem 3.13, we can state results regarding the performance of the MPC closed loop in the following theorem. For more details, we refer to [49, Sects. 8.5–8.7].

Theorem 3.14 (Performance results)**.** *Consider the MPC scheme with a strictly dissipative optimal control problem* ($\mathrm{OCP_N}$) *at* (z^e, u^e) *with a bounded storage function* λ *and let Assumption 3.10(a)-(b) hold.*

(a) Assume that $\ell(z^e, u^e) = 0$*, that* $\mathbb{X}$ *is bounded and let Assumption 3.10(d) hold.*[3] *Then there exists* $\delta_1 \in \mathcal{L}$ *such that the* non-averaged finite-horizon closed-loop performance

$$J_M^{cl}(\mathring{z}, \mathcal{F}) := \sum_{k=0}^{M-1} \ell(z_{\mathcal{F}}(k;\mathring{z}), \mathcal{F}(z_{\mathcal{F}}(k;\mathring{z})))$$

satisfies the inequality

$$J_M^{cl}(z, \mathcal{F}_N) + V_\infty(z_{\mathcal{F}_N}(M; z)) \leq V_\infty(z) + M\delta_1(N) \tag{3.16}$$

for all $z \in \mathbb{X}$*,* $M \in \mathbb{N}$ *and sufficiently large* $N \in \mathbb{N}$*.*

(b) Assume that V_N *is bounded from below on* $\mathbb{X}$*. Then there exists* $\delta_1 \in \mathcal{L}$ *such that for any* $N \geq 2$ *and any* $z \in \mathbb{X}$ *the* averaged infinite-horizon closed-loop performance

$$\overline{J}_\infty^{cl}(\mathring{z}, \mathcal{F}) := \limsup_{M\to\infty} \frac{1}{M} J_M^{cl}(\mathring{z}, \mathcal{F})$$

satisfies the inequality

$$\overline{J}_\infty^{cl}(z, \mathcal{F}_N) \leq \ell(z^e, u^e) + \delta_1(N). \tag{3.17}$$

[3] One can always satisfy $\ell(z^e, u^e) = 0$ by translating ℓ. This does not affect the optimal trajectory.

(c) Let $\mathbb{U}_\kappa^M(z) := \left\{u \in \mathbb{U}^M \mid z_{\mathbf{u}}(M;z) \in \overline{\mathcal{B}}_\kappa(z^e)\right\}$*, where* $\overline{\mathcal{B}}_\kappa(z^e)$ *denotes the closed ball around* z^e *with radius* κ*. Assume that* $\mathbb{X}$ *is bounded and let Assumption 3.10(c) hold. Then there exist* $\delta_1, \delta_2, \delta_3 \in \mathcal{L}$ *such that for all* $z \in \mathbb{X}$ *the inequality*

$$J_M^{cl}(z, \mathcal{F}_N) \leq \inf_{u \in \mathbb{U}_\kappa^M(z)} J_M(z,u) + \delta_1(N) + M\delta_2(N) + \delta_3(M) \tag{3.18}$$

holds with $\kappa \geq 0$*, where* κ *depends on* M, N *(each monotonically decreasing), and* $|z|_{z^e}$ *(monotonically increasing).*

Theorem 3.14(a) states that by following the MPC closed loop up until step M and then switching to the infinite-horizon optimal control starting from that point, the error made compared to using the infinite-horizon optimal control from the beginning can be quantified by $M\delta_1(N)$ with $\delta_1(N) \to 0$ as $N \to \infty$. For fixed N and increasing M, this error increases. However, from Theorem 3.14(b) we infer that the MPC closed-loop solution does not entirely deteriorate, as the average performance behaves well even for $M \to \infty$. Finally, we remark that the assumptions of Theorem 3.14(a)-(c) imply those of Theorem 3.13, i.e., the MPC closed loop is semiglobally practically asymptotically stable with respect to N. The phase until the closed-loop system reaches the δ-neighborhood of z^e is called the transient phase. The conclusion from Theorem 3.14(c) is that—up to some error terms—the MPC closed loop has the best transient performance.

In summary, strict dissipativity is *the* decisive structural property that makes MPC work. This is the main motivation why we analyze it in Chapter 6. Thereby, its relation to another important property of optimal control problems, the so-called *turnpike property*, will be utilized. This classical property in optimal control originated in mathematical economy, cf. [26], and recently attracted significant attention in the PDE control community, cf., e.g., [93]. It demands that there exists a function $\sigma \in \mathcal{L}$ such that for all $N, P \in \mathbb{N}$, $z \in \mathbb{X}$, and the optimal trajectories $z^*(k;z)$ with horizon N, the set

$$\mathcal{Q}(z,u,P,N) := \{k \in \{0,\dots,N-1\} \mid |z^*(k;z)|_{z^e} \geq \sigma_\delta(P)\} \tag{3.19}$$

has at most P elements. In words, most of the time the finite-horizon optimal trajectories stay close to the optimal equilibrium z^e.[4] This is exemplarily illustrated in Figure 3.3.

Under a boundedness condition on the optimal value function (known as *cheap reachability*, for which Assumption 3.10(b) is sufficient), it can be shown that strict dissipativity implies the turnpike property and under a controllability condition, these two properties are even equivalent [48]. Unsurprisingly, the turnpike property can be used to deduce stability[5] and performance results, see [47, 94]. Moreover, it is often a good indicator for strict dissipativity. In contrast to strict dissipativity, the turnpike property is more difficult to check analytically, because it involves the knowledge of optimal trajectories. On the other hand, the turnpike property is more easily checked numerically by means of simulating optimal trajectories. Hence, these two properties complement each other in a nice way when analyzing strict dissipativity of optimal control problems. Figure 3.4 gives an overview of the relations between strict dissipativity, the turnpike property and the above-discussed desired properties of the MPC closed loop.

[4] There are several distinctions of turnpike behavior, see, e.g., [48, Def. 2.2] and [49, Props. 8.15, 8.18]

[5] under additional assumptions such as terminal constraints.

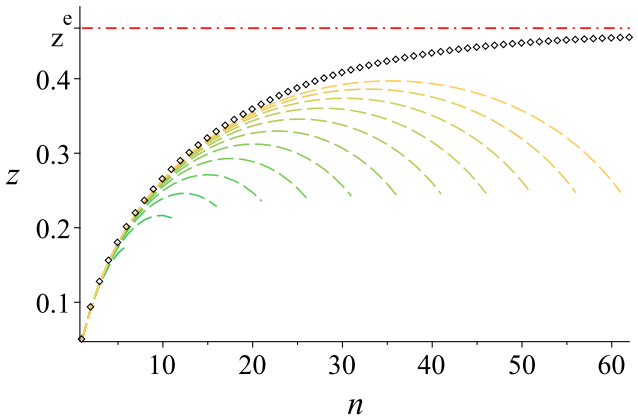


Figure 3.3: Open-loop optimal trajectories for $N = 2, 6, 11, 16, 21, ..., 61$ (dashed), closed-loop trajectory (black dots), and optimal equilibrium z^e (red dash-dot) for Example 6.23.

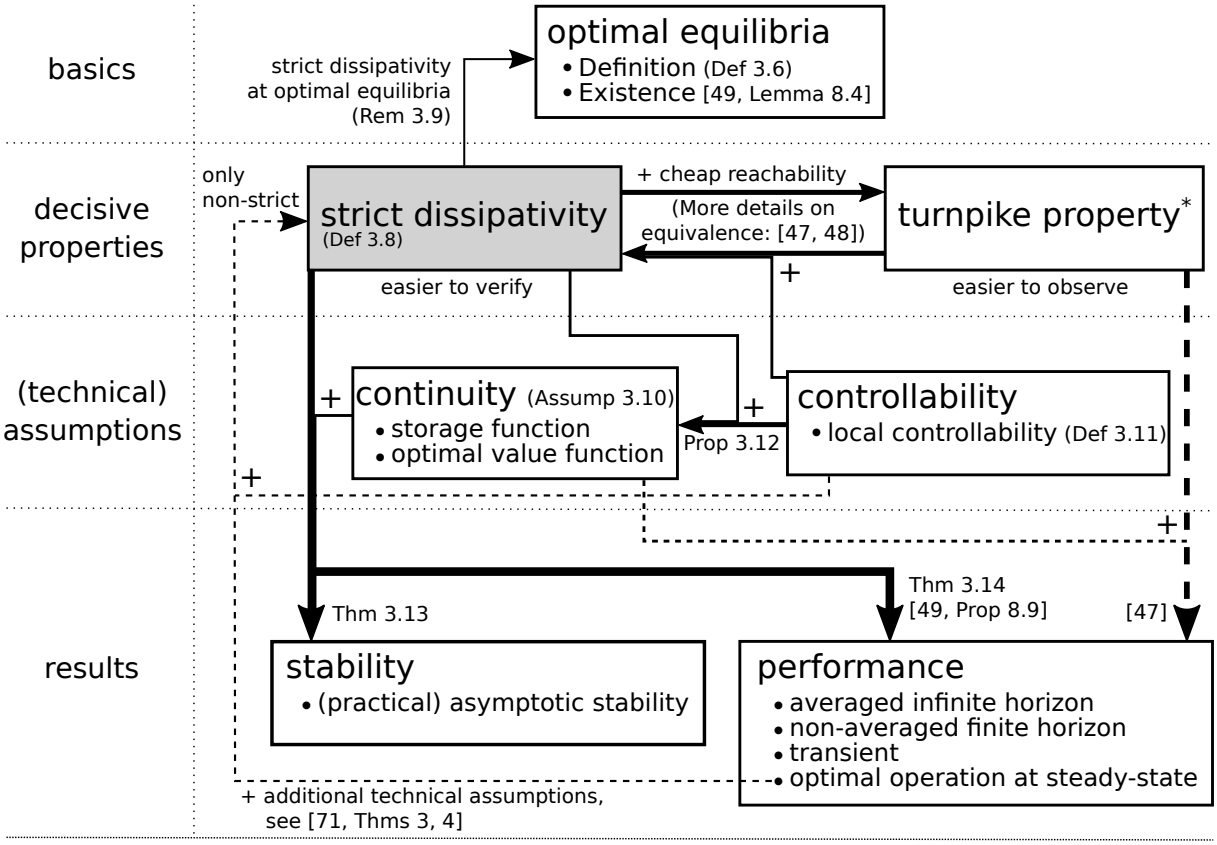


Figure 3.4: Relations between strict dissipativity, the turnpike property, and stability and performance of the MPC closed loop in the economic MPC setting.

Stabilizing MPC – Space-independent control

4

Having introduced MPC, we begin our study of the behavior of the MPC closed loop corresponding to OCPs of type (1.9). In this chapter we consider the task of steering the state, in this case the PDF, asymptotically to a desired equilibrium. Different classes of control functions can be used in this setting. Those that do not depend on space in the Fokker–Planck equation, i.e., control inputs that are independent of the current state of the stochastic process, are particularly easy to implement. This class of functions was used in [4, 5] and is also considered in this chapter, as a starting point.

In the MPC scheme, cf. Section 3.1, the length of the finite optimization horizon directly influences the numerical effort required for solving these problems: the shorter the horizon, the faster the numerical solution. On the other hand, long horizons may be needed in order to obtain stability of the resulting MPC closed loop, cf. [49, 2]. While numerical results in [4, 5, 36] indicate that for the Fokker–Planck equation very short optimization horizons are sufficient for obtaining stability, a formal proof of this fact is to the best of our knowledge missing up to now.

In this chapter, we close this gap for the Fokker–Planck equation corresponding to the controlled Ornstein–Uhlenbeck process using an L^2 cost and control functions that are constant in space but may be time-dependent. We show that for normally distributed PDFs, stability can always be achieved, even when looking only one time step into the future, thus resulting in the simplest possible optimal control problem with a constant control function in each MPC iteration. Our analysis relies on an exponential controllability condition for the considered stage cost, which is established for three different cases depending on the ratio of the variance of the initial PDF to the variance of the desired PDF. We employ a suitably chosen equivalent stage cost for one of the cases.

The remainder of the chapter is organized as follows. Section 4.1 defines the problem setting, particularly the Fokker–Planck equation we are going to control. Section 4.2 contains the main stability result, which is obtained by checking the exponential controllability condition from Section 3.2. Our results are illustrated by numerical examples in Section 4.3 before we conclude this chapter in Section 4.4.

4.1 Problem Setting

In this chapter we consider the (d-dimensional extension of the) Ornstein–Uhlenbeck process (1.7) introduced in Section 1.1. In contrast to Chapter 2, the control u is assumed

to be only time-dependent. The associated Fokker–Planck equation (1.2) reads

$$\partial_t \rho(x,t) - \frac{1}{2}\sum_{i=1}^{d} \varsigma_i^2 \partial_{ii}^2 \rho(x,t) + \sum_{i=1}^{d} \partial_i \left([-\theta_i x_i + u_i(t)]\, \rho(x,t) \right) = 0 \quad \text{in } Q, \tag{4.1a}$$

$$\rho(\cdot,0) = \mathring{\rho}(\cdot) \text{ in } \Omega. \tag{4.1b}$$

We consider the Gaussian setting with $\Omega = \mathbb{R}^d$ (and thus $Q = \mathbb{R}^d \times]0,T[$) and moreover assume that the initial PDF $\mathring{\rho}$ is a (multivariate) Gaussian PDF with mean $\mathring{\mu} \in \mathbb{R}^d$ and covariance matrix $\mathring{\Sigma} = \operatorname{diag}(\mathring{\sigma}_1^2, ..., \mathring{\sigma}_d^2)$ with $\mathring{\sigma}_i > 0$, $i = 1, ..., d$, i.e.,

$$\mathring{\rho}(x) = \left((2\pi)^d \prod_{i=1}^{d} \mathring{\sigma}_i^2 \right)^{-1/2} \exp\left(-\sum_{i=1}^{d} \frac{(x_i - \mathring{\mu}_i)^2}{2\mathring{\sigma}_i^2} \right).$$

For constant controls $u_i(t) \equiv \bar{u}_i \in \mathbb{R}$, the solution of the Fokker–Planck equation (4.1) exists in closed form, cf. [4] for the 1D case, which can be straightforwardly extended to the d-dimensional setting:

$$\rho(x,t;\bar{u}) = \left((2\pi)^d \prod_{i=1}^{d} \sigma_i^2(t) \right)^{-1/2} \exp\left(-\sum_{i=1}^{d} \frac{(x_i - \mu_i(t;\bar{u}_i))^2}{2\sigma_i^2(t)} \right), \tag{4.2}$$

where

$$\mu_i(t;\bar{u}_i) := \frac{\bar{u}_i}{\theta_i} + \left(\mathring{\mu}_i - \frac{\bar{u}_i}{\theta_i} \right) e^{-\theta_i t} \qquad \text{and} \qquad \sigma_i^2(t) := \frac{\varsigma_i^2}{2\theta_i} + \left(\mathring{\sigma}_i^2 - \frac{\varsigma_i^2}{2\theta_i} \right) e^{-2\theta_i t}.$$

Note that since the control is space-independent, it only affects the mean of the distribution, not its variance. For $i = 1, ..., d$ we define

$$\bar{\mu}_i := \frac{\bar{u}_i}{\theta_i} \quad \text{and} \quad \bar{\sigma}_i^2 := \frac{\varsigma_i^2}{2\theta_i}.$$

Then as $t \to \infty$, $\rho(x,t;\bar{u})$ converges to

$$\begin{aligned} \bar{\rho}(x;\bar{u}) :=& \left((2\pi)^d \prod_{i=1}^{d} \frac{\varsigma_i^2}{2\theta_i} \right)^{-1/2} \exp\left(-\sum_{i=1}^{d} \frac{\left(x_i - \frac{\bar{u}_i}{\theta_i} \right)^2}{\frac{\varsigma_i^2}{\theta_i}} \right) \\ =& \left((2\pi)^d \prod_{i=1}^{d} \bar{\sigma}_i^2 \right)^{-1/2} \exp\left(-\sum_{i=1}^{d} \frac{(x_i - \bar{\mu}_i)^2}{2\bar{\sigma}_i^2} \right). \end{aligned}$$

In particular, given any constant control $u \equiv \bar{u} \in \mathbb{R}^d$, the PDF $\bar{\rho}$ is an equilibrium solution of (4.1). We want to steer from some given initial PDF $\mathring{\rho}$ to such a target PDF $\bar{\rho}$. Of course, this can be achieved simply by applying the corresponding constant control $\bar{u}$. However, our goal is to reach the target quicker and/or more cheaply with respect to some cost function. To calculate a control that achieves this, we use MPC, cf. Chapter 3. Thus, the problem we consider is, given $\mathring{\rho}$ and $\bar{\rho}$, we want to solve ($\mathrm{OCP_N}$) for stage costs of type (3.4). In this chapter, the stage cost is defined by

$$\ell(\rho(k), u(k)) = \frac{1}{2} \left\| \rho(k) - \bar{\rho} \right\|_{L^2(\mathbb{R}^d)}^2 + \frac{\gamma}{2} |u(k) - \bar{u}|^2, \tag{4.3}$$

where $\rho(k)$ denotes the solution ρ to (4.1), sampled at discrete time step $k \in \mathbb{N}_0$, and $|\cdot|$ is the Euclidean norm. Hence, we want to minimize

$$J_N(\mathring{\rho}, \mathbf{u}) := \sum_{k=0}^{N-1} \left[\frac{1}{2} \|\rho(k) - \bar{\rho}\|^2_{L^2(\mathbb{R}^d)} + \frac{\gamma}{2} |u(k) - \bar{u}|^2 \right].$$

For $N = 2$, the resulting objective function

$$J_2(\mathring{\rho}, \mathbf{u}) = \frac{1}{2} \|\rho(0) - \bar{\rho}\|^2_{L^2(\mathbb{R}^d)} + \frac{\gamma}{2} |u(0) - \bar{u}|^2 + \frac{1}{2} \|\rho(1) - \bar{\rho}\|^2_{L^2(\mathbb{R}^d)} + \frac{\gamma}{2} |u(1) - \bar{u}|^2 \quad (4.4)$$

is equivalent to

$$\hat{J}_2(\mathring{\rho}, \mathbf{u}) := \frac{1}{2} \|\rho(1) - \bar{\rho}\|^2_{L^2(\mathbb{R}^d)} + \frac{\gamma}{2} |u(0) - \bar{u}|^2, \quad (4.5)$$

since the first term in (4.4) is a constant that cannot be influenced and the last term is always zero.[1]

The objective function (4.5) with $\bar{u} = 0$ is the type of cost functional used in [4, 5], albeit for general target probability density functions, which are not necessarily (equilibrium) solutions to the Ornstein–Uhlenbeck process. Often $|u|^2$ is used in the objective function rather than $|u - \bar{u}|^2$. Due to $\ell(\bar{\rho}, \bar{u}) \neq 0$, this case leads to economic MPC, see Section 3.3. Investigating the MPC closed loop in the framework of economic MPC is the topic of Chapter 6.

In this chapter, however, the question at hand is whether the MPC scheme yields a stabilizing control and if so, how to choose the horizon length N to guarantee stability of the MPC closed loop. The state space Z, cf. Section 3.1, is the space of normally distributed PDFs. To simplify the presentation, we focus on the one-dimensional case.

4.2 Stability of the MPC Closed-Loop Solution

In this section we analyze exponential controllability with respect to the stage cost (4.3) according to Definition 3.3 in order to estimate the minimal stabilizing horizon length depending on the overshoot C and the decay rate δ in (3.6).

One promising candidate for an exponentially stabilizing control sequence in (3.6) is the constant control $\bar{u}$. In this case, the second term in the stage cost (4.3) vanishes and the left-hand side of (3.6), which is given by $\ell(\rho(k), \bar{u})$, can be calculated explicitly[2] thanks to (4.2):

$$\ell(\rho(k), \bar{u}) = \frac{\sqrt{\theta}}{2\sqrt{2\pi\varsigma^2}} \left(1 + \frac{1}{\sqrt{\zeta(t_k)}} - \frac{2\sqrt{2}\exp(-\eta(t_k))}{\sqrt{\zeta(t_k) + 1}} \right), \quad (4.6)$$

where

$$\zeta(t) := 1 + (\alpha - 1)e^{-2\theta t} > 0, \qquad \alpha := \frac{2\theta\mathring{\sigma}^2}{\varsigma^2} = \frac{\mathring{\sigma}^2}{\bar{\sigma}^2} > 0,$$

$$\eta(t) := \frac{\beta e^{-2\theta t}}{\zeta(t) + 1} \geq 0, \qquad \beta := \frac{(\mathring{\mu} - \frac{\bar{u}}{\theta})^2}{\frac{\varsigma^2}{\theta}} = \frac{(\mathring{\mu} - \bar{\mu})^2}{2\bar{\sigma}^2} \geq 0.$$

[1] Since the control $u(1)$ only influences the subsequent states, which are not included in the objective function, choosing $u(1) = \bar{u}$ is always the best option when minimizing J_2.

[2] In this chapter, we rely on the explicit solution formula. The more general case, which is independent of such formulas, is provided in Lemma 5.5.

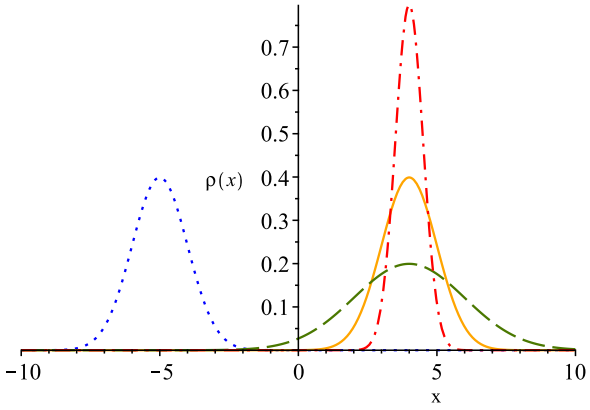


Figure 4.1: A sample desired PDF $\bar{\rho}(x)$ (dotted blue) and three initial PDFs $\mathring{\rho}(x)$ for $\alpha = 1$ (solid orange), $\alpha < 1$ (dashed green) and $\alpha > 1$ (dot-dashed red).

On the right-hand side of (3.6) we have

$$C\delta^k \min_u \ell(\mathring{\rho}, u) = C\delta^k \ell(\mathring{\rho}, \bar{u}).$$

Thus, to prove the exponential controllability property (3.6), we show that

$$\ell(\rho(t), \bar{u}) \leq Ce^{-\kappa t}\ell(\rho(0), \bar{u}) \tag{4.7}$$

in continuous time for some $\kappa > 0$, $C \geq 1$ and define $\delta := e^{-\kappa T_s}$, where T_s is the MPC sampling time, to arrive at (3.6). The constant C is the overshoot bound from Definition 3.3, δ is the decay rate. Since we can ignore the constant factor $\sqrt{\theta}/2\sqrt{2\pi\varsigma^2}$ in (4.6), this is equivalent to proving

$$W_\alpha(t) \leq Ce^{-\kappa t}W_\alpha(0), \tag{4.8}$$

where

$$W_\alpha(t) := 1 + \frac{1}{\sqrt{\zeta(t)}} - \frac{2\sqrt{2}\exp(-\eta(t))}{\sqrt{\zeta(t)+1}}. \tag{4.9}$$

Hence, in the following we show that (4.8) holds, and moreover with $C = 1$, as we then get stability for the shortest meaningful horizon length $N = 2$.

Before that, however, we give an interpretation of the above parameters α and β. These depend on the model parameters θ and ς^2 as well as on the initial PDF, which is characterized by $(\mathring{\mu}, \mathring{\sigma}^2)$. The value of β indicates the distance between the initial mean $\mathring{\mu}$ and the mean of the target equilibrium PDF $\bar{\rho}$. Similarly, the former parameter, α, relates the initial variance $\mathring{\sigma}^2$ to that of the target equilibrium PDF $\bar{\rho}$. If $\alpha = 1$, the variance does not change in time since $\mathring{\sigma}^2 = \varsigma^2/(2\theta)$ in (4.2). For $\alpha < 1$, the variance of the distribution is increasing in time since $\mathring{\sigma}^2 < \varsigma^2/(2\theta)$. Analogously, it shrinks in time if $\alpha > 1$. All cases are illustrated in Figure 4.1. We recall that we cannot control the variance, only the mean, see (4.2).

In order to conclude stability of the MPC closed loop from the exponential controllability condition (3.6), an exponentially stabilizing control needs to exist for the initial state $\mathring{z} = z_{\mathcal{F}}(n) = \rho(t_n, \cdot)$ in every MPC iteration. Hence, the value of α may change from

one step to the next, i.e., $\alpha_{n+1} \neq \alpha_n$, where α_n denotes the value of α in the n-th MPC iteration. It is important to note, however, that for space-independent control the sign of $\alpha_n - 1$ does not change with n. This is due to the monotone convergence of α_n to 1 that we get from reformulating the change in the variance in (4.2),

$$\mathring{\sigma}^2_{n+1} = \frac{\varsigma^2}{2\theta} + \left(\mathring{\sigma}^2_n - \frac{\varsigma^2}{2\theta}\right) e^{-2\theta T_s},$$

to

$$\alpha_{n+1} = 1 + (\alpha_n - 1)e^{-2\theta T_s}. \tag{4.10}$$

In order to prove (4.8) we now consider the three cases $\alpha = 1$, $\alpha < 1$, and $\alpha > 1$ separately.

The case $\alpha = 1$:

In this case, the shape of the PDF stays the same since the space-independent control can only move the PDF as a whole. We have

$$W_1(t) = 2 - 2e^{-\beta e^{-2\theta t}/2} \tag{4.11}$$

and we can prove the following proposition.

Proposition 4.1. *For $W_1(t)$, inequality (4.8) holds with $C = 1$ and $\kappa = 2\theta e^{-\beta/2}$.*

Proof. We show $W_1'(t) \leq -\kappa W_1(t)$ to conclude our assertion. To this end, consider

$$\begin{aligned} W_1'(t) + \kappa W_1(t) &= -4\theta \left(\frac{\beta}{2} e^{-2\theta t} e^{-\beta e^{-2\theta t}/2} - e^{-\beta/2} + e^{-\beta/2} e^{-\beta e^{-2\theta t}/2} \right) \\ &= -4\theta \left(e^{-\beta e^{-2\theta t}/2} \left[\frac{\beta}{2} e^{-2\theta t} + e^{-\beta/2} \right] - e^{-\beta/2} \right) \\ &= -4\theta \left(e^{-\tilde{\beta}\tau} \left[\tilde{\beta}\tau + e^{-\tilde{\beta}} \right] - e^{-\tilde{\beta}} \right), \end{aligned}$$

where $\tilde{\beta} := \beta/2 \geq 0$ and $\tau := e^{-2\theta t} \in]0, 1]$. For arbitrary but fixed $\tilde{\beta}$ we define the C^∞ function

$$h_1(\tau) := e^{-\tilde{\beta}\tau}(\tilde{\beta}\tau + e^{-\tilde{\beta}}) - e^{-\tilde{\beta}}.$$

It can easily be shown that $h_1(0) = 0$ and $h_1(1) \geq 0$. By calculating $h_1'(\tau)$, one can show that $h_1(\tau)$ is monotonously increasing on $]0, \tau^*[$, with $\tau^* := (1 - e^{-\tilde{\beta}})/\tilde{\beta}$ being the unique root of $h_1'(\tau)$, and monotonously decreasing on $]\tau^*, 1]$. Therefore, $h_1(\tau) \geq 0$ on $]0, 1]$, which concludes the proof. □

Since $C = 1$, the MPC closed loop is asymptotically stable even for the shortest possible horizon N.

The case $\alpha < 1$:

For $\alpha < 1$, the shape of the PDF becomes wider in time. Due to the nature of the L^2 cost (4.6), initially, the cost may be higher compared to $\alpha = 1$, i.e., $W_\alpha(0) \geq W_1(0)$. However, it also drops more quickly, i.e., $W'_\alpha(t) \leq W'_1(t)$. The idea is to prove

$$h_2(t) := W_1(0)W_\alpha(t) - W_\alpha(0)W_1(t) \leq 0, \tag{4.12}$$

since for $W_1(0) \neq 0$, which we can assume w.l.o.g., we then use Proposition 4.1 to get

$$W_\alpha(t) \leq \frac{W_\alpha(0)}{W_1(0)} W_1(t) \leq \frac{W_\alpha(0)}{W_1(0)} e^{-\kappa t} W_1(0) = e^{-\kappa t} W_\alpha(0)$$

for κ as in Proposition 4.1.

Obviously, $h_2(0) = 0$ and $\lim_{t\to\infty} h_2(t) = 0$. Analogously to the proof of Proposition 4.1, one can show there exists at most one root $t^* \in [0, \infty[$ of $h'_2(t)$ and that $h_2(t)$ is monotonously decreasing on $[0, t^*[$ (or $[0, \infty[$ in case there is no root of $h'_2(t)$ in $[0, \infty[$) and monotonously increasing on $]t^*, \infty[$. Hence, (4.12) holds and we have shown the following.

Proposition 4.2. *For $\alpha < 1$, $W_\alpha(t)$ satisfies (4.8) with C and κ from Proposition 4.1.*

The case $\alpha > 1$:

If $\alpha > 1$, the shrinking variance of the distribution may lead to increasing stage costs at the beginning, i.e., $W'_\alpha(t) > 0$ for $t \in [0, t^*[$ and some $t^* > 0$. This occurs, for instance, for $\theta = \mathring{\mu} = \varsigma = 1$, $\mathring{\sigma} = 100$, and control $\bar{u} = 2000$, cf. Figure 4.2. It is due to the L^2 norm used in the stage cost (4.3). Obviously, condition (4.8) does not hold for $C = 1$.

To circumvent this issue, we can add (time-dependent) control-independent terms to $W_\alpha(t)$. One possibility is to add

$$2\sqrt{2}|\alpha - 1|e^{-2\theta t} + 1 - \frac{1}{\sqrt{\zeta(t)}}$$

to $W_\alpha(t)$, which results in

$$\tilde{W}_\alpha(t) := 2\sqrt{2}\hat{W}_\alpha(t) := 2\sqrt{2}\left(|\alpha - 1|e^{-2\theta t} + \frac{1}{\sqrt{2}} - \frac{\exp(-\eta(t))}{\sqrt{\zeta(t) + 1}}\right). \tag{4.13}$$

Just as $W_\alpha(t)$ from (4.9) stemmed from the stage cost ℓ from (4.3), there exists a stage cost $\tilde{\ell}$ that, for $u \equiv \bar{u}$, yields $\tilde{W}_\alpha(t)$ in the one-dimensional case. A short calculation reveals that the terms added to $W_\alpha(t)$ can be formulated in terms of ρ and $\bar{\rho}$:

$$\tilde{\ell}(\rho, u) := \ell(\rho, u) + \frac{1}{2}\left(\|\bar{\rho}\|^2_{L^2(\mathbb{R})} - \|\rho\|^2_{L^2(\mathbb{R})}\right) + \sqrt{2}\,\|\bar{\rho}\|^2_{L^2(\mathbb{R})}\left|\frac{\|\bar{\rho}\|^4_{L^2(\mathbb{R})}}{\|\rho\|^4_{L^2(\mathbb{R})}} - 1\right|. \tag{4.14}$$

Note that $\tilde{\ell}$ yields the same optimal control sequence as ℓ from (4.3) and thus Theorem 3.4 can be applied to $\tilde{\ell}$. This is because the added terms to ℓ do not depend on the control u. At first glance, this might seem counterintuitive, as the PDF ρ of course does depend on u. The $L^2(\mathbb{R})$ norm of ρ, however, does not. At second glance the reason is clear: All the control u can do is shift the PDF ρ to the left or to the right. But moving a function does not change its L^2 norm on $\mathbb{R}$.

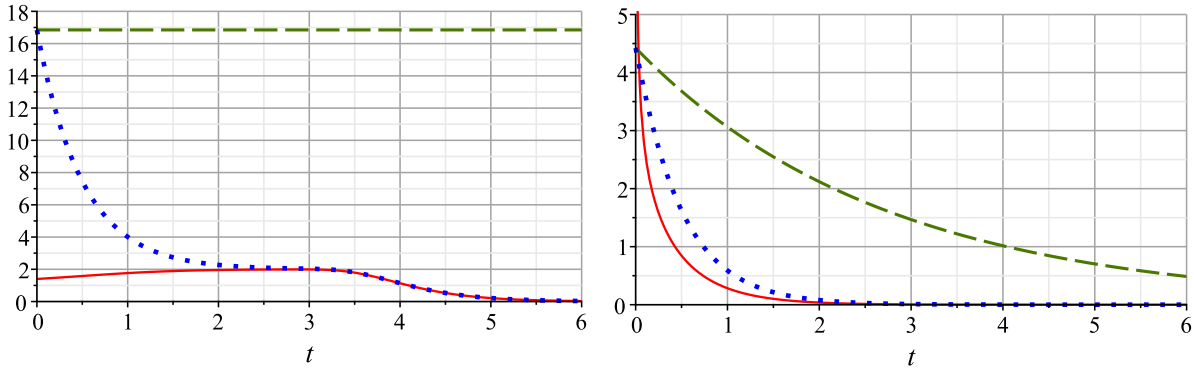


Figure 4.2: $W_\alpha(t)$ (solid red), $\tilde{W}_\alpha(t)$ (dotted blue), and $\tilde{W}_\alpha(0)e^{-\kappa t}$ (dashed green) with κ from Proposition 4.3 for $(\theta, \mathring{\mu}, \varsigma^2, \mathring{\sigma}^2, \bar{u}) = (1, 0, 1, 25, 200)$ (left) and for $(\theta, \mathring{\mu}, \varsigma^2, \mathring{\sigma}^2, \bar{u}) = (1, 0, 16, 1/10000, 1/4)$ (right), giving $(\alpha, \beta) = (25/4, 5000)$ and $(\alpha, \beta) = (4/625, 2)$, respectively.

Proposition 4.3. *For $\tilde{W}_\alpha(t)$ with $\alpha > 1$, inequality (4.8) holds with $C = 1$ and $\kappa = \theta e^{-\beta/2}$.*

Proof. To prove that inequality (4.8) holds for $\tilde{W}_\alpha(t)$ we can equivalently consider $\hat{W}_\alpha(t)$ from (4.13). Since $\alpha > 1$ we may drop the absolute value in $\hat{W}_\alpha(t)$. Then we can rewrite $\hat{W}_\alpha(t)$ due to $(\alpha - 1)e^{-2\theta t} = \zeta(t) - 1$:

$$\hat{W}_\alpha(t) = \zeta(t) - 1 + \frac{1}{\sqrt{2}} - \frac{\exp(-\eta(t))}{\sqrt{\zeta(t) + 1}}.$$

As in the proof of Proposition 4.1 we show $\hat{W}'_\alpha(t) + \kappa \hat{W}_\alpha(t) \leq 0$ to conclude our assertion. To keep notation brief, we introduce the variables

$$\tau := e^{-2\theta t} \in]0, 1], \qquad \omega := \zeta(t) + 1 > 2, \qquad \tilde{\kappa} := \frac{\kappa}{2\theta} = \frac{e^{-\beta/2}}{2} > 0.$$

This yields

$$\begin{aligned}
\zeta'(t) &= -2\theta(\alpha - 1)e^{-2\theta t} = -2\theta(\omega - 2), \\
\eta(t) &= \frac{\beta e^{-2\theta t}}{\zeta(t) + 1} = \frac{\beta\tau}{\omega}, \\
\eta'(t) &= -\frac{2\theta\beta e^{-2\theta t}}{\zeta(t) + 1} - \frac{\zeta'(t)\beta e^{-2\theta t}}{(\zeta(t) + 1)^2} = -\frac{2\theta\beta\tau}{\omega} + \frac{2\theta(\omega - 2)\beta\tau}{\omega^2} = -2\theta\left[\frac{\beta\tau}{\omega} - \frac{(\omega - 2)\beta\tau}{\omega^2}\right] \\
&= -2\theta \cdot \frac{2\beta\tau}{\omega^2}.
\end{aligned}$$

We have

$$\begin{aligned}
&\hat{W}'_\alpha(t) + \kappa \hat{W}_\alpha(t) \\
&= \zeta'(t) + \frac{\eta'(t)\exp(-\eta(t))}{\sqrt{\zeta(t) + 1}} + \frac{\zeta'(t)\exp(-\eta(t))}{2(\zeta(t) + 1)^{3/2}} + \kappa\left[\zeta(t) - 1 + \frac{1}{\sqrt{2}} - \frac{\exp(-\eta(t))}{\sqrt{\zeta(t) + 1}}\right] \\
&= -2\theta\left(\omega - 2 + \frac{e^{-\beta\tau/\omega}}{\sqrt{\omega}}\left[\frac{2\beta\tau}{\omega^2} + \frac{\omega - 2}{2\omega}\right]\right) + \kappa\left[\omega - 2 + \frac{1}{\sqrt{2}} - \frac{e^{-\beta\tau/\omega}}{\sqrt{\omega}}\right] \\
&= \underbrace{-2\theta\left(\omega - 2 + \frac{e^{-\beta\tau/\omega}}{\sqrt{\omega}}\left[\frac{2\beta\tau}{\omega^2} + \frac{\omega - 2}{2\omega}\right] - \frac{\kappa}{2\theta}\left[\omega - 2 + \frac{1}{\sqrt{2}} - \frac{e^{-\beta\tau/\omega}}{\sqrt{\omega}}\right]\right)}_{=: h_3}
\end{aligned}$$

and now need to show that $h_3 \geq 0$:

$$\begin{aligned}
h_3 &= \omega - 2 + \frac{e^{-\beta\tau/\omega}}{\sqrt{\omega}}\left[\frac{2\beta\tau}{\omega^2} + \frac{\omega-2}{2\omega}\right] - \frac{\kappa}{2\theta}\left[\omega - 2 + \frac{1}{\sqrt{2}} - \frac{e^{-\beta\tau/\omega}}{\sqrt{\omega}}\right] \\
&= \omega - 2 + \frac{e^{-\beta\tau/\omega}}{\sqrt{\omega}}\left[\frac{2\beta\tau}{\omega^2} + \frac{\omega-2}{2\omega}\right] - \tilde{\kappa}\left[\omega - 2 + \frac{1}{\sqrt{2}} - \frac{e^{-\beta\tau/\omega}}{\sqrt{\omega}}\right] \\
&= (\omega-2)(1-\tilde{\kappa}) + \frac{e^{-\beta\tau/\omega}}{\sqrt{\omega}}\left[\frac{2\beta\tau}{\omega^2} + \frac{\omega-2}{2\omega} + \tilde{\kappa}\right] - \frac{\tilde{\kappa}}{\sqrt{2}} \\
&= \underbrace{(\omega-2)}_{\geq 0}\ \underbrace{(1-2\tilde{\kappa})}_{=1-\exp(-\beta/2)\geq 0} + \frac{e^{-\beta\tau/\omega}}{\sqrt{\omega}}\left[\frac{2\beta\tau}{\omega^2} + \tilde{\kappa}\right] + \underbrace{\frac{e^{-\beta\tau/\omega}}{\sqrt{\omega}}\cdot\frac{\omega-2}{2\omega}}_{\geq 0} - \frac{\tilde{\kappa}}{\sqrt{2}} + (\omega-2)\tilde{\kappa} \\
&\geq \frac{e^{-\beta\tau/\omega}}{\sqrt{\omega}}\left[\frac{2\beta\tau}{\omega^2} + \tilde{\kappa}\right] + \tilde{\kappa}\left[\omega - 2 - \frac{1}{\sqrt{2}}\right] =: h_4.
\end{aligned}$$

If $\omega \geq 2 + 1/\sqrt{2}$ then the assertion follows. Hence, it remains to show that $h_4 \geq 0$ for $2 < \omega < 2 + 1/\sqrt{2}$. To this end, we have

$$\begin{aligned}
h_4 &= \frac{e^{-\beta\tau/\omega}}{\sqrt{\omega}}\left[\frac{2\beta\tau}{\omega^2} + \tilde{\kappa}\right] + \tilde{\kappa}\left[\omega - 2 - \frac{1}{\sqrt{2}}\right] \\
&= \tilde{\kappa}\left(\frac{e^{-\beta\tau/\omega}}{\sqrt{\omega}}\left[\frac{2\beta\tau}{\tilde{\kappa}\omega^2} + 1\right] + \omega - 2 - \frac{1}{\sqrt{2}}\right) \\
&= \underbrace{\tilde{\kappa}}_{>0}\ \underbrace{\left(\frac{e^{-\beta\tau/\omega}}{\sqrt{\omega}}\left[\frac{4\beta\tau e^{\beta/2}}{\omega^2} + 1\right] + \omega - 2 - \frac{1}{\sqrt{2}}\right)}_{=: h_5(\beta)}.
\end{aligned}$$

We assume $2 < \omega < 2 + 1/\sqrt{2}$. The idea now is to show $h_5(0) \geq 0$ as well as $h_5'(\beta) \geq 0$.[3] First, we have

$$h_5(0) = \frac{1}{\sqrt{\omega}} + \omega - 2 - \frac{1}{2} \geq 0.$$

Second, the derivative $h_5'(\beta)$ is given by

$$\begin{aligned}
h_5'(\beta) &= -\frac{\tau}{\omega}\left[\frac{4\beta\tau e^{\beta/2}}{\omega^2} + 1\right]\frac{e^{-\beta\tau/\omega}}{\sqrt{\omega}} + \frac{e^{-\beta\tau/\omega}}{\sqrt{\omega}}\left[\frac{4\tau e^{\beta/2}}{\omega^2} + \frac{2\beta\tau e^{\beta/2}}{\omega^2}\right] \\
&= \underbrace{-\frac{\tau}{\omega^{7/2}}e^{-\beta\tau/\omega}}_{\leq 0}\ \underbrace{\left(4\beta\tau e^{\beta/2} + \omega^2 - 4\omega e^{\beta/2} - 2\beta\omega e^{\beta/2}\right)}_{=: h_6}.
\end{aligned}$$

In the final step we show that $h_6 \leq 0$ for $2 < \omega < 2 + 1/\sqrt{2}$:

$$\begin{aligned}
h_6 &= \omega^2 + 4\beta\tau e^{\beta/2} - 2\beta\omega e^{\beta/2} - 4\omega\ \underbrace{e^{\beta/2}}_{\geq 1} \\
&\leq \omega^2 + 2\beta e^{\beta/2}\ \underbrace{(2\tau - \omega)}_{\leq 2-\omega \leq 0} - 4\omega \\
&\leq \omega(\omega - 4) < 0.
\end{aligned}$$

In conclusion, we have shown $\hat{W}_\alpha'(t) + \kappa\hat{W}_\alpha(t) \leq 0$ and thus the assertion. □

[3] We recall that τ and ω are independent of β.

To summarize, in all three cases we can apply Theorem 3.4 in order to conclude asymptotic stability of the MPC closed loop for the shortest possible horizon $N = 2$.

Remark 4.4. *(a) Figure 4.2 suggests that—at least in some cases—a much better decay rate can be obtained. However, this is irrelevant if $C = 1$ and the goal is to show asymptotic stability of the MPC closed loop for $N = 2$.*

(b) It is possible to employ $\tilde{W}_\alpha(t)$ for all three cases of α. For $\alpha = 1$, $\tilde{W}_1(t)$ coincides with $W_1(t)$. For $\alpha < 1$, the proof is structurally similar to the one of Proposition 4.3. Figure 4.2 (right) depicts exemplarily the case of $\alpha < 1$.

(c) As we will see more clearly in Chapter 5, in the case of Gaussian PDFs we can replace the PDF ρ in the stage costs ℓ and $\tilde{\ell}$, cf. (4.3) *and* (4.14)*, by its mean μ and its covariance matrix Σ. In the one-dimensional case Σ corresponds to the variance σ^2. Figure 4.3 depicts the terms penalizing the* state *in both stage costs—i.e., $\ell(\rho, \bar{u})$ and $\tilde{\ell}(\rho, \bar{u})$—in terms of (μ, Σ), where the desired PDF $\bar{\rho}$ is a Gaussian PDF with $(\bar{\mu}, \bar{\Sigma}) = (0, 1)$. The exact formulas can be obtained from Lemma 5.5 and the proof thereof.*

4.3 Numerical Simulations

For our numerical study, we consider the Ornstein–Uhlenbeck process on $Q := \mathbb{R} \times]0, 5[$. We use the explicit solution formula (4.2) and solve the optimal control problem using the program `OU-MPC`, cf. Section 7.2.[4]

We fix $\theta = 1$, $\mathring{\mu} = -3.5$, and $\bar{u} = 3.5$. For $\alpha = 1$, the remaining model parameters are $(\varsigma, \mathring{\sigma}) = (1/\sqrt{8}, 1/4)$. The cases $\alpha < 1$ and $\alpha > 1$ are modeled by $(\varsigma, \mathring{\sigma}) = (0.5, 0.1)$ and $(\varsigma, \mathring{\sigma}) = (0.1, 0.5)$, yielding $(\alpha, \beta) = (0.08, 196)$ and $(\alpha, \beta) = (50, 4900)$, respectively. In the MPC algorithm, we only look one time step into the future. The sampling time T_s is 0.1. We use the cost defined by (4.5) with $\gamma = 0.25$. The gradient of the cost was computed analytically.

Figure 4.4 shows the PDF $\rho(x, t)$ at various times, the desired equilibrium solution $\bar{\rho}$, and the corresponding controls for all three cases of α. The optimal control stays near $\bar{u} = 3.5$ until the PDF ρ is close enough to $\bar{\rho}$, when a higher control value helps reaching the target faster at reasonable cost. Table 4.1 displays the total cost $\sum_{n=0}^{49} \hat{J}_2(\rho_{\mathbf{u}_n}(n), \mathbf{u}_n)$ for the constant control $\mathbf{u}_n \equiv \bar{u}$ as well as for $\mathbf{u}_n = \mathbf{u}_n^*$, which denotes the optimal control sequence $\mathbf{u}^*$ calculated at the n-th MPC step. It shows the sub-optimality of $\bar{u}$.

	$\alpha = 1$	$\alpha < 1$	$\alpha > 1$
$\bar{u}$	32.43	21.57	136.15
$\mathbf{u}^*$	27.45 (-15.36%)	19.59 (-9.16%)	90.86 (-33.26%)

Table 4.1: Total cost for the constant control $\mathbf{u}_n \equiv \bar{u}$ and for $\mathbf{u}_n = \mathbf{u}_n^*$.

The cost $\hat{J}_2(\rho_{\mathbf{u}_n^*}(n), \mathbf{u}_n^*)$ in each MPC step n is illustrated in Figure 4.5 for all three cases of α and develops as predicted. We note that for $\alpha > 1$, even the optimal sequence

[4] It is also possible to numerically solve the Fokker–Planck equation directly, e.g., using the program `PDE-MPC`, cf. Section 7.1. A sufficiently fine discretization in space and time yields the same results.

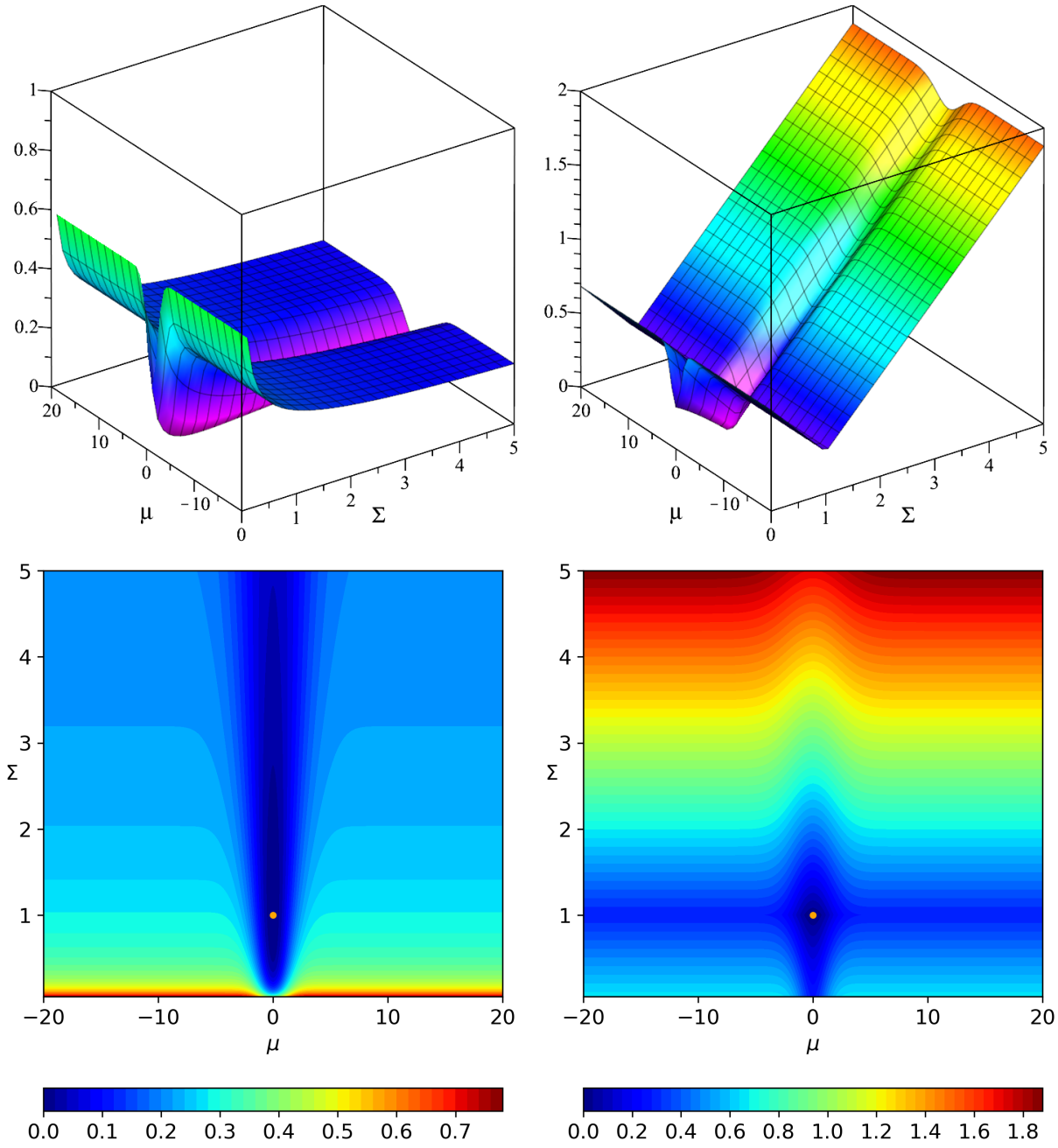


Figure 4.3: State costs $\ell(\rho, \bar{u}) = \ell(\rho, u) - \frac{\gamma}{2}|u - \bar{u}|^2$ (left) and $\tilde{\ell}(\rho, \bar{u}) = \tilde{\ell}(\rho, u) - \frac{\gamma}{2}|u - \bar{u}|^2$ (right) from (4.3) and (4.14), respectively, in the one-dimensional case expressed in terms of mean μ and covariance matrix Σ. The desired PDF $\bar{\rho}$ is a Gaussian PDF with $(\bar{\mu}, \bar{\Sigma}) = (0, 1)$. The orange dot in the bottom pictures at $(\mu, \Sigma) = (\bar{\mu}, \bar{\Sigma})$ marks the minimum.

leads to an increasing cost at the beginning. Since for the optimal control sequence $\mathbf{u}^* = (u^*(0), ..., u^*(N-1))$ we have $J_2(\mathring{\rho}, \mathbf{u}^*) = V_2(\mathring{\rho})$, cf. Theorem 3.4, this also shows that the optimal value function V_2 grows. In particular, V_2 cannot be a Lyapunov function for $N = 2$. Thus, based on this numerical evidence, Theorem 3.4 implies that exponential controllability with $C = 1$ cannot hold for the running cost (4.3). This further highlights the need of an equivalent stage cost in the proof since clearly, in the numerical simulations, the equilibrium solution $\bar{\rho}$ is asymptotically stable for the MPC closed loop, even for the

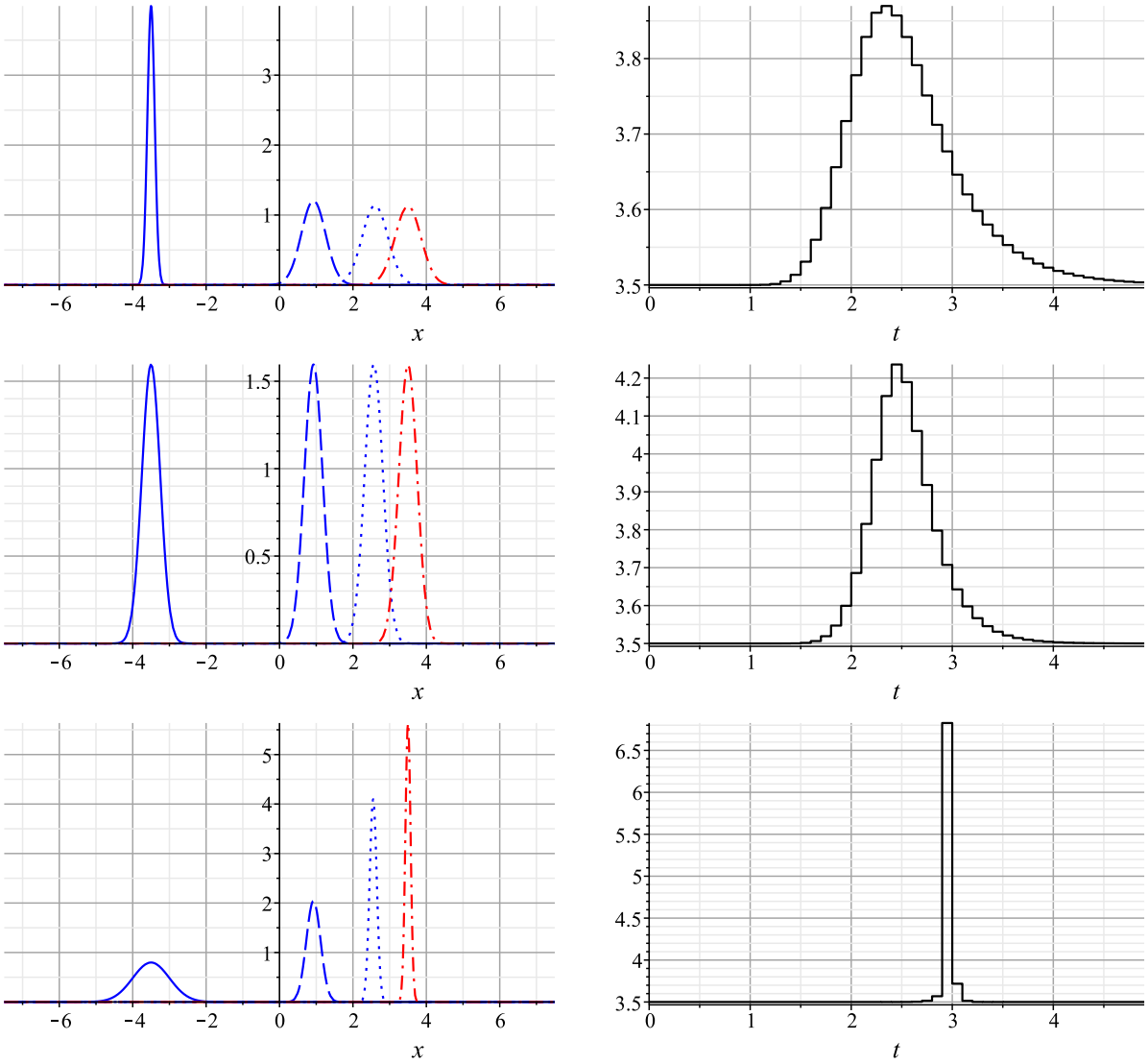


Figure 4.4: PDFs $\rho(x,0)$ (solid blue), $\rho(x,1)$ (dashed blue), $\rho(x,2)$ (dotted blue) and $\bar{\rho}(x)$ (dot-dashed red) on the left and the corresponding optimal MPC control $u^*(t)$ on the right for $\alpha < 1$, $\alpha = 1$, and $\alpha > 1$ (from top to bottom).

shortest possible optimization horizon. If we consider the stage cost $\tilde{\ell}$ and use an objective function $\hat{J}_2^{\tilde{\ell}}$ that is derived analogously to how $\hat{J}_2$ was derived from J_2, cf. (4.4) and (4.5), but with $\tilde{\ell}$ instead of ℓ, then we do see the exponential decay, see Figure 4.6. In conclusion, the numerical simulations coincide with our theoretical findings.

4.4 Conclusion

This chapter provides first insights and results regarding the stability of the MPC closed loop in the Fokker–Planck optimal control framework. For the Fokker–Planck equation associated with the Ornstein–Uhlenbeck process we can conclude asymptotic stability of the MPC closed loop even for the shortest possible horizon $N = 2$, if the control u does not depend on space. These findings coincide with our numerical simulations.

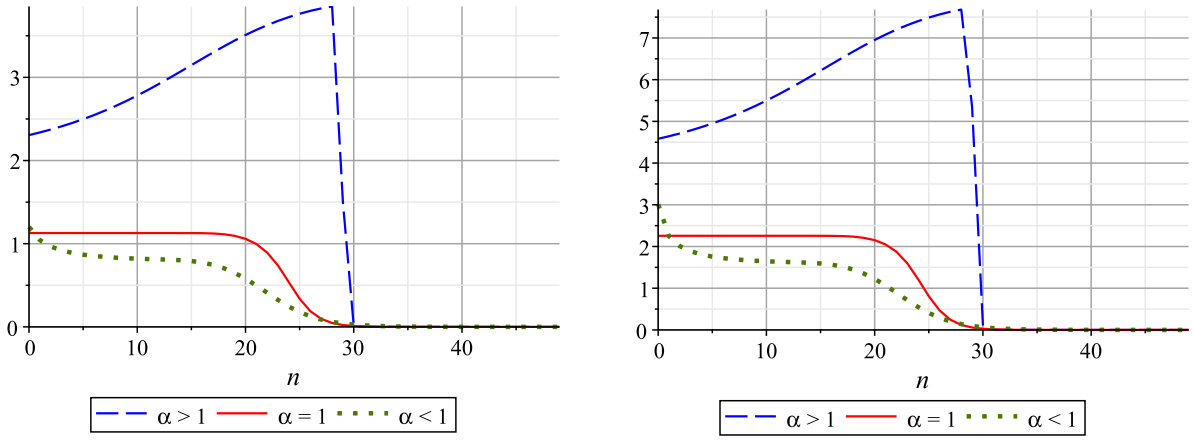


Figure 4.5: Objective functions $\hat{J}_2(\rho_{\mathbf{u}_n^*}(n), \mathbf{u}_n^*)$ from (4.5) (left) and $J_2(\rho_{\mathbf{u}_n^*}(n), \mathbf{u}_n^*)$ from (4.4) (right) for $\alpha = 1$ (solid red), $\alpha < 1$ (dotted green) and $\alpha > 1$ (dashed blue).

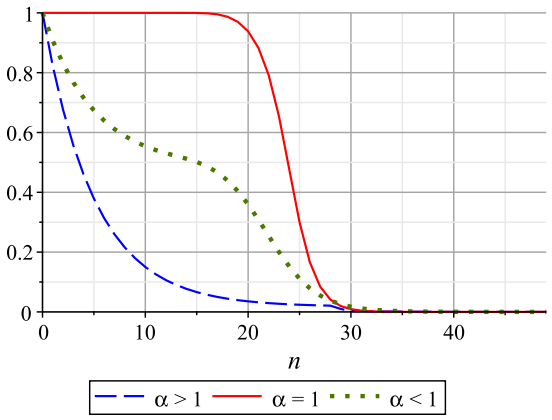


Figure 4.6: Objective function $\hat{J}_2^{\tilde{\ell}}(\rho_{\mathbf{u}_n^*}(n), \mathbf{u}_n^*)$ for $\alpha = 1$ (solid red), $\alpha < 1$ (dotted green) and $\alpha > 1$ (dashed blue), normalized to 1 at the beginning for better comparison.

Depending on the model parameters and the target PDF (more specifically, the relation between the initial and the target variance as well as the distance between the initial and the target mean), an adjustment of the stage cost ℓ is required in order to prove asymptotic stability. This is particularly apparent in Figure 4.5, where for $\alpha > 1$ the cost increases even for the optimal control u^*, providing strong numerical evidence that (4.8) with $C = 1$ cannot be concluded with ℓ. The workaround is to use an equivalent stage cost $\tilde{\ell}$, where the intuition is to remove problematic parts from the stage cost that cannot be controlled. Note that $\tilde{\ell}$ is only required in the proof. Since $\tilde{\ell}$ yields the same optimal control sequence as ℓ, one can still use the original stage cost ℓ in the numerical simulations.

In this chapter, the equivalent stage cost was obtained by adding terms related to the (evolution of the) variance, which cannot be influenced by a control that acts on the drift term and does not depend on space. This specific strategy does not work if the control does influence the variance, e.g., with a space-dependent control. However, unsurprisingly, much better tracking results are obtained with a space-dependent control, see Figure 2.1. Hence, in the subsequent chapter, we study a more general setting, where we consider a whole class of stochastic processes and (a class of) space-dependent controls in particular.

Stabilizing MPC – Linear Control

5

In this chapter we continue the stability analysis of Model Predictive Control schemes applied to the Fokker–Planck equation for tracking probability density functions in the stabilizing MPC case, cf. Section 3.2. The analysis is carried out for linear dynamics and Gaussian distributions, where, as in the previous chapter, the distance to the desired reference is measured in the L^2 norm. We present results for general such systems with and without control penalization. Refined results are given for the special case of the Ornstein–Uhlenbeck process—this time with a space-dependent control—, where we establish stability for the shortest possible optimization horizon $N = 2$.

As before, the results in this chapter are based on general MPC stability and performance guarantees from [45, 50] and [49, Ch. 6], which rely on appropriate controllability properties of the stage cost along the controlled dynamics, i.e., the L^2 distance to the reference PDF along the solutions of the Fokker–Planck PDE. More specifically, we rely on the exponential controllability property from Definition 3.3. However, we will see that even in the simplifying linear and Gaussian setting of this chapter, the assumptions from [45, 50] and [49, Ch. 6] are not always satisfied. Hence, for some of our results, we need to develop new arguments for proving stability of the MPC closed loop, cf. Section 5.3.2.

The remainder of this chapter is structured as follows. The precise problem formulation and assumptions are presented in Section 5.1. Section 5.2 collects important auxiliary results for the L^2 stage cost used in this chapter. The main results are presented in Section 5.3, which is divided into results for general linear stochastic control systems in Subsection 5.3.1 and results for the Ornstein–Uhlenbeck process in Subsection 5.3.2. The latter results demonstrate in which sense the general results can be further improved for a particular form of the stochastic dynamics. Section 5.4 concludes this chapter.

5.1 Problem Formulation and Assumptions

The problem setting in this section is a generalization of the one in Chapter 4: Instead of one specific stochastic process, we look at a whole class of stochastic processes. Again, we want to focus on Gaussian distributions. More precisely, we look at solutions of the Fokker–Planck equation (1.2) that have the form (1.6).

While it is entirely possible to work directly with the Fokker–Planck equation, see, for example, [85, 36], in general, it is hard to find conditions on the diffusion matrix (a_{ij}) and drift coefficients b as well as conditions on the structure of the control $u(x,t)$ that guarantee solutions of the form (1.6). Therefore, as a special case, let us consider linear

stochastic systems of the form

$$dX_t = AX_t dt + Bu(t)dt + DdW_t, \quad t \in]0, T[, \tag{5.1}$$

with an initial condition $\mathring{X} \in \mathbb{R}^d$ and where $A \in \mathbb{R}^{d\times d}$, $B \in \mathbb{R}^{d\times l}$, $D \in \mathbb{R}^{d\times m}$, and the control $u(t)$ is defined by

$$u(t) := -K(t)X_t + c(t) \tag{5.2}$$

for functions $K\colon \mathbb{R}_{\geq 0} \to \mathbb{R}^{l\times d}$ and $c\colon \mathbb{R}_{\geq 0} \to \mathbb{R}^l$. This results in

$$dX_t = (A - BK(t))\, X_t dt + Bc(t)dt + DdW_t, \quad t \in]0, T[, \tag{5.3}$$

i.e., a stochastic process (1.1) with constant diffusion $\tilde{a}(X_t, t) \equiv D$ and a linear drift term $b(X_t, t; u) = (A - BK(t))X_t + Bc(t)$, from which the coefficients for the associated Fokker–Planck equation (1.2) can be derived.

As before, for a matrix $A \in \mathbb{R}^{d\times d}$, we write $|A| := \det(A)$. If $\mathring{X} \sim \mathcal{N}(\mathring{\mu}, \mathring{\Sigma})$ with mean $\mathring{\mu} \in \mathbb{R}^d$ and covariance matrix $\mathring{\Sigma} \in \mathbb{R}^{d\times d} > 0$, then the corresponding initial PDF in (1.2b) is given by

$$\mathring{\rho}(x) := |2\pi\mathring{\Sigma}|^{-1/2} \exp\left(-\frac{1}{2}(x - \mathring{\mu})^\top \mathring{\Sigma}^{-1}(x - \mathring{\mu})\right). \tag{5.4}$$

Then, due to linearity of the process, the solution of the corresponding Fokker–Planck equation (1.2), $\rho(x, t)$, is also a Gaussian PDF of form (1.6), cf. [80, 22, 18]. The same holds if A, B, and D are time-dependent, cf. [84, Sect. 6.5]. In particular, for linear processes, the control structure (5.2) is the appropriate choice to preserve Gaussian density functions.

In the rest of this chapter, we consider linear stochastic systems of type (5.3) with corresponding initial PDF (5.4). While it is entirely possible to work in the PDE setting with a control that is linear in space, i.e., $u(x, t) = -K(t)x + c(t)$, we can alternatively characterize these processes via the following ODE system for the corresponding mean $\mu(t)$ and covariance matrix $\Sigma(t)$, see [18, p. 117]:

$$\begin{aligned} \dot{\mu}(t) &= (A - BK(t))\mu(t) + Bc(t), & \mu(0) &= \mathring{\mu}, \\ \dot{\Sigma}(t) &= (A - BK(t))\Sigma(t) + \Sigma(t)(A - BK(t))^\top + DD^\top, & \Sigma(0) &= \mathring{\Sigma}. \end{aligned} \tag{5.5}$$

Note that even though the control (5.2) enters through the drift term, cf. (5.1), since it is linear in space, both mean and covariance matrix are affected. Moreover, since $K(t)$ and $c(t)$ are to be optimized, we remind that the resulting OCP is *bilinear*.

Due to the fixed form of the control, (5.2), in the following, we will use the term "control" for both $u(x, t)$ and the pair of coefficients $(K(t), c(t))$, depending on the context. Likewise, our objective to steer the solution $\rho(x, t; u)$ to a Gaussian PDF

$$\bar{\rho}(x) := |2\pi\bar{\Sigma}|^{-1/2} \exp\left(-\frac{1}{2}(x - \bar{\mu})^\top \bar{\Sigma}^{-1}(x - \bar{\mu})\right) \tag{5.6}$$

and remain there is equivalent to steer the pair $(\mu(t), \Sigma(t))$ to $(\bar{\mu}, \bar{\Sigma})$ and maintain that state.

One particular process of type (5.5) is the already known (controlled) Ornstein–Uhlenbeck process, which we briefly reintroduce due to the new linear control structure.

Example 5.1 (Ornstein–Uhlenbeck)**.** *(a) For given parameters $\theta, \varsigma > 0$ and a control of type* (5.2)*, the controlled Ornstein–Uhlenbeck process* (1.7) *reads:*

$$dX_t = [-(\theta + K(t))\, X_t + c(t)]\, dt + \varsigma dW_t, \quad t \in]0, T[, \tag{5.7}$$

i.e., (5.3) *with $A = -\theta$, $B = 1$, and $D = \varsigma$. To keep the properties of the process, we require $\theta + K(t) > 0$ for all $t \geq 0$. We do not (need to) impose any constraints on $c(t)$.*

(b) An easy extension to the d-dimensional case is made by considering

$$\begin{aligned} A &= \operatorname{diag}(-\theta_1, \ldots, -\theta_d), \\ B &= I, \\ D &= \operatorname{diag}(\varsigma_1, \ldots, \varsigma_d), \\ K(t) &= \operatorname{diag}(k_1(t), \ldots, k_d(t)), \\ c(t) &= (c_1(t), \ldots, c_d(t)), \end{aligned} \tag{5.8}$$

where, analogously, we require that $k_i(t) > -\theta_i$ for all $t \geq 0$, $i = 1, ..., d$.

Let us assume that $\mathring{\rho}$ is a Gaussian PDF with mean $\mathring{\mu} \in \mathbb{R}^d$ and covariance matrix $\mathring{\Sigma} = \operatorname{diag}(\mathring{\sigma}_1^2, ..., \mathring{\sigma}_d^2)$ with $\mathring{\sigma}_i > 0$, $i = 1, ..., d$. Furthermore, let us view the control coefficients $(K(t), c(t))$ as parameters for the moment and assume that they are constant, i.e., $k_i(t) \equiv \bar{k}_i$ and $c_i(t) \equiv \bar{c}_i$, $i = 1, \ldots, d$. Then, analogously to the space-independent control case in Section 4.1, the ODE system (5.5) *can be solved analytically, with the mean given by*

$$\mu_i(t) = \frac{\bar{c}_i}{\theta_i + \bar{k}_i} + \left(\mathring{\mu}_i - \frac{\bar{c}_i}{\theta_i + \bar{k}_i} \right) e^{-(\theta_i + \bar{k}_i)t} \tag{5.9}$$

and covariance matrix

$$\Sigma(t) = \operatorname{diag}(\sigma_1^2(t), ..., \sigma_d^2(t)), \tag{5.10}$$

where

$$\sigma_i^2(t) := \frac{\varsigma_i^2}{2(\theta_i + \bar{k}_i)} + \left(\mathring{\sigma}_i^2 - \frac{\varsigma_i^2}{2(\theta_i + \bar{k}_i)} \right) e^{-2(\theta_i + \bar{k}_i)t}, \tag{5.11}$$

for $i = 1, ..., d$. We define $\bar{\mu} := (\bar{\mu}_1, ..., \bar{\mu}_d)$ and $\bar{\Sigma} := \operatorname{diag}(\bar{\sigma}_1^2, ..., \bar{\sigma}_d^2)$, where

$$\lim_{t \to \infty} \mu_i(t) = \frac{\bar{c}_i}{\theta_i + \bar{k}_i} =: \bar{\mu}_i \quad \text{and} \quad \lim_{t \to \infty} \sigma_i^2(t) = \frac{\varsigma_i^2}{2(\theta_i + \bar{k}_i)} =: \bar{\sigma}_i^2. \tag{5.12}$$

While in Example 5.1 it is easy to see that any desired state of type (5.6) can be reached by choosing appropriate functions $(K(t), c(t))$ and stabilized with constant $(\bar{K}, \bar{c})$, in general, this is not the case. To ensure the existence of controls $(K(t), c(t))$ such that at some given time $T > 0$, $\bar{\rho}$ is reached, it is necessary and sufficient to require (A, B) to be a controllable pair, see [22, Sects. II and III] or [18, Theorems 2.10.5 and 2.10.6]. After having reached $\bar{\rho}$, the aim is to stay there. In this chapter, we want to focus on stationary states that can be stabilized by applying static-state feedback, i.e., (5.2) with some constant $(\bar{K}, \bar{c})$. In general, not every desired PDF $\bar{\rho}$ can be stabilized in this manner. To this end, some conditions on $\bar{\Sigma}$ and the dynamics were derived in [22, Sect. III-B]. Overall, we end up with the following conditions, which we assume throughout the chapter:

Assumption 5.2. *(a) The pair (A, B) is controllable.*

(b) The covariance matrix of the desired Gaussian PDF $\bar{\rho}$, $\bar{\Sigma}$, is such that the equation

$$0 = A\bar{\Sigma} + \bar{\Sigma}A^\top + BX^\top + XB^\top + DD^\top \tag{5.13}$$

can be solved for X.

(c) $A - B\bar{K}$ is a Hurwitz matrix for $\bar{K} = -X^\top\bar{\Sigma}^{-1}$ and X the solution of (5.13).

(d) The equation

$$0 = (A - B\bar{K})\bar{\mu} + B\bar{c}$$

has a solution $(\bar{K}, \bar{c})$ with $\bar{K}$ as in (c).

As mentioned above, the first condition guarantees the existence of controls $(K(t), c(t))$ such that a given Gaussian PDF $\bar{\rho}$, characterized by the pair $(\bar{\mu}, \bar{\Sigma})$, can be reached. From (5.5) we see that Assumption 5.2(b) is a necessary condition such that $\bar{\Sigma}$ can be stabilized using a constant $\bar{K}$. On the other hand, if it holds, then it is possible to choose $\bar{K} = -X^\top\bar{\Sigma}^{-1}$, which satisfies the algebraic Lyapunov equation

$$(A - B\bar{K})\bar{\Sigma} + \bar{\Sigma}(A - B\bar{K})^\top = -DD^\top. \tag{5.14}$$

Hence, if, additionally, Assumption 5.2(c) holds, then $\bar{\Sigma}$ is an admissible stationary state covariance in the sense that it can be stabilized using a constant control $\bar{K}$. In order to stabilize a desired mean $\bar{\mu}$ as well, in addition to the previous assumptions, we require Assumption 5.2(d) to hold. This condition is sufficient due to (5.5) and the fact that $A - B\bar{K}$ is Hurwitz according to Assumption 5.2(c). For more details, see [22].

Remark 5.3. *(a) The solvability of (5.13) is equivalent to the rank condition*

$$\operatorname{rank}\begin{pmatrix} A\bar{\Sigma} + \bar{\Sigma}A^\top + DD^\top & B \\ B & 0 \end{pmatrix} = \operatorname{rank}\begin{pmatrix} 0 & B \\ B & 0 \end{pmatrix},$$

cf. [22] or [42, Prop. 1].

(b) Since $\bar{\Sigma}$ is positive definite, if the symmetric matrix $DD^\top$ is positive definite, too, then Assumption 5.2(c) is guaranteed. In the general case, in which $DD^\top$ is only positive semi-definite, however, this is not true, cf. Example 5.4. Yet, a sufficient condition for Assumption 5.2(c) to hold is that the range of B is a subset of the range of D, i.e., $\mathcal{R}(B) \subseteq \mathcal{R}(D)$, which one can verify without knowing $\bar{K}$, cf. [22].

(c) If one ignores the mean or considers the case where it is constant for all times, then one can drop Assumption 5.2(d).

Example 5.4. *Consider*

$$A := \begin{pmatrix} -\frac{13}{2} & -\frac{11}{4} \\ \frac{13}{4} & \frac{7}{8} \end{pmatrix},\ B := \begin{pmatrix} 0 & 3 \\ 1 & -3 \end{pmatrix},\ D := \begin{pmatrix} 1 \\ -\frac{1}{2} \end{pmatrix},\ \bar{\Sigma} = I,\ X := \begin{pmatrix} 3 & 2 \\ 2 & 1 \end{pmatrix},$$

for which (5.13) holds.[1] *However, the matrix $A - B\bar{K}$ with $\bar{K} = -X^\top\bar{\Sigma}^{-1}$ is not Hurwitz since one of the Eigenvalues of $A - B\bar{K}$ is zero.*

[1] Note also that D has full rank.

To summarize, we consider stochastic processes (5.3) with corresponding initial PDF (5.4). Our objective is to steer to and remain at a certain stationary PDF $\bar{\rho}$ from (5.6), which can be characterized by its mean $\bar{\mu}$ and covariance matrix $\bar{\Sigma}$. Therefore, we can equivalently study the dynamics (5.5). With Assumption 5.2 we ensure the feasibility of the problem.

In a next step, we want to solve this problem using Model Predictive Control, cf. Chapter 3. Formulated in the MPC setting, we want to solve ($\mathrm{OCP_N}$) subject to dynamics that are sampled from (5.5) and where the stage cost ℓ in the cost function $J_N((\mathring{\mu}, \mathring{\Sigma}), \mathbf{u})$ is yet to be defined. We consider stabilizing MPC, cf. Section 3.2, i.e., a positive definite ℓ with respect to a stationary Gaussian PDF (5.6) that is characterized by $(\bar{\mu}, \bar{\Sigma})$. As in Chapter 4, we are mainly interested in the stability of the MPC closed loop. Since the choice of the stage cost ℓ is crucial, we take a closer look at designing a suitable stage cost ℓ in the next section before we turn to the analysis of the MPC closed loop.

5.2 Design and Properties of the Stage Cost ℓ

In light of Section 3.2, the standard choice of using quadratic costs in the state and the control penalization as in (3.4) appears to be viable. As before, we use the L^2 norm for the term penalizing the state. Applying the L^2 norm to the control penalization term as well—which we did in Chapter 2—is a common choice in PDE-constrained optimization, cf. [95]. However, since here the control (5.2) acts on the whole domain $\Omega = \mathbb{R}^d$ and is linear in space, using, e.g., $\|u(t) - \bar{u}\|^2_{L^2(\mathbb{R}^d)}$ is not meaningful. Here, $\bar{u}$ is of form (5.2) and can be characterized by its coefficients $(\bar{K}, \bar{c})$ that satisfy Assumption 5.2. Therefore, we penalize the deviation of the control coefficients $(K(t), c(t))$ from $(\bar{K}, \bar{c})$, which results in

$$\ell(\rho, u) := \frac{1}{2} \|\rho - \bar{\rho}\|^2_{L^2(\mathbb{R}^d)} + \frac{\gamma}{2} \left\|BK - B\bar{K}\right\|^2_F + \frac{\gamma}{2} \|Bc - B\bar{c}\|_2^2 \tag{5.15}$$

for some weight $\gamma \geq 0$ and where $\|\cdot\|_F$ denotes the Frobenius norm. Using the Frobenius norm for $K \in \mathbb{R}^{l \times d}$ fits well with the Euclidian norm used for $c \in \mathbb{R}^l$. We will use the appearing B in (5.15) in the following. Yet, for the Ornstein–Uhlenbeck process presented in Example 5.1 it does not matter since $B = I$ in that case.

In our setting, $\rho = \rho(x, t; u)$ is a Gaussian PDF of form (1.6) with mean $\mu(t)$ and covariance matrix $\Sigma(t)$. If we turn our focus from the Fokker–Planck equation (1.2) to the associated dynamics (5.5), it is sensible to rewrite the term penalizing the state in (5.15) in terms of μ and Σ. In the following, we may drop the argument u in $\rho(x, t; u)$, $\Sigma(t; u)$, and $\mu(t; u)$ for better readability.

Lemma 5.5. *Let $\rho(x, t; u)$ and $\bar{\rho}(x)$ be given by* (1.6) *and* (5.6)*, respectively. Then for all $t \geq 0$:*

$$\begin{aligned} \|\rho(\cdot, t) - \bar{\rho}(\cdot)\|^2_{L^2(\mathbb{R}^d)} = 2^{-d}\pi^{-d/2} \Big[&|\Sigma(t)|^{-1/2} + |\bar{\Sigma}|^{-1/2} \\ &-2 \left| \frac{1}{2}(\Sigma(t) + \bar{\Sigma}) \right|^{-1/2} \exp\left(-\frac{1}{2} (\mu(t) - \bar{\mu})^\top (\Sigma(t) + \bar{\Sigma})^{-1} (\mu(t) - \bar{\mu}) \right) \Big]. \end{aligned} \tag{5.16}$$

We recall that $|A| = \det(A)$ for $A \in \mathbb{R}^{d \times d}$.

Proof. We split the L^2 norm into

$$\|\rho(t) - \bar{\rho}\|^2_{L^2(\mathbb{R}^d)} = \|\rho(t)\|^2_{L^2(\mathbb{R}^d)} + \|\bar{\rho}\|^2_{L^2(\mathbb{R}^d)} - 2\int_{\mathbb{R}^d} \rho(t)\bar{\rho}\, dx \tag{5.17}$$

and consider the three terms separately. Since only spatial integrals are involved while the time t remains fixed, in the following, we may drop the argument whenever it is clear from the context, i.e., instead of $\rho(x,t)$ we write $\rho(x)$.

We can apply standard results regarding integrals of Gaussians, cf. [75, Sect. 8.1.1], to

$$\rho(x)^2 = |2\pi\Sigma|^{-1} \exp\left(-(x-\mu)^\top \Sigma^{-1}(x-\mu)\right)$$

to get

$$\|\rho\|^2_{L^2(\mathbb{R}^d)} = |2\pi\Sigma|^{-1} \left|2\pi\left(\frac{1}{2}\Sigma\right)\right|^{1/2} = 2^{-d}\pi^{-d/2}|\Sigma|^{-1/2}.$$

Analogously, we have

$$\|\bar{\rho}\|^2_{L^2(\mathbb{R}^d)} = 2^{-d}\pi^{-d/2}|\bar{\Sigma}|^{-1/2}.$$

The last term in (5.17) is a bit more involved. First, we note that

$$\begin{aligned}\rho\bar{\rho} &= |2\pi\Sigma|^{-1/2}|2\pi\bar{\Sigma}|^{-1/2} \exp\left[-\frac{1}{2}(x-\mu)^\top\Sigma^{-1}(x-\mu) - \frac{1}{2}(x-\bar{\mu}^\top)\bar{\Sigma}^{-1}(x-\bar{\mu})\right] \\ &= |2\pi\Sigma|^{-1/2}|2\pi\bar{\Sigma}|^{-1/2} e^C \exp\left[-\frac{1}{2}(x-\mu_c)^\top\Sigma_c^{-1}(x-\mu_c)\right],\end{aligned} \tag{5.18}$$

where the second equality holds with

$$\begin{aligned}\Sigma_c^{-1} &:= \Sigma^{-1} + \bar{\Sigma}^{-1}, \\ \mu_c &:= (\Sigma^{-1} + \bar{\Sigma}^{-1})^{-1}(\Sigma^{-1}\mu + \bar{\Sigma}^{-1}\bar{\mu}), \\ C &:= \frac{1}{2}(\mu^\top\Sigma^{-1} + \bar{\mu}^\top\bar{\Sigma}^{-1})(\Sigma^{-1} + \bar{\Sigma}^{-1})^{-1}(\Sigma^{-1}\mu + \bar{\Sigma}^{-1}\bar{\mu}) - \frac{1}{2}(\mu^\top\Sigma^{-1}\mu + \bar{\mu}^\top\bar{\Sigma}^{-1}\bar{\mu}),\end{aligned}$$

cf. [75, Sect. 8.1.7]. Now we can apply the standard results from above to (5.18) in order to get

$$\begin{aligned}\int_{\mathbb{R}^d} \rho\bar{\rho}\, dx &= |2\pi\Sigma|^{-1/2}|2\pi\bar{\Sigma}|^{-1/2}|2\pi\Sigma_c|^{1/2} e^C \\ &= (2\pi)^{-d/2}|\Sigma|^{-1/2}|\bar{\Sigma}|^{-1/2}\left|\left(\Sigma^{-1} + \bar{\Sigma}^{-1}\right)^{-1}\right|^{1/2} e^C \\ &= (2\pi)^{-d/2}|\Sigma|^{-1/2}|\bar{\Sigma}|^{-1/2}\left|\Sigma^{-1} + \bar{\Sigma}^{-1}\right|^{-1/2} e^C \\ &= (2\pi)^{-d/2}\left|\Sigma\left(\Sigma^{-1} + \bar{\Sigma}^{-1}\right)\bar{\Sigma}\right|^{-1/2} e^C \\ &= (2\pi)^{-d/2}|\bar{\Sigma} + \Sigma|^{-1/2} e^C \\ &= 2^{-d}\pi^{-d/2}\left|\frac{1}{2}\left(\Sigma + \bar{\Sigma}\right)\right|^{-1/2} e^C.\end{aligned}$$

Therefore, it is left to show that

$$C = -\frac{1}{2}(\mu - \bar{\mu})^\top(\Sigma + \bar{\Sigma})^{-1}(\mu - \bar{\mu}).$$

To this end, we note that, since both Σ and $\bar{\Sigma}$ are symmetric positive definite and in particular invertible,

$$\bar{\Sigma}^{-1}\left(\Sigma^{-1}+\bar{\Sigma}^{-1}\right)^{-1}\Sigma^{-1} = \left(\Sigma\left(\Sigma^{-1}+\bar{\Sigma}^{-1}\right)\bar{\Sigma}\right)^{-1} = \left(\bar{\Sigma}+\Sigma\right)^{-1}. \tag{5.19}$$

Furthermore, we have that

$$\Sigma^{-1}\left(\Sigma^{-1}+\bar{\Sigma}^{-1}\right)^{-1}\Sigma^{-1} - \Sigma^{-1} = -\left(\Sigma+\bar{\Sigma}\right)^{-1}$$

due to

$$\begin{aligned}
&\Sigma^{-1}\left(\Sigma^{-1}+\bar{\Sigma}^{-1}\right)^{-1}\Sigma^{-1} - \Sigma^{-1} + \left(\Sigma+\bar{\Sigma}\right)^{-1} \\
\overset{(5.19)}{=}\ &\Sigma^{-1}\left(\Sigma^{-1}+\bar{\Sigma}^{-1}\right)^{-1}\Sigma^{-1} - \Sigma^{-1} + \bar{\Sigma}^{-1}\left(\Sigma^{-1}+\bar{\Sigma}^{-1}\right)^{-1}\Sigma^{-1} \\
=\ &\left[\left(\Sigma^{-1}+\bar{\Sigma}^{-1}\right)\left(\Sigma^{-1}+\bar{\Sigma}^{-1}\right)^{-1} - I\right]\Sigma^{-1} = 0.
\end{aligned}$$

These two results allow us to calculate C. We have

$$\begin{aligned}
C &= \frac{1}{2}(\mu^\top\Sigma^{-1}+\bar{\mu}^\top\bar{\Sigma}^{-1})(\Sigma^{-1}+\bar{\Sigma}^{-1})^{-1}(\Sigma^{-1}\mu+\bar{\Sigma}^{-1}\bar{\mu}) - \frac{1}{2}(\mu^\top\Sigma^{-1}\mu+\bar{\mu}^\top\bar{\Sigma}^{-1}\bar{\mu}) \\
&= \frac{1}{2}\mu^\top\Sigma^{-1}(\Sigma^{-1}+\bar{\Sigma}^{-1})^{-1}\Sigma^{-1}\mu + \frac{1}{2}\bar{\mu}^\top\bar{\Sigma}^{-1}(\Sigma^{-1}+\bar{\Sigma}^{-1})^{-1}\bar{\Sigma}^{-1}\bar{\mu} \\
&\quad - \frac{1}{2}(\mu^\top\Sigma^{-1}\mu+\bar{\mu}^\top\bar{\Sigma}^{-1}\bar{\mu}) + \frac{1}{2}\mu^\top\underbrace{\Sigma^{-1}(\Sigma^{-1}+\bar{\Sigma}^{-1})^{-1}\bar{\Sigma}^{-1}}_{=(\Sigma+\bar{\Sigma})^{-1}}\bar{\mu} \\
&\quad + \frac{1}{2}\bar{\mu}^\top\underbrace{\bar{\Sigma}^{-1}(\Sigma^{-1}+\bar{\Sigma}^{-1})^{-1}\Sigma^{-1}}_{=(\Sigma+\bar{\Sigma})^{-1}}\mu \\
&= \frac{1}{2}\mu^\top\underbrace{\left[\Sigma^{-1}\left(\Sigma^{-1}+\bar{\Sigma}^{-1}\right)^{-1}\Sigma^{-1} - \Sigma^{-1}\right]}_{=-(\Sigma+\bar{\Sigma})^{-1}}\mu \\
&\quad + \frac{1}{2}\bar{\mu}^\top\underbrace{\left[\bar{\Sigma}^{-1}\left(\Sigma^{-1}+\bar{\Sigma}^{-1}\right)^{-1}\bar{\Sigma}^{-1} - \bar{\Sigma}^{-1}\right]}_{=-(\Sigma+\bar{\Sigma})^{-1}}\bar{\mu} + \mu^\top(\Sigma+\bar{\Sigma})^{-1}\bar{\mu} \\
&= -\frac{1}{2}\mu^\top(\Sigma+\bar{\Sigma})^{-1}\mu - \frac{1}{2}\bar{\mu}^\top(\Sigma+\bar{\Sigma})^{-1}\bar{\mu} + \mu^\top(\Sigma+\bar{\Sigma})^{-1}\bar{\mu} \\
&= -\frac{1}{2}(\mu-\bar{\mu})^\top(\Sigma+\bar{\Sigma})^{-1}(\mu-\bar{\mu}),
\end{aligned}$$

which concludes the proof. □

In the course of this chapter, it will be useful to restrict the target PDF $\bar{\rho}$ of form (5.6) to

$$\bar{\rho}(x) = (2\pi)^{-d/2}\exp\left(-\frac{1}{2}x^\top x\right), \tag{5.20}$$

i.e., to set $\bar{\mu} = 0 \in \mathbb{R}^d$ and $\bar{\Sigma} = I \in \mathbb{R}^{d\times d}$. Then, due to Assumption 5.2(d), we have that $B\bar{c} = 0$, cf. (5.5). In this case, expressing the stage cost (5.15) in terms of the state (μ, Σ)

and control (K, c) using Lemma 5.5 leads to

$$\begin{aligned}\ell((\mu,\Sigma),(K,c)) = 2^{-d}\pi^{-d/2}\left[|\Sigma|^{-1/2} + 1 - 2\left|\frac{1}{2}(\Sigma+I)\right|^{-1/2}\exp\left(-\frac{1}{2}\mu^\top(\Sigma+I)^{-1}\mu\right)\right] \\ + \frac{\gamma}{2}\left\|BK - B\bar{K}\right\|_F^2 + \frac{\gamma}{2}\left\|Bc\right\|_2^2.\end{aligned} \tag{5.21}$$

This restriction on $\bar{\rho}$ in (5.20), i.e., assuming $(\bar{\mu}, \bar{\Sigma}) = (0, I)$, does not affect the generality of this chapter, as the following lemma shows.

Lemma 5.6. *Consider the optimal control problem* ($\mathrm{OCP_N}$) *subject to dynamics that are sampled from* (5.5)*. The problem of steering to a general target* $(\bar{\mu}, \bar{\Sigma})$ *can be transformed into a problem of steering to the target* $(0, I)$*. Using the stage cost* (5.15) *in the transformed problem yields the same cost as using the modified stage cost*

$$\begin{aligned}\ell_2(\rho,u) := \frac{1}{2}|\bar{\Sigma}|^{1/2}\left\|\rho-\bar{\rho}\right\|^2_{L^2(\mathbb{R}^d)} + \frac{\gamma}{2}\left\|\bar{\Sigma}^{-1/2}\left(BK - B\bar{K}\right)\bar{\Sigma}^{1/2}\right\|_F^2 \\ + \frac{\gamma}{2}\left\|\bar{\Sigma}^{-1/2}\left[(A-BK)\,\bar{\mu} + Bc\right]\right\|_2^2\end{aligned} \tag{5.22}$$

in the original problem.

The idea of the proof is to first consider (5.20) and work with the corresponding stage cost (5.15) and then encompass arbitrary target normal distributions $\bar{\rho}$—characterized by some mean $\bar{\mu}$ and some covariance matrix $\bar{\Sigma}$—by transforming the system dynamics and modifying the stage cost (5.15) in a suitable way. For example, it should make no difference in cost and in the control sequence whether we steer the expected value of a normal distribution from 10 to zero or from 11 to 1 in the one-dimensional case.

Proof. Starting from the SDE (5.3) and some arbitrary target normal distribution $\bar{\rho}$ characterized by its mean $\bar{\mu}$ and covariance matrix $\bar{\Sigma}$, we introduce a new random variable $Y_t := \bar{\Sigma}^{-1/2}(X_t - \bar{\mu})$. Then, due to linearity of the expected value, we get

$$\mu_Y(t) = \mathbb{E}[Y_t] = \mathbb{E}\left[\bar{\Sigma}^{-1/2}(X_t-\bar{\mu})\right] = \bar{\Sigma}^{-1/2}(\mathbb{E}[X_t]-\bar{\mu}) = \bar{\Sigma}^{-1/2}(\mu(t)-\bar{\mu})$$

and with

$$Y_t - \mu_Y(t) = \bar{\Sigma}^{-1/2}(X_t-\bar{\mu}) - \mu_Y(t) = \bar{\Sigma}^{-1/2}(X_t-\mu(t))$$

we get

$$\begin{aligned}\Sigma_Y(t) &= \mathbb{E}\left[(Y_t-\mu_Y(t))(Y_t-\mu_Y(t))^\top\right] \\ &= \mathbb{E}\left[\bar{\Sigma}^{-1/2}(X_t-\mu(t))(X_t-\mu(t))^\top\bar{\Sigma}^{-1/2}\right] \\ &= \bar{\Sigma}^{-1/2}\mathbb{E}\left[(X_t-\mu(t))(X_t-\mu(t))^\top\right]\bar{\Sigma}^{-1/2} = \bar{\Sigma}^{-1/2}\Sigma(t)\bar{\Sigma}^{-1/2}.\end{aligned}$$

Transforming (5.5) into the new variables (μ_Y, Σ_Y) yields

$$\begin{aligned}\dot{\mu}_Y(t) &= \bar{\Sigma}^{-1/2}(A-BK(t))\bar{\Sigma}^{1/2}\mu_Y(t) + \bar{\Sigma}^{-1/2}\left[(A-BK(t))\bar{\mu} + Bc(t)\right], \\ \mu_Y(0) &= \bar{\Sigma}^{-1/2}(\mathring{\mu}-\bar{\mu}), \\ \dot{\Sigma}_Y(t) &= \bar{\Sigma}^{-1/2}(A-BK(t))\bar{\Sigma}^{1/2}\Sigma_Y(t) + \Sigma_Y(t)\bar{\Sigma}^{1/2}(A-BK(t))^\top\bar{\Sigma}^{-1/2} \\ &\quad + \bar{\Sigma}^{-1/2}DD^\top\bar{\Sigma}^{-1/2}, \\ \Sigma_Y(0) &= \bar{\Sigma}^{-1/2}\mathring{\Sigma}\bar{\Sigma}^{-1/2}.\end{aligned} \tag{5.23}$$

Therefore, steering the system (5.23) to $(\bar{\mu}_Y, \bar{\Sigma}_Y) = (0, I)$ is equivalent to steering (5.5) to $(\bar{\mu}, \bar{\Sigma})$. In particular, if Assumption 5.2 holds for (5.5), then (5.23) can be steered towards $(0, I)$.

For the moment, let us assume that $(\bar{\mu}, \bar{\Sigma}) = (0, I)$. Then the stage cost (5.15) results in (5.21). The idea now is to compare the system (5.5) in the special case $(\bar{\mu}, \bar{\Sigma}) = (0, I)$ to (5.23) and adjust the stage cost accordingly. For instance, $\bar{\Sigma}^{-1/2}(A - BK(t))\bar{\Sigma}^{1/2}$ takes the role of $(A - BK(t))$.[2] Instead of $Bc(t)$, we have $\bar{\Sigma}^{-1/2}\left[(A - BK(t))\bar{\mu} + Bc(t)\right]$. Therefore, we adjust the stage cost (5.21) accordingly:

$$\|Bc\|_2^2 \rightsquigarrow \left\|\bar{\Sigma}^{-1/2}\left[(A - BK)\bar{\mu} + Bc\right]\right\|_2^2$$

and

$$\begin{aligned} &\left\|BK - B\bar{K}\right\|_F^2 = \left\|(A - BK) - (A - B\bar{K})\right\|_F^2 \\ \rightsquigarrow\ &\left\|\bar{\Sigma}^{-1/2}\left(A - BK\right)\bar{\Sigma}^{1/2} - \bar{\Sigma}^{-1/2}\left(A - B\bar{K}\right)\bar{\Sigma}^{1/2}\right\|_F^2 = \left\|\bar{\Sigma}^{-1/2}\left(BK - B\bar{K}\right)\bar{\Sigma}^{1/2}\right\|_F^2. \end{aligned}$$

The only term left to adjust is $\|\rho - \bar{\rho}\|_{L^2(\mathbb{R}^d)}^2$. Since $\Sigma(t) = \bar{\Sigma}^{1/2}\Sigma_Y(t)\bar{\Sigma}^{1/2}$ and $\Sigma(t) + \bar{\Sigma} = \bar{\Sigma}^{1/2}\left(\Sigma_Y(t) + I\right)\bar{\Sigma}^{1/2}$, we have

$$\begin{aligned} |\Sigma(t)|^{-1/2} &= |\bar{\Sigma}^{1/2}\Sigma_Y(t)\bar{\Sigma}^{1/2}|^{-1/2} = |\bar{\Sigma}|^{-1/2}|\Sigma_Y(t)|^{-1/2}, \\ \left|\frac{1}{2}(\Sigma(t) + \bar{\Sigma})\right|^{-1/2} &= \left|\frac{1}{2}(\bar{\Sigma}^{1/2}\left(\Sigma_Y(t) + I\right)\bar{\Sigma}^{1/2})\right|^{-1/2} = |\bar{\Sigma}|^{-1/2}\left|\frac{1}{2}(\Sigma_Y(t) + I)\right|^{-1/2}. \end{aligned}$$

Furthermore, since $\bar{\mu} = 0$ and therefore $\mu_Y(t) = \bar{\Sigma}^{-1/2}(\mu(t) - \bar{\mu}) = \bar{\Sigma}^{-1/2}\mu(t)$, we have

$$\begin{aligned} &|\Sigma(t)|^{-1/2} + |\bar{\Sigma}|^{-1/2} - 2\left|\frac{1}{2}(\Sigma(t) + \bar{\Sigma})\right|^{-1/2} \exp\left(-\frac{1}{2}\mu(t)^\top(\Sigma(t) + \bar{\Sigma})^{-1}\mu(t)\right) \\ &= |\bar{\Sigma}|^{-1/2}\left[|\Sigma_Y(t)|^{-1/2} + 1 - 2\left|\frac{1}{2}(\Sigma_Y(t) + I)\right|^{-1/2} \exp\left(-\frac{1}{2}\mu_Y(t)^\top(\Sigma_Y(t) + I)^{-1}\mu_Y(t)\right)\right]. \end{aligned}$$

This, together with (5.21), explains the last necessary adjustment, namely the factor $|\bar{\Sigma}|^{1/2}$ in front of the term penalizing the state in (5.22). □

In the special case of $\mu(t) \equiv \bar{\mu}$, i.e., if the state has reached the target mean and stays at that target, the restriction to $\bar{\Sigma} = I$ gives rise to the following result.

Lemma 5.7. *Let $\mu(t) \equiv \bar{\mu}$ and $\bar{\Sigma} = I$. Define $\Phi(t) := \operatorname{diag}(\phi_1(t), \dots, \phi_d(t))$, where $\phi_i(t)$, $i = 1, \dots, d$, are the Eigenvalues of $\Sigma(t)$. Furthermore, we define*

$$g(\Phi) := 1 + |\Phi|^{-1/2} - 2\left|\frac{1}{2}(\Phi + I)\right|^{-1/2} = 1 + \left(\prod_{i=1}^d \phi_i\right)^{-1/2} - 2^{1+d/2}\prod_{i=1}^d(\phi_i + 1)^{-1/2}. \quad (5.24)$$

Then

$$\|\rho(\cdot, t) - \bar{\rho}(\cdot)\|_{L^2(\mathbb{R}^d)}^2 = 2^{-d}\pi^{-d/2}g(\Phi(t)).$$

[2] To see this in the equation for $\dot{\Sigma}_Y(t)$, it is helpful to use (5.14), which holds due to Assumption 5.2(b).

Proof. Since $\bar{\Sigma} = I$ and $\mu(t) \equiv \bar{\mu}$, the state cost (5.16) becomes

$$\|\rho(\cdot,t) - \bar{\rho}(\cdot)\|^2_{L^2(\mathbb{R}^d)} = 2^{-d}\pi^{-d/2}\left[|\Sigma(t)|^{-1/2} + 1 - 2\left|\frac{1}{2}(\Sigma(t)+I)\right|^{-1/2}\right].$$

If $\phi_1(t), \dots, \phi_d(t)$ are the Eigenvalues of $\Sigma(t)$, then $\phi_i(t)+1$, $i = 1, \dots, d$, are the Eigenvalues of $\Sigma(t)+I$. Since $|\Sigma(t)| = |\Phi(t)|$ and $|\Sigma(t)+I| = |\Phi(t)+I|$, the assertion follows. □

5.3 Minimal Stabilizing Horizon Estimates

In this section, we want to study the behavior of the MPC closed loop. More precisely, we are interested in estimating minimal horizon lengths N such that our desired equilibrium $\bar{\rho}$, respectively $(\bar{\mu}, \bar{\Sigma})$, is asymptotically stable for the MPC closed loop.

Whether we consider the Fokker–Planck equation (1.2) with state $\rho(x,t)$ or, equivalently, the dynamics (5.5) with state $(\mu(t), \Sigma(t))$, they are always sampled in order to obtain the discrete-time system described in Section 3.1. That is, if $(\mu(t), \Sigma(t))$ is the solution trajectory of (5.5), then we denote by $\Sigma(n)$ the evaluation of $\Sigma(t)$ at time $t = t_n := t_0 + nT_s$, where $T_s > 0$ is the sampling rate and $n \in \mathbb{N}_0$. Similarly, we will write $\Sigma(k)$, where the difference between k and n is the same as in the MPC scheme in Section 3.1: The "global" time will be denoted by n, while k will indicate the "local" time, i.e., the time in the open-loop optimal control problem ($\mathrm{OCP_N}$) that needs to be solved in every MPC step. We will use the same notation for $\mu(n)$ and $\mu(k)$.

In order to prove asymptotic stability, we can use the exponential controllability property, cf. Theorem 3.4. A suitable stage cost ℓ is given by (5.15) or (5.22). In both cases, the state ρ is penalized in the L^2 norm, which, as already mentioned before, is well suited for PDE-constrained optimization. However, expressing the stage cost (5.15) in terms of the state $(\mu(t), \Sigma(t))$ instead of $\rho(x,t)$ leads to rather uncommon expressions, cf. Lemma 5.5. Yet, we strive to show that MPC does cope with these types of cost in this setting.

To this end, in Subsection 5.3.1, we present results for general stochastic processes (5.3) with $\mathring{X} \sim \mathcal{N}(\mathring{\mu}, \mathring{\Sigma})$, i.e., general dynamics of type (5.5). Then, in Subsection 5.3.2, we try to improve these results for a special case, namely the Ornstein–Uhlenbeck process that was introduced in Example 5.1.

5.3.1 General Dynamics of Type (5.3)

In this section, we consider general dynamics given by (5.1) with control (5.2), leading to the controlled linear dynamics (5.3) and the equivalent dynamics (5.5) for the Fokker–Planck equation (1.2). We start with the most simple case, in which there are no state constraints, no control constraints, and no control costs.

Theorem 5.8. *Consider the system* (5.5) *associated to a linear stochastic process defined by* (5.3) *with a Gaussian initial condition* (5.4) *and a desired PDF* $\bar{\rho}$ *given by* (5.6)*. Let the stage cost be given by* $\ell(\rho) := \frac{1}{2}\|\rho - \bar{\rho}\|^2_{L^2(\mathbb{R}^d)}$*, which corresponds to* (5.15) *with* $\gamma = 0$*. Then the equilibrium* $\bar{\rho}$ *is globally asymptotically stable for the MPC closed loop for each optimization horizon* $N \geq 2$*.*

Proof. In absence of state or control constraints, it is obvious that any system of type (5.5) that satisfies Assumption 5.2(a) can reach any desired state $\bar{\rho}$, which is characterized by

some mean $\bar{\mu}$ and some covariance matrix $\bar{\Sigma}$, in an arbitrarily short time $\tilde{T}$. In particular, in the continuous-time setting, one can choose a control coefficient $\tilde{K}(t)$ such that the desired covariance $\bar{\Sigma}$ is reached in $\tilde{T}/2$ time units. At that point in time, we switch to $\bar{K}$ and use an appropriate control coefficient $\tilde{c}(t)$ to arrive at the desired mean $\bar{\mu}$.

In the sampled system, in order to arrive at the desired state within one MPC time step, the control $(\tilde{K}(t), \tilde{c}(t))$ from the continuous time needs to be discretized adequately, i.e., every element in the open-loop control sequence $(K(k), c(k))_{k=0,\dots,N-1}$ of the first MPC time step may be a time-dependent function on $[t_k, t_{k+1}[$. In particular, the first element of that sequence, $(K(0), c(0))$, may be a time-dependent function on $[t_0, t_1[$. Having reached the desired state $\bar{\rho}$ in the first MPC time step, we then switch the control to $(\bar{K}, \bar{c})$, thus staying at $\bar{\rho}$ due to Assumptions 5.2(b)-(d) and invoking zero cost from then on. □

Remark 5.9. *While non-constant coefficients $(K(0), c(0))$ are no issue in theory, in practice the discretization of the control sequence $u(k)$ is often coupled with the discretization of the dynamics, leading to control sequences that are constant in every MPC time step. If the system cannot be steered towards the desired state within one discrete step using constant $(K(0), c(0))$, then one should adjust the discretization of the control in time.*

Now we turn to the more interesting case where $\gamma > 0$ and/or control constraints are present. In this case, in general, we cannot guarantee that the target $\bar{\rho}$ is asymptotically stable for $N = 2$. Yet, we can recover the asymptotic stability by choosing $N \geq 2$ sufficiently large, cf. Theorem 5.11. In the proof thereof, we will need the following result.

Lemma 5.10. *Consider* (5.5) *for $K(t) \equiv \bar{K}$. Then*

$$\left\|\Sigma(t) - \bar{\Sigma}\right\|_F \leq Ce^{-\kappa t} \left\|\Sigma(0) - \bar{\Sigma}\right\|_F \tag{5.25}$$

for some constants $C, \kappa > 0$.

Proof. Due to Assumption 5.2, $A - B\bar{K}$ is a Hurwitz matrix and (5.14) holds. Therefore,

$$\begin{aligned}\dot{\Sigma}(t) &= (A - B\bar{K})\Sigma(t) + \Sigma(t)(A - B\bar{K})^\top + DD^\top \\ &\overset{(5.14)}{=} (A - B\bar{K})(\Sigma(t) - \bar{\Sigma}) + (\Sigma(t) - \bar{\Sigma})(A - B\bar{K})^\top.\end{aligned}$$

Defining $M := A - B\bar{K}$ and $S(t) := \Sigma(t) - \bar{\Sigma}$, we can rewrite the above equation to

$$\dot{S}(t) = MS(t) + S(t)M^\top.$$

Then we vectorize this equation by going through the matrix $S(t)$ row by row, i.e., for

$$S(t) = \begin{pmatrix} s_{11}(t) & \dots & s_{1d}(t) \\ \vdots & & \vdots \\ s_{d1}(t) & \dots & s_{dd}(t) \end{pmatrix},$$

we define yet another variable

$$s_v(t) := (s_{11}(t), \dots, s_{1d}(t), s_{21}(t), \dots, s_{2d}(t), \dots, s_{d1}(t), \dots, s_{dd}(t))$$

and arrive at

$$\dot{s}_v(t) = \tilde{A}s_v(t), \tag{5.26}$$

with $\tilde{A} \in \mathbb{R}^{d^2 \times d^2}$ defined by

$$\tilde{A} := \begin{pmatrix} m_{11}(t)I & \dots & m_{1d}(t)I \\ \vdots & & \vdots \\ m_{d1}(t)I & \dots & m_{dd}(t)I \end{pmatrix} + \begin{pmatrix} M & & \\ & \ddots & \\ & & M \end{pmatrix}.$$

Let $\epsilon(M)$ be the set of all Eigenvalues of M. Then one can calculate that the set of all Eigenvalues of $\tilde{A}$, $\epsilon(\tilde{A})$, consists of all possible sums $\phi_1^m + \phi_2^m$, where $\phi_1^m, \phi_2^m \in \epsilon(M)$. In particular, $\epsilon(\tilde{A}) \subset \mathbb{C}_-$ since $\epsilon(M) \subset \mathbb{C}_-$. For the linear system (5.26) this implies exponential stability, and thus

$$\|s_v(t)\|_2 \leq Ce^{-\kappa t} \|s_v(0)\|_2$$

for some constants $C, \kappa > 0$. Since $\|s_v(t)\|_2 = \|S(t)\|_F = \left\|\Sigma(t) - \bar{\Sigma}\right\|_F$, we arrive at (5.25). □

Theorem 5.11. *Consider the dynamic system* (5.5) *associated to a linear stochastic process defined by* (5.3) *with a Gaussian initial condition* (5.4) *and a desired PDF* $\bar{\rho}$ *given by* (5.20). *Let the stage cost be given by* (5.15) *with* $\gamma \geq 0$. *Then there exists some* $\bar{N} \geq 2$ *such that the equilibrium* $\bar{\rho}$ *is asymptotically stable for the MPC closed loop for each optimization horizon* $N \geq \bar{N}$ *on recursively feasible sets that contain a neighborhood of* $\bar{\rho}$. *These sets are characterized in Remark 5.12 below.*

Proof. We want to prove exponential controllability of the system (5.5) with respect to the stage cost defined by (5.15), cf. Definition 3.3, at least locally. Then our assertion follows from Theorem 3.5.

Having Assumption 5.2 in mind, a natural control candidate to prove exponential controllability is $(\bar{K}, \bar{c})$. In this case, our stage cost reduces to $\frac{1}{2}\|\rho - \bar{\rho}\|^2_{L^2(\mathbb{R}^d)}$, i.e., the term penalizing the state. We will use the control candidate $(\bar{K}, \bar{c})$ throughout the proof. To prove local exponential controllability, we will show that

$$\|\rho(t) - \bar{\rho}\|^2_{L^2(\mathbb{R}^d)} \leq Ce^{-\kappa t} \|\rho(0) - \bar{\rho}\|^2_{L^2(\mathbb{R}^d)} \tag{5.27}$$

holds in continuous time for some $C, \kappa > 0$ and for initial PDFs $\rho(0)$ close to $\bar{\rho}$. Then with $\delta := e^{-\kappa T_s}$ we arrive at (3.6). With

$$V(\mu, \Sigma) := |\Sigma|^{-1/2} + 1 - 2\left|\frac{1}{2}(\Sigma + I)\right|^{-1/2} \exp\left(-\frac{1}{2}\mu^\top(\Sigma + I)^{-1}\mu\right), \tag{5.28}$$

due to (5.16), proving (5.27) is equivalent to showing

$$V(\mu(t), \Sigma(t)) \leq Ce^{-\kappa t} V(\mu(0), \Sigma(0)). \tag{5.29}$$

To this end, we take a closer look at the control $(\bar{K}, \bar{c})$. Since $A - B\bar{K}$ is a Hurwitz matrix and $\bar{\mu} = 0$, $B\bar{c}$ equals zero, cf. Assumption 5.2(d). Therefore, it is easy to see from the dynamics (5.5) that there exist some constants $C_1, \kappa_1 > 0$ such that

$$\|\mu(t)\|_2 \leq C_1 e^{-\kappa_1 t} \|\mu(0)\|_2 . \tag{5.30}$$

Furthermore, let $\phi_i(t)$, $i = 1, \dots, d$, be the Eigenvalues of $\Sigma(t)$ that we collect in the vector $\phi := (\phi_1, \dots, \phi_d)$ as well as in the matrix $\Phi := \operatorname{diag}(\phi_1, \dots, \phi_d)$. Since the initial

condition is Gaussian, the positivity of $\phi_i(t)$ is preserved, i.e., $\phi_i(t) > 0$ for all $i = 1, ..., d$ and all $t \geq 0$. Moreover, due to Lemma 5.10 we have $\|\Sigma(t) - I\|_F \leq C_2 e^{-\kappa_2 t} \|\Sigma(0) - I\|_F$ for some $C_2, \kappa_2 > 0$. This can be expressed in terms of the Eigenvalues: Denoting by $\vec{1}$ the d-dimensional vector of ones, we have $\|\Sigma(t) - I\|_F = \|\Phi(t) - I\|_F = \|\phi(t) - \vec{1}\|_2$, where the first equation holds because $\Sigma(t) - I$ is a real and symmetric and therefore normal matrix[3] and the Eigenvalues of $\Sigma(t) - I$ coincide with those of $\Phi(t) - I$, and the second equality holds since Φ is diagonal. Consequently,

$$\|\phi(t) - \vec{1}\|_2 \leq C_2 e^{-\kappa_2 t} \|\phi(0) - \vec{1}\|_2. \tag{5.31}$$

In the following, we want to use (5.30) and (5.31) to deduce (5.29).

Since C_2 is independent of the initial value $\phi(0)$, by limiting $\phi(0)$ to a (small enough) neighborhood of the target $\vec{1}$ we can bound $\sup_{t\geq 0}\|\phi(t) - \vec{1}\|_2$ to an arbitrarily small positive number. The analogous is true for $\|\mu(t)\|_2$. Thus, we denote by $B_r(x)$ a ball of radius $r > 0$ around $x \in \mathbb{R}^d$. Then from (5.30) and (5.31) we deduce that, for a given $\varepsilon \in]0, 1[$, there exist $r_\mu, r_\phi \in]0, \varepsilon[$ such that for any $(\mu(0), \phi(0)) \in \mathcal{B}_{r_\mu}(0) \times \mathcal{B}_{r_\phi}(\vec{1})$ we have $-\varepsilon \leq \phi_i(t) - 1 \leq \varepsilon$ and $-\varepsilon \leq \mu_i(t) \leq \varepsilon$ for all $t \geq 0$ and all $i = 1, ..., d$.

If $\phi_i(t)$, $i = 1, \ldots, d$, are the Eigenvalues of $\Sigma(t)$, then $\phi_i(t) + 1$ are the Eigenvalues of $\Sigma(t)+I$ and $(\phi_i(t)+1)^{-1}$ are the Eigenvalues of $(\Sigma(t)+I)^{-1}$. Since $0 < 1-\varepsilon \leq \phi_i(t) \leq 1+\varepsilon$, we have

$$1 > \frac{1}{\phi_i(t) + 1} \geq \frac{1}{2 + \varepsilon}.$$

Then we can bound the exponential term of V in (5.28):

$$\frac{1}{2+\varepsilon} \|\mu(t)\|_2^2 \leq \mu(t)^\top (\Sigma(t) + I)^{-1} \mu(t) \leq \|\mu(t)\|_2^2 .$$

Therefore,

$$\begin{aligned} 1 - \exp\left(-\frac{1}{2(2+\varepsilon)} \|\mu(t)\|_2^2\right) &\leq 1 - \exp\left(-\frac{1}{2}\mu(t)^\top (\Sigma(t) + I)^{-1} \mu(t)\right) \\ &\leq 1 - \exp\left(-\frac{1}{2} \|\mu(t)\|_2^2\right). \end{aligned}$$

Since

$$V(\mu, \Sigma) = |\Sigma|^{-1/2} + 1 - 2\left|\frac{1}{2}(\Sigma + I)\right|^{-1/2} + 2\left|\frac{1}{2}(\Sigma + I)\right|^{-1/2} \left[1 - \exp\left(-\frac{1}{2}\mu^\top (\Sigma + I)^{-1}\mu\right)\right],$$

we can bound $V(\mu(t), \Sigma(t))$:

$$V_l(\mu(t), \Sigma(t)) \leq V(\mu(t), \Sigma(t)) \leq V_u(\mu(t), \Sigma(t)), \tag{5.32}$$

where

$$V_l(\mu, \Sigma) := |\Sigma|^{-1/2} + 1 - 2\left|\frac{1}{2}(\Sigma + I)\right|^{-1/2} + 2\left|\frac{1}{2}(\Sigma + I)\right|^{-1/2} \left[1 - \exp\left(-\frac{1}{2(2+\varepsilon)} \|\mu\|_2^2\right)\right],$$

$$V_u(\mu, \Sigma) := |\Sigma|^{-1/2} + 1 - 2\left|\frac{1}{2}(\Sigma + I)\right|^{-1/2} + 2\left|\frac{1}{2}(\Sigma + I)\right|^{-1/2} \left[1 - \exp\left(-\frac{1}{2} \|\mu\|_2^2\right)\right].$$

[3] A normal matrix A is unitarily diagonalizable, i.e., has a factorization $A = U\Lambda U^T$, where $U^\top U = I$. Λ is a diagonal matrix consisting of the Eigenvalues of A. Then $\|A\|_F^2 = \mathrm{tr}(A^\top A) = \mathrm{tr}(U\Lambda^\top U^\top U\Lambda U^\top) = \mathrm{tr}(U\Lambda^\top \Lambda U^\top) = \mathrm{tr}(\Lambda^\top \Lambda) = \|\Lambda\|_F^2$. Thus, the Frobenius norm of normal matrices only depends on their Eigenvalues.

Note that $V_l(\mu, \Sigma) \geq 0$. With $\Phi = \operatorname{diag}(\phi_1, \ldots, \phi_d)$ we have $V_l(\mu, \Sigma) = V_l(\mu, \Phi)$ and $V_u(\mu, \Sigma) = V_u(\mu, \Phi)$. Moreover, since

$$|\Sigma| = |\Phi| = \prod_{i=1}^{d} \phi_i \quad \text{and} \quad \left|\frac{1}{2}(\Sigma + I)\right| = \left|\frac{1}{2}(\Phi + I)\right| = \prod_{i=1}^{d} \frac{\phi_i + 1}{2},$$

we can view the functions V_l and V_u as functions of the vector $\phi = (\phi_1, \ldots, \phi_d)$ instead of the matrix Φ and calculate for all $j = 1, \ldots, d$:

$$\partial_{\phi_j} V_l(\mu, \phi) = \frac{1}{2}\left[\left(\prod_{i=1}^{d} \frac{\phi_i + 1}{2}\right)^{-1/2} \left(\frac{\phi_j + 1}{2}\right)^{-1} \exp\left(-\frac{1}{2(2+\varepsilon)} \|\mu\|_2^2\right) - \left(\prod_{i=1}^{d} \phi_i\right)^{-1/2} \phi_j^{-1}\right],$$

$$\partial_{\mu_j} V_l(\mu, \phi) = \left(\prod_{i=1}^{d} \frac{\phi_i + 1}{2}\right)^{-1/2} \frac{1}{2+\varepsilon} \mu_j \exp\left(-\frac{1}{2(2+\varepsilon)} \|\mu\|_2^2\right).$$

This yields

$$V_l(0, \vec{1}) = 0, \quad \partial_{\phi_j} V_l(0, \vec{1}) = 0, \quad \partial_{\mu_j} V_l(0, \vec{1}) = 0$$

and, analogously,

$$V_u(0, \vec{1}) = 0, \quad \partial_{\phi_j} V_u(0, \vec{1}) = 0, \quad \partial_{\mu_j} V_u(0, \vec{1}) = 0.$$

As a consequence, no constant or linear terms appear in the Taylor expansion of either $V_l(\mu, \phi)$ or $V_u(\mu, \phi)$ around $(0, \vec{1})$. Moreover, one can easily verify that the respective Hessian matrices at $(0, \vec{1})$, i.e., $\nabla^2 V_l(0, \vec{1})$ and $\nabla^2 V_u(0, \vec{1})$, are positive definite. Thus, there are symmetric positive definite matrices $P_1, P_2 \in \mathbb{R}^{2d \times 2d}$ such that for all $-\varepsilon \leq \mu_i \leq \varepsilon$ and $0 < 1 - \varepsilon \leq \phi_i \leq 1 + \varepsilon$:

$$(\mu, \phi - \vec{1})^\top P_1 (\mu, \phi - \vec{1}) \leq V_l(\mu, \phi),$$
$$(\mu, \phi - \vec{1})^\top P_2 (\mu, \phi - \vec{1}) \geq V_u(\mu, \phi).$$

All in all, then, we have:

$$(\mu, \phi - \vec{1})^\top P_1 (\mu, \phi - \vec{1}) \leq V_l(\mu, \phi) \overset{(5.32)}{\leq} V(\mu, \Sigma) \overset{(5.32)}{\leq} V_u(\mu, \phi) \leq (\mu, \phi - \vec{1})^\top P_2 (\mu, \phi - \vec{1}). \tag{5.33}$$

Due to equivalence of norms, there are constants $C_3, C_4 > 0$ such that

$$\|(\mu, \phi - \vec{1})\|_2^2 \leq \frac{1}{C_3} (\mu, \phi - \vec{1})^\top P_1 (\mu, \phi - \vec{1}), \tag{5.34}$$

$$(\mu, \phi - \vec{1})^\top P_2 (\mu, \phi - \vec{1}) \leq C_4 \|(\mu, \phi - \vec{1})\|_2^2. \tag{5.35}$$

Recalling the constants from (5.30) and (5.31), we define $C_5 := \max\{C_1, C_2\}$ and $\kappa :=$

$\min\{\kappa_1, \kappa_2\}$. Then with $C := \frac{C_4}{C_3}C_5^2$, we finally have that

$$\begin{aligned}
V(\mu(t), \Sigma(t)) &\overset{(5.33)}{\leq} (\mu(t), \phi(t) - \vec{1})^\top P_2 (\mu(t), \phi(t) - \vec{1}) \\
&\overset{(5.35)}{\leq} C_4 \|(\mu(t), \phi(t) - \vec{1})\|_2^2 \\
&= C_4 \left(\|\mu(t)\|_2^2 + \|\phi(t) - \vec{1}\|_2^2 \right) \\
&\overset{(5.30),(5.31)}{\leq} C_4 \left(C_1^2 e^{-2\kappa_1 t} \|\mu(0)\|_2^2 + C_2^2 e^{-2\kappa_2 t} \|\phi(0) - \vec{1}\|_2^2 \right) \\
&\leq C_4 C_5^2 e^{-2\kappa t} \left(\|\mu(0)\|_2^2 + \|\phi(0) - \vec{1}\|_2^2 \right) \\
&= C_4 C_5^2 e^{-2\kappa t} \|(\mu(0), \phi(0) - \vec{1})\|_2^2 \\
&\overset{(5.34)}{\leq} \frac{C_4}{C_3} C_5^2 e^{-2\kappa t} (\mu(0), \phi(0) - \vec{1})^\top P_1 (\mu(0), \phi(0) - \vec{1}) \\
&\overset{(5.33)}{\leq} C e^{-2\kappa t} V(\mu(0), \Sigma(0))
\end{aligned}$$

for all $(\mu(0), \phi(0)) \in \mathcal{B}_{r_\mu}(0) \times \mathcal{B}_{r_\phi}(\vec{1})$ with $r_\mu, r_\phi \in]0, \varepsilon[$ such that $-\varepsilon \leq \phi_i(t) - 1 \leq \varepsilon$ and $-\varepsilon \leq \mu_i(t) \leq \varepsilon$ for all $t \geq 0$ and all $i = 1, ..., d$. □

Remark 5.12. *In the proof of Theorem 5.11 we have shown that for a given $\varepsilon \in]0, 1[$ there exist $r_\mu, r_\phi \in]0, \varepsilon[$ such that for any $(\mu(0), \phi(0)) \in \mathcal{B}_{r_\mu}(0) \times \mathcal{B}_{r_\phi}(\vec{1}) =: \mathcal{I}$ we have $-\varepsilon \leq \phi_i(t) - 1 \leq \varepsilon$ and $-\varepsilon \leq \mu_i(t) \leq \varepsilon$ for all $t \geq 0$ and all $i = 1, ..., d$. For this set of initial states, i.e., for $(\mu(0), \phi(0)) \in \mathcal{I}$, the optimal value function $V_\infty(\mu(0), \phi(0))$ is finite due to the exponential decay of the stage cost, see* (5.29)*. Thus, the use of Theorem 3.5 in the proof of Theorem 5.11 implies that MPC "works" for initial values in $\mathcal{I}$ and a sufficiently large horizon N (in the sense that the desired equilibrium $\bar{\rho}$ is asymptotically stable for the MPC closed loop and that the closed loop trajectory stays in a recursively feasible set). The set $\mathcal{I}$ seems rather limited, but can be (greatly) extended: Given any (large) bound $\Gamma > 0$, any set $\tilde{\mathcal{I}} \supseteq \mathcal{I}$ in which* all *initial values $(\mu(0), \phi(0))$ can be steered inside $\mathcal{I}$ with total costs less than Γ is not a subset of the problematic set $\mathcal{O}$ from Theorem 3.5 because we can bound the value function uniformly on $\tilde{\mathcal{I}}$. Hence, we can find a compact set $\mathcal{C} \subset \tilde{\mathcal{I}}$ as required in Theorem 3.5, which gives us a basin of attraction $\mathcal{S} \supseteq \mathcal{C}$ on which MPC "works".*

Remark 5.13. *If $\Sigma(t)$ in Theorem 5.11 is a diagonal matrix for all $t \geq 0$, then the function $V(\mu, \Sigma)$ from* (5.28) *can be viewed as a function of the vector $\phi = (\phi_1, ..., \phi_d) = (\Sigma_{11}, ..., \Sigma_{dd})$. Then in the Taylor expansion of $V(\mu, \phi)$ around $(0, \vec{1})$ no constant or linear terms appear. As such, with the same arguments as in the proof of Theorem 5.11, we arrive at* (5.29) *without needing the bounds V_l and V_u. Consequently, we do not need to impose bounds on ϕ_i or μ_i (as long as $\Sigma(t)$ is positive definite) and hence get the exponential controllability property globally.*

5.3.2 The Ornstein–Uhlenbeck Process

For more specific dynamics, the results of Theorem 5.11 can be improved by determining the constants C and κ or at least (tighter) estimates of those. To this end, we look

more closely at the Ornstein–Uhlenbeck process introduced in Example 5.1, i.e., we consider (5.5) with $A, B, D, K(t), c(t)$ as in (5.8). We recall that, as in Example 5.1, we impose control constraints $k_i(t) > -\theta_i$, $i = 1, ..., d$.

Due to Lemma 5.6, we assume that the target probability density function is characterized by $(\bar{\mu}, \bar{\Sigma}) = (0, I)$, i.e., $\bar{\rho}$ is given by (5.20). The stage cost is given by (5.15). Numerical simulations suggest that $(\bar{\mu}, \bar{\Sigma}) = (0, I)$ is globally asymptotically stable for the MPC closed loop for the shortest possible horizon $N = 2$ also for $\gamma > 0$. Although performance degrades with shorter N and depends on the sampling time T_s, the stability property is maintained for various initial conditions $\mathring{\rho}$, sampling times T_s, and weights $\gamma \geq 0$, cf. the examples in this section. If we could prove exponential controllability of the system with respect to stage cost (5.15) with $C = 1$ independent of the weight γ, then Theorem 3.4 would confirm our conjecture drawn from numerical findings. A canonical control candidate in this matter is $(\bar{K}, \bar{c})$ because it induces no control cost. However, as shown in the following, this simple solution often does not work.

The rest of this section is divided into two parts. In the first, we state results for general weights $\gamma \geq 0$. In particular, for the one-dimensional Ornstein–Uhlenbeck process, we prove that $(\bar{\mu}, \bar{\Sigma}) = (0, 1)$ is globally asymptotically stable for the MPC closed loop for $N \geq 2$. The multi-dimensional case is more involved and thus, we consider the special case $\gamma = 0$ in the second part. Note that although control costs are eliminated, this scenario is not covered by Theorem 5.8 due to the control constraints $k_i(t) > -\theta_i$, $i = 1, ..., d$.

The Case of $\gamma \geq 0$

To simplify the notation, in this part we focus on control sequences that are piecewise constant in time, i.e., for fixed $k \in \mathbb{N}_0$, $K(k)$ and $c(k)$ are constant. These piecewise constant control sequences fit well with the notation of $\Sigma(k)$ introduced in the beginning of Section 5.3. All simulations were carried out with such controls. Otherwise one should specify how to evaluate the stage cost (5.15) in every discrete time step. For instance, one could integrate over time, e.g., use $\int_{t_k}^{t_{k+1}} \left\| BK(t) - B\bar{K} \right\|_F^2 \, dt$ or a discretized version thereof. The results presented in this part extend to controls that are not piecewise constant if the above integral is used.

We start by illustrating the problems when using the canonical control candidate $(\bar{K}, \bar{c})$, see the following example.

Example 5.14. *Consider the 1D Ornstein–Uhlenbeck process with (model) parameters*

$$A = -\theta = -4, \quad B = 1, \quad D = \varsigma = \sqrt{6}, \quad (\mathring{\mu}, \mathring{\Sigma}) = (14, 12), \quad (\bar{\mu}, \bar{\Sigma}) = (0, 1)$$

and some $\gamma > 0$. From (5.9), (5.11), and (5.12) we can calculate the "equilibrium control"

$$(\bar{K}, \bar{c}) = (\varsigma^2/(2\bar{\Sigma}) - \theta, 0) = (\varsigma^2/2 - \theta, 0) = (-1, 0)$$

that can be used to asymptotically stabilize $(\bar{\mu}, \bar{\Sigma})$. We set the MPC horizon N to 2, the sampling rate T_s to 0.1, and use the stage cost (5.15), which, in this case, coincides with (5.21). In Figure 5.1 (left), we illustrate the incurring cost $J_2((\mu_{\mathbf{u}_n}(n), \Sigma_{\mathbf{u}_n}(n)), \mathbf{u}_n)$ in every MPC step $n = 0, ..., 14$, where $\mathbf{u}_n$ denotes the (open-loop) control sequence in the n-th MPC step. We consider the equilibrium control $\mathbf{u}_n \equiv (\bar{K}, \bar{c}) =: \bar{u}$ (blue circle) as well as optimal open-loop control sequences $\mathbf{u}_n^*$ for $\gamma = 0.015$ (red cross) and for $\gamma = 10^{-5}$ (green diamond). For a high enough weight $\gamma > 0$, even the optimal sequence leads

to temporarily increasing cost. Since for the optimal open-loop control sequence $\mathbf{u}^*$ *we have* $J_2((\mathring{\mu}, \mathring{\Sigma}), \mathbf{u}^*) = V_2((\mathring{\mu}, \mathring{\Sigma}))$*, cf. Definition 3.2, the figure also shows that the optimal value function* V_2 *grows. In particular, this function cannot be a Lyapunov function for* $N = 2$*. Hence, based on this numerical evidence, Theorem 3.4 implies that exponential controllability with* $C = 1$ *cannot hold.*

Yet, from Figure 5.1 (right), which depicts the normalized Euclidean distances

$$\Delta(\mu) := \|\mu - \bar{\mu}\|_2^2 \,/\, \|\mathring{\mu} - \bar{\mu}\|_2^2 \quad \text{and} \quad \Delta(\Sigma) := \left\|\Sigma - \bar{\Sigma}\right\|_F^2 \,/\, \|\mathring{\Sigma} - \bar{\Sigma}\|_F^2 \tag{5.36}$$

of the mean $\mu(n)$ *(filled) and the variance* $\Sigma(n)$ *(empty) from the target* $(\bar{\mu}, \bar{\Sigma})$ *in every MPC step for the equilibrium control* $(\bar{K}, \bar{c})$ *(blue circle) and for the optimal open-loop control sequences* $\mathbf{u}_n^*$ *for* $\gamma = 0.015$ *(red square) and for* $\gamma = 10^{-5}$ *(green diamond), we see that the target is reached in all cases.*

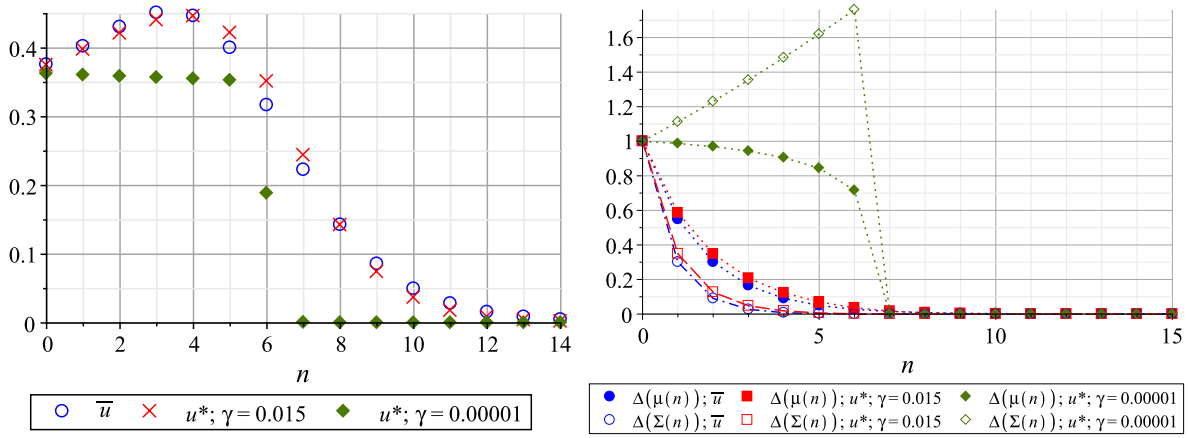


Figure 5.1: Objective function J_2 with the stage cost given by (5.15) (left) and normalized differences (5.36) (right) for Example 5.14.

In light of Example 5.14 it is apt to explore other means of proving (global) asymptotic stability of the MPC closed loop (for $N = 2$). Already in the proof of Theorem 5.11 we needed to treat the mean $\mu(t)$ and the covariance matrix $\Sigma(t)$ separately. For the dynamics given by the Ornstein–Uhlenbeck process, we can indeed decouple these two. Note that the (d-dimensional) Ornstein–Uhlenbeck process from Example 5.1 satisfies the requirements of the following proposition due to the constraints on $K(t)$, i.e., $k_i(t) > -\theta_i$ for $i = 1, \dots, d$.

Proposition 5.15. *Consider the system* (5.5) *associated to a linear stochastic process defined by* (5.3) *with a Gaussian initial condition* (5.4) *and a desired PDF* $\bar{\rho}$ *given by* (5.20)*. Assume that* $A - BK(t)$ *is a negative definite diagonal matrix for all* $t \geq 0$ *and that* B *is a square and invertible matrix. Furthermore, let the stage cost be given by* (5.15) *with* $\gamma \geq 0$*. Then each component of the mean* $\mu_i(t)$ *converges exponentially towards* $\bar{\mu}_i = 0$ *in the MPC closed loop for each optimization horizon* $N \geq 2$*.*

Proof. Let $N \geq 2$. If we express the stage cost (5.15) in terms of (μ, Σ), cf. (5.16), then the objective function J_N, cf. ($\mathrm{OCP_N}$), can be written as

$$J_N((\mathring{\mu}, \mathring{\Sigma}), \mathbf{u}) = J_N((\mathring{\mu}, \mathring{\Sigma}), (\mathbf{K}, \mathbf{c})) = \sum_{k=0}^{N-1} \ell((\mu(k), \Sigma(k), (K(k), c(k))) \tag{5.37}$$

with

$$\ell((\mu(k), \Sigma(k), (K(k), c(k))) = 2^{-d}\pi^{-d/2}\ell_{\Sigma,\mu}(k) + \frac{\gamma}{2}\ell_{K,c}(k),$$

where

$$\ell_{\Sigma,\mu}(k) := |\Sigma(k)|^{-1/2} + 1 - 2\left|\frac{1}{2}(\Sigma(k) + I)\right|^{-1/2} \exp\left(-\frac{1}{2}\mu(k)^\top(\Sigma(k) + I)^{-1}\mu(k)\right), \quad (5.38a)$$

$$\ell_{K,c}(k) := \left\|BK(k) - B\bar{K}\right\|_F^2 + \|Bc(k)\|_2^2, \quad (5.38b)$$

cf. (5.21). Let $(\mathbf{K}^*, \mathbf{c}^*) := (K^*(k), c^*(k))_{k=0,\dots,N-1}$ be the optimal control sequence that, together with the corresponding state trajectory $(\mu^*(k), \Sigma^*(k))_{k=0,\dots,N-1}$, minimizes (5.37) given some initial value $(\mathring{\mu}, \mathring{\Sigma})$.

Looking at the continuous-time dynamics (5.5), we note that $K(t)$ influences both the mean $\mu(t)$ and the covariance matrix $\Sigma(t)$, while $c(t)$ has an impact on $\mu(t)$ only. Therefore, we are able to control the mean $\mu(t)$ independently of the covariance matrix $\Sigma(t)$. Moreover, since $A - BK(t) =: M(t)$ is a (negative definite) diagonal matrix, i.e., $M(t) = \operatorname{diag}(m_1(t), \dots, m_d(t))$, defining $\tilde{c}(t) := Bc(t)$ yields

$$\dot{\mu}_i(t) = m_i(t)\mu_i(t) + \tilde{c}_i(t), \quad \mu_i(0) = \mathring{\mu}_i$$

for $i = 1, \dots, d$. This ODE can be solved to obtain the sampled system

$$\mu_i(k+1) = \underbrace{\exp\left(\int_{t_k}^{t_{k+1}} m_i(s)\,ds\right)}_{\in]0,1[}\left[\mu_i(k) + \int_{t_k}^{t_{k+1}} \tilde{c}_i(s) \underbrace{\exp\left(-\int_{t_k}^{s} m_i(\delta)\,d\delta\right)}_{>1} ds\right]$$

for $i = 1, \dots, d$. In the case of piecewise constant controls, this simplifies to

$$\begin{aligned}\mu_i(k+1) &= \exp(m_i(k)T_s)\left[\mu_i(k) + \frac{\tilde{c}_i(k)}{m_i(k)}(1 - \exp(-m_i(k)T_s))\right] \\ &= \underbrace{e^{m_i(k)T_s}}_{\in]0,1[}\mu_i(k) + \tilde{c}_i(k)\underbrace{\frac{e^{m_i(k)T_s} - 1}{m_i(k)}}_{>0},\end{aligned} \quad (5.39)$$

where we remind that $T_s = t_{k+1} - t_k$.

To prove our assertion, it is sufficient to exclude two things in the sampled system:

1. It is optimal to not approach or to deviate from the target zero in any component of the mean at any time, i.e., $\exists \tilde{k} \in \{1, \dots, N-1\}$, $j \in \{1, \dots, d\}$:

$$\begin{cases}\mu_j^*(\tilde{k}) \geq \mu_j^*(\tilde{k}-1) & \text{if } \mu_j^*(\tilde{k}-1) > 0, \quad (5.40a)\\ \mu_j^*(\tilde{k}) \leq \mu_j^*(\tilde{k}-1) & \text{if } \mu_j^*(\tilde{k}-1) < 0, \quad (5.40b)\\ \mu_j^*(\tilde{k}) \neq \mu_j^*(\tilde{k}-1) & \text{if } \mu_j^*(\tilde{k}-1) = 0. \quad (5.40c)\end{cases}$$

2. It is optimal to overshoot the target zero in any component of the mean at any time, i.e., $\exists \tilde{k} \in \{1, \dots, N-1\}, j \in \{1, \dots, d\}$:

$$\begin{cases}\mu_j^*(\tilde{k}) < 0 & \text{if } \mu_j^*(\tilde{k}-1) > 0, \quad (5.41a)\\ \mu_j^*(\tilde{k}) > 0 & \text{if } \mu_j^*(\tilde{k}-1) < 0. \quad (5.41b)\end{cases}$$

We now prove that none of these two points occur in the sampled system. Due to (5.39), we may look at one component μ_j at a time. For a given $j \in \{1, \dots, d\}$, let $\tilde{k}$ be the smallest $k \in \{1, \dots, N-1\}$ for which either (5.40) or (5.41) holds. Analogous to $\tilde{c}(t)$, we define $\tilde{c}^*(k) := Bc^*(k)$. Due to $m_j(k) < 0$ we see from (5.39) that with $\tilde{c}_j(k) = 0$ we always get $|\mu_j(\tilde{k})| \leq |\mu_j(\tilde{k}-1)|$, with equality if and only if $\mu_j(\tilde{k}-1) = 0$. For any given $\Sigma(k)$, a lower $|\mu_j(k)|$ yields a lower state cost $\ell_{\Sigma,\mu}(k)$, cf. (5.38). Moreover, for any given $M(k)$ (and thus $K(k)$), choosing $\tilde{c}_j(k) = 0$ is optimal with respect to the control cost $\ell_{K,c}(k)$.

With these preliminary considerations in mind, let us first assume that (5.41) holds for $\tilde{k}$. In the following, we construct a control sequence that performs strictly better, contradicting optimality of the current control sequence and thus excluding (5.41). To this end, we note that there exists some $\tilde{c}_j(\tilde{k}-1) =: \tilde{c}_j^0(\tilde{k}-1)$ such that $\mu_j(\tilde{k}) = 0$, cf. (5.39). Since $|B| \neq 0$, there is a sequence $\mathbf{c}^{**} := (c^{**}(k))_{k=0,\dots,N-1}$ such that

$$\tilde{c}_i^{**}(k) := \begin{cases} \tilde{c}_j^0(\tilde{k}-1), & i = j \wedge k = \tilde{k}-1, \\ 0, & i = j \wedge k \geq \tilde{k}, \\ \tilde{c}_i^*(k), & \text{otherwise}, \end{cases} \qquad i = 1, \dots, d,\ k = 0, \dots, N-1. \tag{5.42}$$

From (5.39) we see that $|\tilde{c}_j^0(\tilde{k}-1)| < |\tilde{c}_j^*(\tilde{k}-1)|$. Thus, the new control sequence $(K^*(k), c^{**}(k))_{k=0,\dots,N-1}$ outperforms the optimal control sequence in terms of (total) control cost:

$$\ell_{K^*,c^{**}}(k) \begin{cases} = \ell_{K^*,c^*}(k), & k = 0, \dots, \tilde{k}-2, \\ < \ell_{K^*,c^*}(k), & k = \tilde{k}-1, \\ \leq \ell_{K^*,c^*}(k), & k = \tilde{k}, \dots, N-1. \end{cases}$$

For the corresponding state trajectory $(\mu^{**}(k), \Sigma^{**}(k))_{k=0,\dots,N-1}$ and all $k \in \{0, \dots, N-1\}$, we have $\Sigma^{**}(k) = \Sigma^*(k)$ and

$$\mu_i^{**}(k) = \begin{cases} 0, & i = j \wedge k \geq \tilde{k}, \\ \mu_i^*(k), & \text{otherwise}. \end{cases}$$

Therefore, we have reduced the (total) state cost as well:

$$\ell_{\Sigma^{**},\mu^{**}}(k) \begin{cases} = \ell_{\Sigma^*,\mu^*}(k), & k = 0, \dots, \tilde{k}-1, \\ < \ell_{\Sigma^*,\mu^*}(k), & k = \tilde{k}, \\ \leq \ell_{\Sigma^*,\mu^*}(k), & k = \tilde{k}+1, \dots, N-1. \end{cases}$$

In conclusion,

$$J_N((\mathring{\mu}, \mathring{\Sigma}), (\mathbf{K}^*, \mathbf{c}^{**})) < J_N((\mathring{\mu}, \mathring{\Sigma}), (\mathbf{K}^*, \mathbf{c}^*)), \tag{5.43}$$

which contradicts optimality of $(K^*(k), c^*(k))_{k=0,\dots,N-1}$ and thus excludes (5.41).

To exclude (5.40), we proceed in a similar manner. Assuming (5.40) holds for $\tilde{k}$, we can find a sequence $\mathbf{c}^{**} := (c^{**}(k))_{k=0,\dots,N-1}$ such that

$$\tilde{c}_i^{**}(k) := \begin{cases} 0, & i = j \wedge k = \tilde{k}-1, \\ \tilde{c}_i^*(k), & \text{otherwise}, \end{cases} \quad i = 1, \dots, d,\ k = 0, \dots, N-1.$$

From the preliminary considerations above we know that (5.40) does not occur with $\tilde{c}_j^{**}(\tilde{k}) = 0$, i.e., $\tilde{c}_j^*(\tilde{k}) \neq 0$. Thus, the new control sequence exhibits a lower (total) control

cost. As above, we denote the corresponding state trajectory by $(\mu^{**}(k), \Sigma^{**}(k))_{k=0,\dots,N-1}$ and once again $\Sigma^{**}(k) = \Sigma^*(k)$. Clearly we have $\mu_j^{**}(k) = \mu_j^*(k)$ for $k = 0, \dots, \tilde{k} - 1$ and $|\mu_j^{**}(\tilde{k})| < |\mu_j^*(\tilde{k})|$, which results in

$$\ell_{\Sigma^{**},\mu^{**}}(k) \begin{cases} = \ell_{\Sigma^*,\mu^*}(k), & k = 0, \dots, \tilde{k} - 1, \\ < \ell_{\Sigma^*,\mu^*}(k), & k = \tilde{k}. \end{cases}$$

In addition, we can make sure that $\ell_{\Sigma^{**},\mu^{**}}(k) \leq \ell_{\Sigma^*,\mu^*}(k)$ for $k = \tilde{k} + 1, \dots, N - 1$: Using the new control sequence, there are three distinct cases that can occur in the next time step $\tilde{k} + 1$ for the j-th component. If (5.40) holds, then we repeat this procedure, reducing the cost also for $\tilde{k}+1$. If (5.41) holds, then we construct another control sequence analogous to (5.42), arriving at a lower cost overall. If neither (5.40) nor (5.41) hold, then $|\mu_j^{**}(\tilde{k} + 1)| < |\mu_j^*(\tilde{k} + 1)|$ since $|\mu_j^{**}(\tilde{k})| < |\mu_j^*(\tilde{k})|$, cf. (5.39), which again results in a reduced state cost for $\tilde{k} + 1$. This can be done iteratively until we arrive at (5.43), thus excluding (5.40).

Therefore, we have shown monotone convergence of $\mu_i(t)$ to $\bar{\mu}_i$. Since the ODE for $\mu(t)$ in (5.5) is linear, the convergence is indeed exponential. □

We note that the proof of Proposition 5.15 is the same if we include box constraints on $\tilde{c}(t) = Bc(t)$, i.e., $\tilde{c}_1 \leq \tilde{c}(t) \leq \tilde{c}_u$ with $\tilde{c}_l \leq 0 \leq \tilde{c}_u$. Furthermore, Proposition 5.15 extends to other stochastic processes where the dynamics are given by (5.5) provided that

- each component of the mean can be controlled separately and
- we can approach the target (in each component) invoking zero control cost (with respect to $Bc(k)$) regardless of how $K(k)$ is chosen.

While it is debatable whether the first ingredient is really necessary, Example 5.16 illustrates what happens if the second property is violated.

Example 5.16. *Consider a shifted version of Example 5.14: instead of* $(\bar{\mu}, \mathring{\mu}) = (0, 14)$, *we consider* $(\bar{\mu}, \mathring{\mu}) = (1, 15)$. *The other model parameters remain the same. In order to take the control constraint* $K(t) > -\theta$ *into account, we set* $K(t) + \theta \geq \varepsilon$ *with* $\varepsilon = 10^{-8}$ *in our numerical simulation. Due to* (5.12), *we have* $(\bar{K}, \bar{c}) = (-1, 3)$. *In this example, we specifically use the original stage cost* (5.15), *not the modified cost* (5.22). *Looking at Figure 5.1 from Example 5.14, for low enough values of* γ *we expect the variance to increase at the beginning when using the calculated optimal control, which indeed is the case for* $\gamma = 10^{-5}$, *cf. Table 5.1. However, the mean* μ *also grows in time, which is due to* (5.9)*: with* $\bar{c} = 3$, *the mean does not converge to its target for all admissible* K, *and deviating from* $\bar{c} = 3$ *enough to make a difference seems too expensive. This results in a PDF that is drifting away from its target rather than converging towards it, as desired.*

Remark 5.17. *The effect of drifting away from the target as in Example 5.16 did not occur in Chapter 4 since the variance could not be controlled. In particular, it was impossible to choose "unsuitable" values for* K.

Of course, using the modified stage cost (5.22) restores the second key property: we can again approach the target mean (in each component) while invoking zero control cost with respect to $Bc(k)$ for any admissible $K(k)$. Needless to say, rerunning the numerical

n	0	1	2	3	4	5	6	7	...	199
$\mu(n)$	15	15.23	15.47	15.7	15.93	16.15	16.38	16.61	...	72.38
$\Sigma(n)$	12	12.6	13.2	13.8	14.4	15	15.6	16.2	...	131.4
$K^*(0)$	$\varepsilon-4$	$\varepsilon-4$	$\varepsilon-4$	$\varepsilon-4$	$\varepsilon-4$	$\varepsilon-4$	$\varepsilon-4$	$\varepsilon-4$	...	$\varepsilon-4$
$c^*(0)$	2.34	2.32	2.3	2.29	2.28	2.27	2.27	2.26	...	3
$V_2(n)$	.362	.361	.359	.357	.356	.354	.353	.351	...	.307

Table 5.1: State, associated feedback control (the first value of the optimal control sequence $\mathbf{u}_n^*$, cf. Algorithm 3.1), and optimal value function $V_2((\mu(n), \Sigma(n))) =: V_2(n)$ in each MPC step for Example 5.16 with $\gamma = 10^{-5}$.

simulation of Example 5.16 with the modified stage cost, we end up with the exact same behavior as in Example 5.14 (with μ shifted by 1).

Having established exponential convergence of the mean in Proposition 5.15, we can confirm our numerical findings in the one-dimensional case.

Proposition 5.18. *Consider the one-dimensional Ornstein–Uhlenbeck process from Example 5.1, i.e.,* (5.5) *with* $A = -\theta < 0$, $B = 1$, $D = \varsigma > 0$, $K(t) > -\theta$ *and* $c(t) \in \mathbb{R}$. *Assume that the desired PDF* $\bar{\rho}$ *is given by* (5.20). *Furthermore, let the stage cost be given by* (5.15) *with* $\gamma \geq 0$. *Then the MPC closed loop converges to the equilibrium* $\bar{\rho}$ *for each optimization horizon* $N \geq 2$ *and each initial condition.*

Even though the process in Proposition 5.18 is one-dimensional, the proof is very technical without providing more insight and can therefore be found in the Appendix. In the multi-dimensional case, however, even if $\mathring{\mu} = \bar{\mu}$, we face again the issue of increasing cost, see the following example.

Example 5.19. *Consider the 2D Ornstein–Uhlenbeck process with (model) parameters* $A = -\mathrm{diag}(3.1, 11)$, $B = I$, $D = \mathrm{diag}(0.2, \sqrt{20})$, $\mathring{\mu} = 0 = \bar{\mu}$, $\mathring{\Sigma} = \mathrm{diag}(0.02, 200)$, $\bar{\Sigma} = I$, *and some* $\gamma > 0$. *We set the MPC horizon* N *to* 2, *the sampling rate* T_s *to* 0.2, *and use the stage cost* (5.15). *As in Example 5.14, in Figure 5.2 (left) we depict the cost* $J_2((\mu_{\mathbf{u}_n}(n), \Sigma_{\mathbf{u}_n}(n)), \mathbf{u}_n)$, *where* $\mathbf{u}_n$ *denotes the control sequence in the n-th MPC step. We consider the equilibrium control* $\mathbf{u}_n \equiv (\bar{K}, \bar{c}) =: \bar{u}$ *(blue dash-dot) as well as optimal control sequences* $\mathbf{u}_n^*$ *for* $\gamma = 0.0005$ *(red dash) and for* $\gamma = 10^{-5}$ *(green dot). As above, Figure 5.2 (left) also shows that the optimal value function* V_2 *grows, implying that exponential controllability with* $C = 1$ *cannot hold. Yet, as in Example 5.14, the target is reached in all cases, as Figure 5.2 (right) shows.*[4]

As a consequence, similar to Example 5.14, for a sufficiently large weight $\gamma > 0$, the exponential controllability property does not hold with $C = 1$. Moreover, in contrast to the mean, cf. Proposition 5.15, numerical simulations illustrate that we can neither expect monotone convergence of each component Σ_{ii} to 1, $i = 1, ..., d$, nor monotone convergence of $\|\Sigma(t) - I\|_F$ to zero.

In order to get more insight on how to develop alternative methods to circumvent this issue, we focus on the state cost (5.16) by setting $\gamma = 0$.

[4]In Figure 5.2 (right) we have depicted the normalized differences (5.36) only for the first 10 MPC steps as there are no visual changes afterwards.

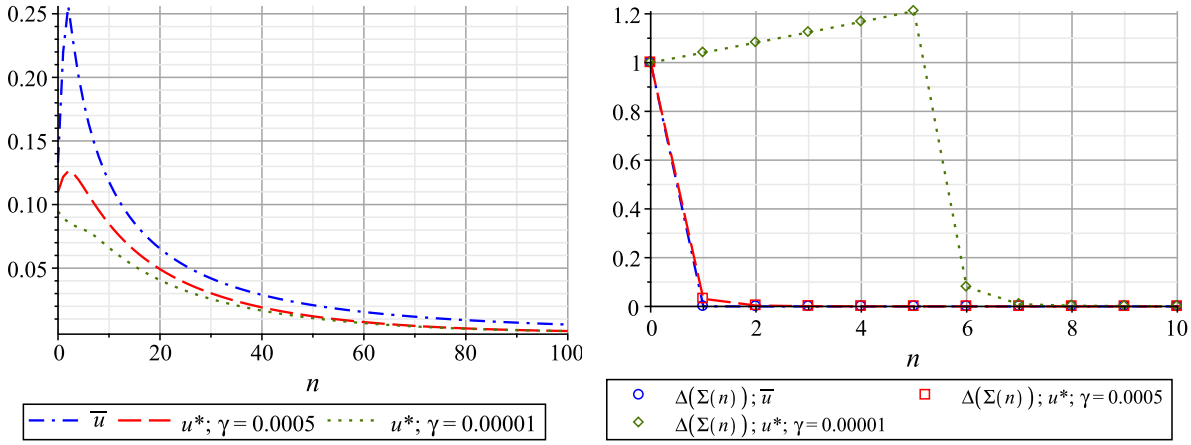


Figure 5.2: Objective function J_2 with the stage cost given by (5.15) (left) and normalized differences (5.36) (right) for Example 5.19.

The Case of $\gamma = 0$

Setting $\gamma = 0$ allows us to focus on the state cost (5.16). We recall that we still impose the control constraints $k_i(t) > -\theta_i$, cf. Example 5.1. These restrictions affect the dynamics: Assuming $\mathring{\Sigma} = \operatorname{diag}(\mathring{\sigma}_1^2, ..., \mathring{\sigma}_d^2)$ as in Example 5.1, one can show from (5.10) and (5.11) that, while $\Sigma_{ii}(t)$ can be decreased to an arbitrarily smaller positive value in one time step, there is an upper bound. More precisely, with $T_s = t_{k+1} - t_k$ one can show that

$$0 < \Sigma_{ii}(t_{k+1}) \leq \Sigma_{ii}(t_k) + 2T_s\varsigma_i^2, \tag{5.44}$$

$i = 1, ..., d$. In particular, the target variance cannot always be reached within one MPC time step, even if we allowed non-constant control coefficients as in Theorem 5.8.

In light of Example 5.19, we want to focus on steering this variance. Hence, in this part we assume that the target mean $\bar{\mu}$ is already reached, i.e., that $\mu(t) \equiv \bar{\mu}$. Moreover, to keep the connection to the previous part, we consider control sequences that are piecewise constant in time. In the case of the Ornstein–Uhlenbeck process considered here, both assumptions are sensible; if the target mean is not reached initially, i.e., $\mathring{\mu} \neq \bar{\mu}$, then the mean converges exponentially (with piecewise constant control sequences), see Proposition 5.15. However, most of the content in this section extends naturally to general dynamics (5.5) with $(\bar{\mu}, \bar{\Sigma}) = (0, I)$ if we assume that the target mean $\bar{\mu}$ is already reached or, alternatively, that it can be reached within one MPC step. This is due to Lemma 5.7, which depicts the state cost (5.16) in terms of the Eigenvalues $\phi_i(t)$ of $\Sigma(t)$. Therefore, in order to keep this generality, instead of looking at $\Sigma(t)$, we look at its Eigenvalues $\phi_i(t)$ collected in the matrix $\Phi(t) = \operatorname{diag}(\phi_i(t), \ldots, \phi_d(t))$. Likewise, instead of (5.16), we consider only the relevant part of the state cost, namely (5.24).

The goal of this section is to understand better the L^2 cost and to show that for $\gamma = 0$ the MPC closed loop is stable with $N = 2$, cf. Corollary 5.23. Regarding the former, we will look at the level sets of (5.24). Regarding the latter, we proceed as follows. First, we show in Proposition 5.20 that heading towards the target $\bar{\Sigma} = I$ leads to a lower cost. Second, since there might be other directions that yield an even lower cost in the short term—and with $N = 2$ we only look one step ahead—we need to rule out that we drift away from the target indefinitely like we did in Example 5.16.

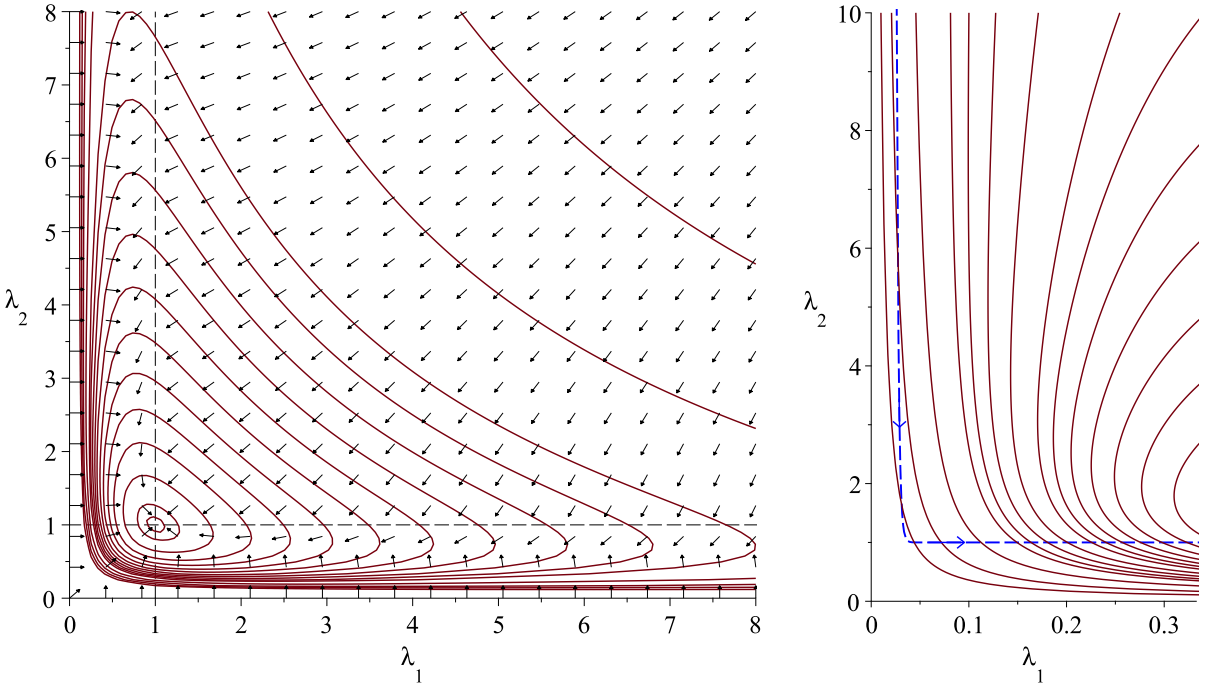


Figure 5.3: Level sets and gradient of $g(\phi)$ in the two-dimensional setting (left) and the trajectory (blue dash) from Example 5.19 (right).

We start by studying the equivalent state cost (5.24). As in the proof of Theorem 5.11 we can interpret the matrix $\Phi = \operatorname{diag}(\phi_1, \dots, \phi_d)$ as a vector $\phi = (\phi_1, \dots, \phi_d)$. In this case, we write $g(\phi)$ instead of $g(\Phi)$. Then the gradient of $g(\phi)$ is given by

$$\nabla g(\phi) = \frac{1}{2} \left(\left(\prod_{i=1}^{d} \frac{\phi_i + 1}{2} \right)^{-1/2} \left(\frac{\phi_j + 1}{2} \right)^{-1} - \left(\prod_{i=1}^{d} \phi_i \right)^{-1/2} \phi_j^{-1} \right)_{j=1,\dots,d} .$$

Figure 5.3 gives an impression of the level sets and gradients of $g(\phi)$ in the two-dimensional case and illustrates the problem that occurs in Example 5.19. First, we note that in the Ornstein–Uhlenbeck process under consideration, $\Sigma(t)$ is diagonal and therefore $\Phi_{ii}(t) = \Sigma_{ii}(t)$. Then, due to (5.11) and (5.12), each component Σ_{ii} respective ϕ_i converges monotonously to 1 when using $\bar{K}$. In particular, if ϕ_1 and ϕ_2 are both greater than 1 or both smaller than 1, the costs do not rise when using $\bar{K}$ and one can prove exponential controllability with $C = 1$ by applying the proof of the one-dimensional case, cf. Proposition 5.18, to each component. However, we may run into problems if $\operatorname{sign}(\phi_1 - 1) \neq \operatorname{sign}(\phi_2 - 1)$ as in Example 5.19.[5] Moreover, as can be seen by the arrows representing the gradient of $g(\phi)$ in Figure 5.3 (left), the optimal control sequence calculated in one MPC iteration might drive the state into the problematic region even if starting from, e.g., $\mathring{\phi}_i > 0$, $i = 1, 2$. Therefore, the sets $\{\phi \in \mathbb{R}^d \mid \forall i = 1, \dots, d : \phi_i > 1\}$

[5] This is connected to the value of α in Chapter 4, describing the relation of current and equilibrium variance. The difference to the situation here is that in Chapter 4 the cost increase stemmed from the mean being too far away from the target. Here, the target mean is already reached, and the cost increase stems from "shifting the mass of the PDF" too fast in some components compared to others, which may occur if (at least) one Eigenvalue is greater than 1 and at least one Eigenvalue is smaller than 1.

and $\{\phi \in \mathbb{R}^d \mid \forall i = 1, \dots, d : \phi_i < 1\}$ are not forward-invariant. Hence, showing the exponential controllability property only for these sets is not fruitful.

In the following, we therefore follow a different path to prove that with $N = 2$, a stable MPC closed loop is obtained.

Proposition 5.20. *Let $\Phi \neq I$. Then, for g defined in (5.24), $I - \Phi$ is a descent direction, i.e., $Dg(\Phi)(I - \Phi) < 0$ for all $\Phi \neq I$.*

Proof. Let $A, H \in \mathbb{R}^{d \times d}$. Due to

$$D(\det A)H = \det(A)\mathrm{tr}(A^{-1}H) = |A|\mathrm{tr}(A^{-1}H),$$

cf. [75, Sect. 2], we have

$$\begin{aligned}
Dg(\Phi)H = & -\frac{1}{2}|\Phi|^{-3/2} D\left(\det(\Phi)\right) H + \left|\frac{1}{2}(\Phi + I)\right|^{-3/2} D\left(\det\left(\frac{1}{2}(\Phi + I)\right)\right) H \cdot \frac{1}{2} \\
= & -\frac{1}{2}|\Phi|^{-3/2}|\Phi|\mathrm{tr}(\Phi^{-1}H) + \frac{1}{2}\left|\frac{1}{2}(\Phi + I)\right|^{-3/2}\left|\frac{1}{2}(\Phi + I)\right| \underbrace{\mathrm{tr}\left(\left(\frac{1}{2}(\Phi + I)\right)^{-1} H\right)}_{=2\mathrm{tr}((\Phi+I)^{-1}H)} \\
= & -\frac{1}{2}|\Phi|^{-1/2}\mathrm{tr}(\Phi^{-1}H) + \left|\frac{1}{2}(\Phi + I)\right|^{-1/2} \mathrm{tr}((\Phi + I)^{-1}H).
\end{aligned}$$

Therefore,

$$\begin{aligned}
Dg(\Phi)(I - \Phi) &= -\frac{1}{2}|\Phi|^{-1/2}\mathrm{tr}(\Phi^{-1}(I - \Phi)) + \left|\frac{1}{2}(\Phi + I)\right|^{-1/2} \mathrm{tr}((\Phi + I)^{-1}(I - \Phi)) \\
&= \frac{1}{2}|\Phi|^{-1/2}\left[-\mathrm{tr}(\Phi^{-1} - I) + 2\left|\frac{1}{2}\left(I + \Phi^{-1}\right)\right|^{-1/2} \mathrm{tr}((I + \Phi^{-1})^{-1}(\Phi^{-1} - I))\right].
\end{aligned}$$

Defining $\Theta := \frac{1}{2}(I + \Phi^{-1}) = \mathrm{diag}(\vartheta_1, \dots, \vartheta_d)$ with $\vartheta_i \geq \frac{1}{2}$, we have that

$$\begin{aligned}
& Dg(\Phi)(I - \Phi) < 0 \\
\Leftrightarrow\ & -\mathrm{tr}(\Phi^{-1} - I) + 2\left|\frac{1}{2}(I + \Phi^{-1})\right|^{-1/2} \mathrm{tr}((I + \Phi^{-1})^{-1}(\Phi^{-1} - I)) < 0 \\
\Leftrightarrow\ & -2\mathrm{tr}(\Theta - I) + 2|\Theta|^{-1/2}\mathrm{tr}((2\Theta)^{-1}(2\Theta - 2I)) < 0 \\
\Leftrightarrow\ & |\Theta|^{1/2}\mathrm{tr}(\Theta - I) > \mathrm{tr}(\Theta^{-1}(\Theta - I)) \\
\Leftrightarrow\ & \left(\prod_{i=1}^{d} \vartheta_i\right)^{1/2} \sum_{i=1}^{d}(\vartheta_i - 1) > \sum_{i=1}^{d}\left(1 - \frac{1}{\vartheta_i}\right).
\end{aligned}$$

For each $i = 1, \dots, d$, the inequality $\vartheta_i - 1 \geq 1 - \frac{1}{\vartheta_i}$ holds, with equality if and only if $\vartheta_i = 1$. In particular, $\sum(\vartheta_i - 1) \leq 0$ implies $\sum\left(1 - \frac{1}{\vartheta_i}\right) \leq 0$. It is therefore sufficient to show that

(a) $\prod \vartheta_i \leq 1$, if $\sum(\vartheta_i - 1) \leq 0$ and

(b) $\prod \vartheta_i \geq 1$, if $\sum \left(1 - \frac{1}{\vartheta_i}\right) \geq 0$.

First, we show (a). To this end, we have

$$\sum_{i=1}^{d} (\vartheta_i - 1) \leq 0 \quad \Leftrightarrow \quad \sum_{i=1}^{d} \vartheta_i \leq d \quad \Leftrightarrow \quad \sum_{i=1}^{d} \frac{\vartheta_i}{d} \leq 1.$$

Due to $\vartheta_i > 0$, by using the inequality of arithmetic and geometric means we get

$$\left(\prod_{i=1}^{d} \vartheta_i\right)^{1/d} \leq \sum_{i=1}^{d} \frac{\vartheta_i}{d} \leq 1,$$

from which the assertion $\prod \vartheta_i \leq 1$ follows, again due to $\vartheta_i > 0$.

To show (b), we recognize that

$$\sum \left(1 - \frac{1}{\vartheta_i}\right) \geq 0 \quad \Leftrightarrow \quad \sum \frac{1}{\vartheta_i} \leq d.$$

In particular, due to (a), we get $\prod \frac{1}{\vartheta_i} \leq 1$, from which the assertion in (b) follows. □

Corollary 5.21. *The equivalent state cost g defined in* (5.24) *has the unique stationary point I, which is the global minimum with $g(I) = 0$. Moreover, the sublevel sets $L_c := \{\Phi : g(\Phi) \leq c\}$, where $\Phi = \operatorname{diag}(\phi_1, \ldots, \phi_d)$ with $\phi_i > 0$ for each $i = 1, \ldots, d$, are connected.*

Note that g defined in (5.24) is not convex, not even in 1D. Moreover, Proposition 5.20 and Corollary 5.21 are not enough to prevent effects similar to the ones observed in Example 5.16, i.e., we cannot exclude that the MPC closed-loop solution drifts away indefinitely (albeit with monotonously decreasing cost), not even for $\gamma = 0$. This is due to possibly unbounded level sets, which we characterize in the following lemma.

Lemma 5.22. *The sublevel sets L_c from Corollary 5.21 are bounded for $c < 1$ and unbounded otherwise.*

Proof. We first show that the sublevel sets are unbounded for $c \geq 1$:

$$\begin{aligned}
g(\Phi) \leq 1 &\Leftrightarrow |\Phi|^{-1/2} - 2\left|\frac{1}{2}(\Phi + I)\right|^{-1/2} \leq 0 \\
&\Leftrightarrow \left|\frac{1}{2}(\Phi + I)\right| \leq 4|\Phi| \\
&\Leftrightarrow |(\Phi + I)| \leq 2^{d+2}|\Phi| \\
&\Leftrightarrow 2^{d+2} \geq \prod_{i=1}^{d} \frac{\phi_i + 1}{\phi_i} = \left(1 + \frac{1}{\phi_1}\right) \prod_{i=2}^{d} \frac{\phi_i + 1}{\phi_i} \\
&\Leftrightarrow \phi_1 \geq \left(2^{d+2} \prod_{i=2}^{d} \frac{\phi_i}{\phi_i + 1} - 1\right)^{-1}.
\end{aligned}$$

In particular, we can find some $\phi_1 > 0$ such that $g(\Phi) = 1$ even as $\phi_i \to \infty, i = 2, \ldots, d$. Clearly, the indexes are interchangeable, i.e., we have lower bounds on each ϕ_i, but no upper bound.

As for the other claim, we have

$$\begin{aligned} g(\Phi) &= 1 + |\Phi|^{-1/2} - 2\left|\frac{1}{2}(\Phi + I)\right|^{-1/2} \\ &> 1 + |\Phi + I|^{-1/2} - 2\left|\frac{1}{2}(\Phi + I)\right|^{-1/2} \\ &= 1 + \left(1 - 2^{1+d/2}\right)|\Phi + I|^{-1/2} =: h(\Phi). \end{aligned}$$

In particular, the sublevel sets of g are contained in those of h, i.e.,

$$\{\Phi \mid g(\Phi) \leq c\} \subset \{\Phi \mid h(\Phi) \leq c\}.$$

Moreover, for $0 \leq c < 1$, we have

$$\begin{aligned} h(\Phi) \leq c &\Leftrightarrow \left(1 - 2^{1+d/2}\right)|\Phi + I|^{-1/2} \leq c - 1 \\ &\Leftrightarrow \frac{1 - 2^{1+d/2}}{c-1} \geq |\Phi + I|^{1/2} \\ &\Leftrightarrow \left(\frac{1 - 2^{1+d/2}}{c-1}\right)^2 \geq |\Phi + I| = \prod_{i=1}^{d}(\phi_i + 1), \end{aligned} \tag{5.45}$$

which results in upper bounds $\phi_i \leq \left(\frac{1-2^{1+d/2}}{c-1}\right)^2 - 1 =: r$, $i = 1, \ldots, d$. Note that the last equivalence in (5.45) holds due to both sides being positive. Moreover, $r \in]0, \infty[$ for fixed $c \in [0, 1[$. Since $\phi_i > 0$, the (sub)level sets of h and, consequently, those of g, are contained in the d-dimensional hypercube $[0, r]^d$. □

Combining the last three results yields the following result. We recall that the state cost (5.16) that appears in the stage cost (5.15) can be expressed in terms of Eigenvalues $\phi_i(t)$ of $\Sigma(t)$, which we collect in the matrix $\Phi(t) = \operatorname{diag}(\phi_i(t), \ldots, \phi_d(t))$. In the case of the (multi-dimensional) Ornstein–Uhlenbeck process from Example 5.1 the covariance matrix $\Sigma(t)$ is diagonal and hence $\Sigma(t) = \Phi(t)$.

Corollary 5.23. *Consider the (multi-dimensional) Ornstein–Uhlenbeck process from Example 5.1, i.e.,* (5.5) *with* $A, B, D, K(t), c(t)$ *as in* (5.8) *and a desired PDF* $\bar{\rho}$ *given by* (5.20)*. Furthermore, let the stage cost be given by* (5.15) *with* $\gamma = 0$*. Consider all initial values for which the target mean is already reached, i.e.,* $\mathring{\mu} = \bar{\mu} = 0$ *and for which*

(a) $g(\mathring{\Phi}) < 1$*, for* g *defined in* (5.24)*, or*

(b) there exists some $\varepsilon \in]0, 1[$ *such that* $\phi_i(t) \leq \frac{1}{\varepsilon}$ *for all* $i = 1, ..., d$ *and all* $t \geq 0$*.*

Then for these initial conditions the equilibrium $\bar{\rho}$*, characterized by* $(0, I)$*, is asymptotically stable for the MPC closed loop for* $N = 2$*.*

Proof. First of all, for any fixed admissible Φ, it is always most beneficial (in terms of cost) to have $\mu(t) \equiv \bar{\mu}$, cf. (5.16). Since $\mathring{\mu} = \bar{\mu} = 0$ and since we can control the mean independently of Φ, we can and will always stay at $\bar{\mu} = 0$ (since MPC chooses the optimal control $c = 0$), regardless of how the control K is chosen, cf. (5.5). Hence, we will only consider the dynamics of Φ and the control K in the following.

With $N = 2$ we only look one time step into the future. Hence, when computing the optimal control sequence $\mathbf{u}_n^* = \mathbf{K}_n^* = (K_n^*(0), K_n^*(1))$ in the n-th MPC time step, we are looking for a control value $K_n^*(0)$ that minimizes the stage cost (5.15) after one (discrete) time step.[6] Due to $\gamma = 0$ and $\mathring{\mu} = \bar{\mu}$, the stage cost (5.15) can be expressed as $2^{-d}\pi^{-d/2}g(\Phi)$, see Lemma 5.7. Hence, we are effectively minimizing $g(\Phi(t_{n+1}))$ in the n-th MPC step and will thus focus on the behavior of $g(\Phi)$.

For all admissible $\Phi = \operatorname{diag}(\phi_1, ..., \phi_d)$ we have $g(\Phi) \geq 0$ and $g(\Phi) = 0 \Leftrightarrow \Phi = I$. Moreover, every component ϕ_i, $i = 1, ..., d$, will stay away from zero: if some $\phi_i \searrow 0$, then $g(\Phi) \to \infty$. Thus, to arrive at the assertion, in the following we show that with the MPC feedback law, $g(\Phi(t_n)) \searrow 0$ for $n \to \infty$.

If an admissible control value exists such that $g(\Phi(t_{n+1})) < g(\Phi(t_n))$, then the optimal control value $K_n^*(0)$ will be chosen such that the stage cost decreases as much as possible. Such an admissible control does exist for all $n \in \mathbb{N}_0$ as long as $g(\Phi(t_n)) > 0$, since the descent direction from Proposition 5.20 is always feasible (for both requirements (a) and (b)).[7] Hence, $g(\Phi(t_n))$ is bounded from below by 0 and is strictly monotonically decreasing in n as long as $g(\Phi(t_n)) > 0$. To conclude that $g(\Phi(t_n)) \searrow 0$ for $n \to \infty$ we show that for any $\varepsilon > 0$ there exists a $\delta > 0$ such that for any $g(\Phi(t_n)) > \varepsilon$ we get $g(\Phi(t_{n+1})) - g(\Phi(t_n)) < -\delta$ with the MPC feedback. To this end, we prove that all $\Phi(t_n)$ belong to a compact set and that the mapping $\Phi(t_n) \mapsto g(\Phi(t_{n+1})) - g(\Phi(t_n))$ has a continuous negative upper bound for $\Phi \neq I$.

Regarding the former, we first consider (a). Then according to Corollary 5.21 and Lemma 5.22 the sublevel set $\mathring{L} = \{\Phi : g(\Phi) \leq g(\mathring{\Phi})\}$ is connected and bounded. Moreover, $\mathring{L}$ is closed and thus compact. Hence, all $\Phi(t_n)$ belong to the compact set $\mathring{L}$. Next, we consider (b). We note that (a) is more restrictive than (b): If (a) is satisfied, then we can always find some $\varepsilon \in]0, 1[$ such that (b) is satisfied as well. Hence, we assume that $1 \leq g(\mathring{\Phi}) < \infty$, i.e., a closed but unbounded sublevel set $\mathring{L}$. However, the set $\mathring{L} \cap [0, 1/\varepsilon]^d$, which includes I, is closed and bounded, thus compact, and all $\Phi(t_n)$ stay in that set.

Regarding the latter, let $\Phi \neq I$. Then according to Proposition 5.20 $I - \Phi$ is a descent direction. In particular, there exists some $\bar{\alpha} > 0$ such that

$$g(\Phi + \alpha(I - \Phi)) < g(\Phi) \quad \text{for all} \quad \alpha \in]0, \bar{\alpha}[.$$

Since g is twice differentiable, cf. (5.24), from the Taylor expansion of g we can choose $\bar{\alpha}$ continuously dependent on Φ. This continuity carries over to $\tilde{\alpha} := \min\{\bar{\alpha}, 1/\varepsilon\}$ and we do not lose this property if we reduce $\tilde{\alpha}$ further (which might be required in order to adhere to (5.44), i.e., to guarantee the existence of a control K that yields the state $\Phi + \tilde{\alpha}(I - \Phi)$). Then $F(\Phi) := \Phi + \tilde{\alpha}(I - \Phi)$ is continuous in Φ. Moreover, $g(F(\Phi)) - g(\Phi) < 0$ for $\Phi \neq I$ and in particular, $F(\Phi)$ is admissible in both cases (a) (since the cost declines) and (b) (by construction of $\tilde{\alpha}$). Hence, F is a continuous negative upper bound for the mapping $\Phi(t_n) \mapsto g(\Phi(t_{n+1})) - g(\Phi(t_n))$. This concludes the proof. □

Since the properties of $g(\Phi)$ were derived regardless of the dynamics of the system, Corollary 5.23 can be extended to other systems (5.5), with one caveat: In each MPC

[6] We recall that the control value $K_n^*(1)$ only influences subsequent states. However, these are not included in the objective function for $N = 2$ and hence $K_n^*(1)$ does not have an impact on the objective function; see also the end of Section 4.1.

[7] Given any fixed sampling time $T_s > 0$, from any state $\Phi(t_n)$ we can find a control such that $\Phi(t_{n+1})$ is in a neighborhood of $\Phi(t_n)$, cf. (5.44). Note that the descent direction from Proposition 5.20 always adheres to the respective requirements (a) or (b) (for suitably chosen step sizes).

time step $n \in \mathbb{N}_0$ we need to be able to reduce the state cost, i.e., there must exist some admissible control such that $g(\Phi(t_{n+1})) < g(\Phi(t_n))$.

5.4 Conclusion

In this chapter, we have analyzed the stability of the closed loop generated by Model Predictive Control schemes applied to tracking problems involving the Fokker–Planck equation. We have considered a setting involving linear dynamics and Gaussian PDFs. Even in this relatively simple setting, the use of the L^2 cost, which is standard in PDE tracking problems, leads to a rather involved analysis. Particularly, stability does not always hold for the shortest possible horizon $N = 2$. Even in some cases where it does hold, the usual exponential controllability condition without overshoot (i.e., with $C = 1$) is not satisfied and a different technique for the stability analysis had to be developed.

This raises the question whether distances other than L^2 could facilitate the analysis. One alternative is to use the Wasserstein metric, which is specifically designed to measure the distance between two PDFs. To the best of our knowledge, in the general PDE setting this metric lacks a sound existence theory regarding optimal controls, in contrast to the L^2 cost, see Chapter 2. By changing the perspective from the Fokker–Planck equation (1.2) (infinite-dimensional) to the ODE system (5.5) (finite-dimensional), however, the Wasserstein cost function simplifies considerably and is even convex, see the subsequent chapter.

Hence, in the following chapter we compare various alternative cost functions in the setting of linear stochastic processes. Moreover, we switch our focus from stabilizing MPC to the more general economic MPC case, cf. Section 3.3.

Appendix

Proof of Proposition 5.18. Due to Proposition 5.15, we can assume that $\mathring{\mu}$ is arbitrarily close to $\bar{\mu} = 0$. For $|\mathring{\mu}|$ sufficiently small, we argue below that the exponential controllability condition (3.6) with respect to stage cost (5.15) holds with $C = 1$ for the control candidate $(\bar{K}, \bar{c})$. Then we apply Theorem 3.4 to conclude the assertion.

First, due to $\bar{\mu} = 0$, we have that $\bar{c} = 0$. Then, due to $\bar{\Sigma} = 1$ we see from (5.9), (5.11), and (5.12) that applying $(\bar{K}, \bar{c})$ results in

$$\mu(t) = \mathring{\mu} e^{-2(\theta+\bar{K})t} \quad \text{and} \quad \Sigma(t) = 1 + (\mathring{\Sigma} - 1)e^{-2(\theta+\bar{K})t} > 0.$$

We define

$$\bar{\theta} := \theta + \bar{K} > 0.$$

Then with Lemma 5.5 the stage cost (5.15) can be written as

$$\tilde{V}(t) := 1 + \left[1 + (\mathring{\Sigma} - 1)e^{-2\bar{\theta}t}\right]^{-1/2} - 2\left[\frac{2 + (\mathring{\Sigma} - 1)e^{-2\bar{\theta}t}}{2}\right]^{-1/2} \exp\left(-\frac{\mathring{\mu}^2 e^{-2\bar{\theta}t}}{2(2 + (\mathring{\Sigma} - 1)e^{-2\bar{\theta}t})}\right).$$

Our aim is to show $\tilde{V}(t) \leq e^{-\kappa t}\tilde{V}(0)$ for some $\kappa > 0$ (for sufficiently small $\mathring{\mu}^2$). Then (3.6) holds with overshoot bound $C = 1$ and decay rate $\delta = e^{-\kappa T_s}$. We claim that $\tilde{V}(t) \leq e^{-\kappa t}\tilde{V}(0)$ with

$$\kappa := \frac{\bar{\theta}}{\mathring{\Sigma} + 1} > 0.$$

To this end, we prove $\tilde{V}'(t) + \kappa\tilde{V}(t) \leq 0$. First, to shorten the notation, we introduce

$$a := \mathring{\Sigma} - 1 \in]-1, \infty[, \quad \tau := e^{-2\bar{\theta}t} \in]0, 1], \quad \chi := \frac{\mathring{\mu}^2\tau}{(a\tau+2)} \geq 0,$$

$$a_1 := 2\sqrt{2}e^{-\chi/2} - \left(\frac{a\tau+2}{a\tau+1}\right)^{3/2}, \text{ and } a_2 := 1 - \frac{1}{(a\tau+1)^{3/2}} - \frac{4\sqrt{2}\chi e^{-\chi/2}(a+2)}{(a\tau+2)^{3/2}}.$$

Then we can express $\frac{1}{\bar{\theta}}(\tilde{V}'(t) + \kappa\tilde{V}(t))$ by $-\frac{h(\tau)}{(a\tau+2)^{3/2}(a+2)}$, where

$$h(\tau) := a_1\left(a\tau(a+2) + a\tau + 2\right) - a_2(a\tau+2)^{3/2},$$

which means we have to show that $h(\tau) \geq 0$. We consider the two cases $\mathring{\Sigma} > 1$ respective $a > 0$ and $\mathring{\Sigma} < 1$ respective $a < 0$. The case $\mathring{\Sigma} = 1$ is trivial.

First, let us assume $a > 0$. In this case, we set $\mathring{\mu}^2 = \varepsilon a$ for some $\varepsilon \geq 0$. Then

$$\begin{aligned} h(\tau) &= a_1\left(a\tau(a+2) + a\tau + 2\right) - a_2(a\tau+2)^{3/2} \\ &\geq a_1\left(a\tau(a+2) + a\tau + 2\right) - a_3(a\tau+2)^{3/2} \end{aligned}$$

with

$$a_3 := 1 - \frac{1}{(a\tau+1)^{3/2}} - \frac{4\sqrt{2}\chi e^{-\chi/2}}{(a\tau+2)^{1/2}} \geq a_2$$

due to $a + 2 \geq a\tau + 2$. If $a_1 \geq 0$, which we prove below, then

$$\begin{aligned} h(\tau) &\geq a_1\left(a\tau+2\right) + \underbrace{a_1 a\tau(a+2)}_{\geq a_1 a\tau(a\tau+2)} - a_3(a\tau+2)^{3/2} \\ &\geq \underbrace{(a\tau+2)}_{>0}(a_1 + a_1 a\tau - a_3\sqrt{a\tau+2}) \\ &= (a\tau+2)(a_1(a\tau+1) - a_3\sqrt{a\tau+2}), \end{aligned}$$

i.e., it is left to show that $a_1(a\tau+1) - a_2\sqrt{a\tau+2} \geq 0$. Furthermore, if $a_3 \geq 0$, then

$$\begin{aligned} a_1(a\tau+1) - a_3\sqrt{a\tau+2} &\geq a_1(a\tau+1) - a_3\left(\frac{a\tau}{2\sqrt{2}} + \sqrt{2}\right) \\ &= a_1(a\tau+1) - \sqrt{2}a_3\left(\frac{a\tau}{4} + 1\right) \\ &= a_1(a\tau+1) - \sqrt{2}a_3\left(a\tau+1\right) + \frac{3}{4}\sqrt{2}a_3 a\tau \\ &\geq (a\tau+1)(a_1 - \sqrt{2}a_3), \end{aligned}$$

reducing the problem further to

$$a_1 - \sqrt{2}a_3 \geq 0. \tag{5.46}$$

Since $a_1 \geq 0$ follows from (5.46), we only need to prove (5.46) and $a_3 \geq 0$. Regarding the latter, with $\bar{a} := a\tau \in [0, \infty[$ and for $\varepsilon \in [0, \frac{1}{2}]$, we have

$$\begin{aligned} a_3 &= 1 - \frac{1}{(a\tau+1)^{3/2}} - \frac{4\sqrt{2}\chi e^{-\chi/2}}{(a\tau+2)^{1/2}} \\ &= 1 - \frac{1}{(\bar{a}+1)^{3/2}} - \frac{4\sqrt{2}\varepsilon\bar{a}}{(\bar{a}+2)^{3/2}}\exp\left(-\frac{\varepsilon\bar{a}}{2(\bar{a}+2)}\right) \\ &\geq 1 - \frac{1}{(\bar{a}+1)^{3/2}} - \frac{2\sqrt{2}\bar{a}}{(\bar{a}+2)^{3/2}}\exp\left(-\frac{\bar{a}}{4(\bar{a}+2)}\right) \geq 0, \end{aligned}$$

where the first inequality follows since a_3 is monotonically decreasing in ε for $\varepsilon \in [0, \frac{1}{2}]$:

$$\frac{\partial a_3}{\partial \varepsilon} = \underbrace{\frac{2\sqrt{2}\bar{a}}{(\bar{a}+2)^{5/2}} \exp\left(-\frac{\varepsilon\bar{a}}{2(\bar{a}+2)}\right)}_{\geq 0} \underbrace{[\varepsilon\bar{a} - 2(\bar{a}+2)]}_{<0} \leq 0.$$

Now, we can turn our attention to (5.46), which we claim holds for $\varepsilon \in [0, \frac{1}{4}]$. With $\bar{a} = a\tau$ as above, we get

$$a_1 - \sqrt{2}a_3 = 2\sqrt{2}\exp\left(-\frac{\varepsilon\bar{a}}{2(\bar{a}+2)}\right)\left(1 + \frac{2\sqrt{2}\varepsilon\bar{a}}{(2+\bar{a})^{3/2}}\right) - \left(\frac{\bar{a}+2}{\bar{a}+1}\right)^{3/2} - \sqrt{2}\left(1 - \frac{1}{(\bar{a}+1)^{3/2}}\right),$$

which unfortunately is not monotone with respect to ε. We know, however, that

$$(a_1 - \sqrt{2}a_3)_{|\bar{a}=0} = 0 \quad \text{and} \quad (a_1 - \sqrt{2}a_3) \to \frac{2\sqrt{2}}{\sqrt{e^{\varepsilon}}} - (\sqrt{2}+1) \text{ as } \bar{a} \to \infty, \tag{5.47}$$

where the limit is positive for $\varepsilon \in [0, \frac{1}{4}]$. Moreover, in the special case $\varepsilon = 0$, we see that

$$\frac{d(a_1 - \sqrt{2}a_3)}{d\bar{a}} = \frac{3}{2(\bar{a}+1)^2}\left(\sqrt{1 + \frac{1}{\bar{a}+1}} - \frac{\sqrt{2}}{\sqrt{\bar{a}+1}}\right) \geq 0 \Leftrightarrow \frac{\bar{a}}{\bar{a}+1} \geq 0 \Leftrightarrow \bar{a} \geq 0, \tag{5.48}$$

which, together with (5.47), proves that $h(\tau) \geq 0$ for $\varepsilon = 0$. In general, we have

$$\frac{d(a_1 - \sqrt{2}a_2)}{d\bar{a}}_{|\bar{a}=0} = \frac{3}{\sqrt{2}}\varepsilon. \tag{5.49}$$

A similar but more involved argument can be made to show that the derivative has at most one root for $\bar{a} > 0$ and arbitrary but fixed $\varepsilon \in [0, \frac{1}{4}]$. Then from (5.47) and (5.49) follows that $h(\tau) \geq 0$ for $\varepsilon \in [0, \frac{1}{4}]$ and $a > 0$.

For $a \in]-1, 0[$, we cannot choose $\mathring{\mu}^2 = \varepsilon a$. Instead, we set $\mathring{\mu}^2 = \varepsilon \in [0, 1]$ and note that $a\tau \in]-1, 0[$. Then

$$\begin{aligned} h(\tau) &= a_1\left(a\tau(a+2) + a\tau + 2\right) - a_2(a\tau+2)^{3/2} \\ &\geq a_1\left(a\tau(a+2) + a\tau + 2\right) - a_4(a\tau+2)^{3/2} \end{aligned}$$

with

$$a_4 := 1 - \frac{1}{(a\tau+1)^{3/2}} - \frac{4\sqrt{2}\chi e^{-\chi/2}}{(a\tau+2)^{3/2}}$$

due to $a < 1$. If $a_1, a_4 \leq 0$, then due to $a\tau \in]-1, 0[$, we have

$$\begin{aligned} &a_1\left(a\tau+2\right) + \underbrace{a_1 a\tau}_{\geq 0}\underbrace{(a+2)}_{\geq 1} - a_4(a\tau+2)^{3/2} \geq a_1\left(a\tau+2\right) + a_1 a\tau - a_4(a\tau+2)^{3/2} \\ &= 2a_1(a\tau+1) \underbrace{-a_4}_{\geq 0}\underbrace{(a\tau+2)^{3/2}}_{\geq 2\sqrt{2}(a\tau+1)} \geq 2(a\tau+1)\left(a_1 - \sqrt{2}a_4\right). \end{aligned}$$

Note that $(a\tau+2)^{3/2} \geq 2\sqrt{2}(a\tau+1)$ only holds for $a\tau \in]-1,0[$. We only show $a_1-\sqrt{2}a_4 \geq 0$ and $a_1 \leq 0$, since $a_4 \leq 0$ then follows. Regarding the latter, with $\mathring{\mu}^2 = \varepsilon$, we have

$$\begin{aligned} a_1 &= 2\sqrt{2}e^{-\chi/2} - \left(\frac{a\tau+2}{a\tau+1}\right)^{3/2} = 2\sqrt{2}\exp\left(-\frac{\varepsilon\tau}{2(a\tau+2)}\right) - \left(\frac{a\tau+2}{a\tau+1}\right)^{3/2} \\ &\leq 2\sqrt{2} - \left(\frac{a\tau+2}{a\tau+1}\right)^{3/2} \leq 0. \end{aligned}$$

In the last step, we prove $a_1 - \sqrt{2}a_4 \geq 0$:

$$\begin{aligned} &a_1 - \sqrt{2}a_4 = \\ &\quad 2\sqrt{2}\exp\left(-\frac{\varepsilon\tau}{2(a\tau+2)}\right)\left(1+\frac{2\sqrt{2}\varepsilon\tau}{(2+a\tau)^{5/2}}\right) - \left(\frac{a\tau+2}{a\tau+1}\right)^{3/2} - \sqrt{2}\left(1-\frac{1}{(a\tau+1)^{3/2}}\right). \end{aligned}$$

One can set $\varepsilon = -a \in]0,1[$ and use $\bar{a} = a\tau$ to obtain a function depending only on one variable and prove the assertion directly. An alternative approach is to show that $a_1 - \sqrt{2}a_4$ is monotonously decreasing in ε for $\varepsilon \in]0,1[$, which is easy to show. Recall that this property did not hold in case of $a > 1$. Consequently, it suffices to consider $\varepsilon = 0$, for which

$$(a_1 - \sqrt{2}a_4)_{|\varepsilon=0} = (a_1 - \sqrt{2}a_3)_{|\varepsilon=0}.$$

In particular, we can use (5.48). Since the derivative is negative for $\bar{a} < 0$ and the first equation in (5.47) holds, we have $h(\tau) \geq 0$ for $a < 0$. □

Economic MPC – Linear Control

6

The results of the comprehensive analysis in the last two chapters were limited to stabilizing MPC, cf. Section 3.2. The stage cost had the structure of (3.4), i.e., we penalized the distance of the state to a desired equilibrium and of the control to the corresponding control value. In this chapter we consider the more general stage cost (3.5), in which the effort of the control rather than its distance to the—in general difficult to compute—equilibrium control value is penalized. As a result, the closed-loop system should converge to an equilibrium that gives the best trade-off between minimizing the tracking error and the control effort. This is a particular instance of an economic MPC scheme, for which we have argued in Section 3.3 that strict dissipativity of the underlying optimal control problem is the key property for stability and near optimal performance of the closed loop.

For this reason, in this chapter we investigate strict dissipativity for a class of optimal control problems for probability density functions. In order to make the analysis feasible, we again restrict ourselves to linear SDE dynamics governed by the Ornstein–Uhlenbeck process, linear feedback controllers, and Gaussian PDFs. For this setting, motivated by [25, 24], we first explore the opportunities and limitations of obtaining strict dissipativity with a linear storage function, before proposing a nonlinear storage function, which also works for parameter values in which the linear storage function approach fails.

In order to keep the PDE aspect of the problem and make the setting extendable to more complicated dynamics, we first keep the L^2 norm in the cost function. We then extend the analysis to alternative cost functions including the Wasserstein distance W^2. The linear Gaussian setting allows us to compare our results with general purpose cost functions—such as the L^2 or the W^2 cost—to results for a cost function that is particularly tailored to the linear Gaussian setting. In the latter cost function we combine the 2-norm for the mean and the Frobenius norm for the covariance matrix of the Gaussian PDF and thus term it 2F cost. Despite its similarity to the W^2 cost, the results on strict dissipativity are strikingly different for these two cost functions, which is just one important result of this chapter.

This chapter is organized as follows. Section 6.1 introduces the problem and the cost functions under consideration. Section 6.2 collects a few auxiliary results. Our main results concerning strict dissipativity for the L^2 cost, the W^2 cost, and the 2F cost are presented in Section 6.3. We end the chapter—and our analysis of the MPC closed loop—with concluding remarks in Section 6.4.

6.1 Problem Setting

Similar to Chapter 5 we study linear controlled stochastic processes

$$dX_t = (A - BK(t))\, X_t dt + Bc(t)dt + DdW_t, \quad t \in]0, T[,$$

with an initial condition $\mathring{X} \in \mathbb{R}^d$ that is normally distributed, i.e., $\mathring{X} \sim \mathcal{N}(\mathring{\mu}, \mathring{\Sigma})$ with initial mean $\mathring{\mu} \in \mathbb{R}^d$ and covariance matrix $\mathring{\Sigma} \in \mathbb{R}^{d\times d} > 0$. As before, we replace the Fokker–Planck equation by the following system of ODEs for μ and Σ:

$$\begin{aligned} \dot{\mu}(t) &= (A - BK(t))\mu(t) + Bc(t), & \mu(0) &= \mathring{\mu}, \\ \dot{\Sigma}(t) &= (A - BK(t))\Sigma(t) + \Sigma(t)(A - BK(t))^\top + DD^\top, & \Sigma(0) &= \mathring{\Sigma}. \end{aligned} \tag{6.1}$$

Using this ODE system will enable us to analyze strict dissipativity for the optimal control problem we will introduce soon. Particularly, we will carry out the analysis in this chapter for the linearly controlled Ornstein–Uhlenbeck process, which, as in Chapter 5, is defined by

$$dX_t = -\left(\theta + K(t)\right) X_t dt + c(t)dt + \varsigma dW_t, \quad t \in]0, T[, \tag{6.2}$$

with an initial condition $\mathring{X} \sim \mathcal{N}(\mathring{\mu}, \mathring{\Sigma})$, parameters $\theta, \varsigma > 0$ and control constraints

$$0 < \theta + K(t) =: K_\theta(t), \quad t \geq 0. \tag{6.3}$$

Plugging $A - BK(t) = -K_\theta(t) \in \mathbb{R}_{>0}$ and $D = \varsigma \in \mathbb{R}_{>0}$ into (6.1) results in

$$\dot{\mu}(t) = -K_\theta(t)\mu(t) + c(t), \quad \mu(0) = \mathring{\mu}, \tag{6.4a}$$

$$\dot{\Sigma}(t) = -2K_\theta(t)\Sigma(t) + \varsigma^2, \quad \Sigma(0) = \mathring{\Sigma}. \tag{6.4b}$$

The reason to further consider this particular process is its simple, but bilinear structure. The considered systems (6.1) and (6.4) in particular represent a small fraction of problem classes for which MPC is known to yield good numerical results. However, following the common practice in systems and control theory, we first look at these arguably simpler systems, which are more amenable to a rigorous mathematical analysis. Although simplified, we expect several benefits from our study for more general settings: cost functions that turn out to work well for the bilinear problem might also perform well for more general nonlinear problems. Conversely, cost functions that do not perform well for these simpler systems are likely to perform poorly for more general nonlinear problems, as well. Finally, the results for the bilinear case provide the basis for obtaining local nonlinear results via bilinearization.

For the sake of better comparability to [25, 24], in which dissipativity of linear discrete-time dynamics was considered, we would like to keep the bilinear structure in the discrete time setting. Moreover, in any numerical implementation of MPC the dynamics must be approximated by a numerical scheme. In order to allow for a fast computation of the optimal open-loop trajectories, in MPC implementations simple but less accurate schemes are often preferred to more expensive high-order methods. For these reasons, we perform our analysis for the forward Euler approximation of the ODE system (6.4). This discretization both maintains the bilinear structure and defines a scheme that is frequently used in practice. It is given by

$$\mu^+ = \mu + T_s\left(-K_\theta\mu + c\right), \qquad \mu(0) = \mathring{\mu}, \tag{6.5a}$$

$$\Sigma^+ = \Sigma + T_s\left(-2K_\theta\Sigma + \varsigma^2\right), \quad \Sigma(0) = \mathring{\Sigma}. \tag{6.5b}$$

Remark 6.1. *Note that the state constraint* $\Sigma > 0$ *automatically holds for* (6.1) *and* (6.4). *However, when switching to the Euler approximation* (6.5), *we have to impose* $\Sigma(k) > 0$ *as a constraint for all* $k \in \mathbb{N}_0$. *In conjunction with* $K_\theta(k) > 0$, *cf.* (6.3), *this can be incorporated as control constraints*

$$0 < K_\theta(k) < (\Sigma(k) + T_s\varsigma^2)/(2T_s\Sigma(k)). \tag{6.6}$$

The optimal control problem ($\mathrm{OCP_N}$) that is solved in the MPC algorithm then is

$$\begin{aligned} J_N^\mu((\mathring{\mu}, \mathring{\Sigma}), (\mathbf{K}, \mathbf{c})) := \sum_{k=0}^{N-1} \ell((\mu(k), \Sigma(k)), (K(k), c(k))) \to \min! \\ \text{subject to (6.5), (6.6),} \end{aligned} \tag{6.7}$$

where, as usual, we denote the control sequence—$(\mathbf{K}, \mathbf{c})$ in this case—in bold. The stage cost ℓ is of type (3.5) and is specified next.

For the control cost, similar to Chapter 5, we use the Frobenius norm for K and the Euclidean norm for c, which fit well together. For the state cost we consider three options. The first possibility is to penalize the distance between the current probability density function ρ and the desired PDF $\bar{\rho}$ in the L^2 norm, i.e., $\frac{1}{2}\|\rho - \bar{\rho}\|^2_{L^2(\mathbb{R}^d)}$. This is the standard norm used in costs for optimal control problems governed by parabolic PDEs [95]. In terms of Σ and μ, this yields

$$\begin{aligned} \ell^\mu_{L^2}(\mu, \Sigma, K, c) := 2^{-d-1}\pi^{-d/2} \Big[|\Sigma|^{-1/2} + |\bar{\Sigma}|^{-1/2} \\ -2 \left| \frac{1}{2}(\Sigma + \bar{\Sigma}) \right|^{-1/2} \exp\left(-\frac{1}{2} (\mu - \bar{\mu})^\top (\Sigma + \bar{\Sigma})^{-1} (\mu - \bar{\mu}) \right) \Big] + \frac{\gamma}{2}\|K\|_F^2 + \frac{\gamma}{2}\|c\|_2^2, \end{aligned} \tag{6.8}$$

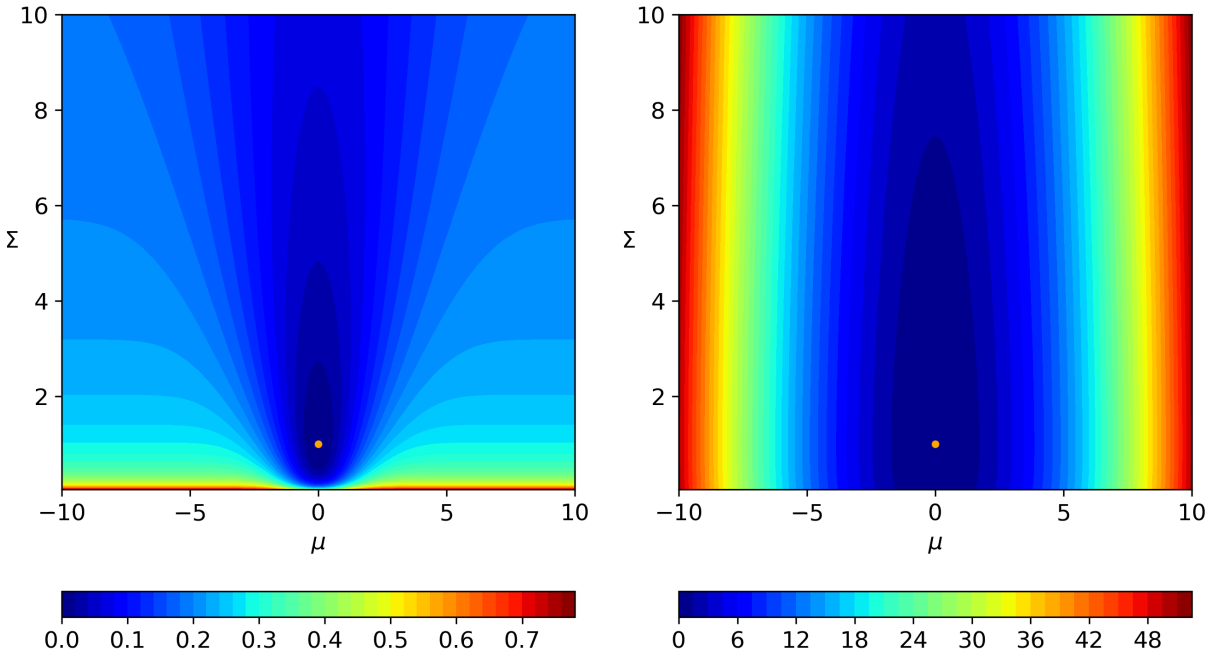
cf. Lemma 5.5. Looking at the cost from the ODE perspective, the L^2 penalization does not seem standard or intuitive at all. One alternative is to use the Wasserstein metric, which is specifically designed to measure the distance between two PDFs. For the general definition of this metric we refer to [44]. Here we only use the formula for the Wasserstein metric for normal distributions derived in [44]. In case Σ and $\bar{\Sigma}$ commute—which does not limit our analysis since w.l.o.g. we can restrict ourselves to $\bar{\Sigma} = I$, see Section 6.2—this formula yields the following stage cost:

$$\ell^\mu_{W^2}(\mu, \Sigma, K, c) := \frac{1}{2}\|\mu - \bar{\mu}\|_2^2 + \frac{1}{2}\left\|\Sigma^{1/2} - \bar{\Sigma}^{1/2}\right\|_F^2 + \frac{\gamma}{2}\|K\|_F^2 + \frac{\gamma}{2}\|c\|_2^2. \tag{6.9}$$

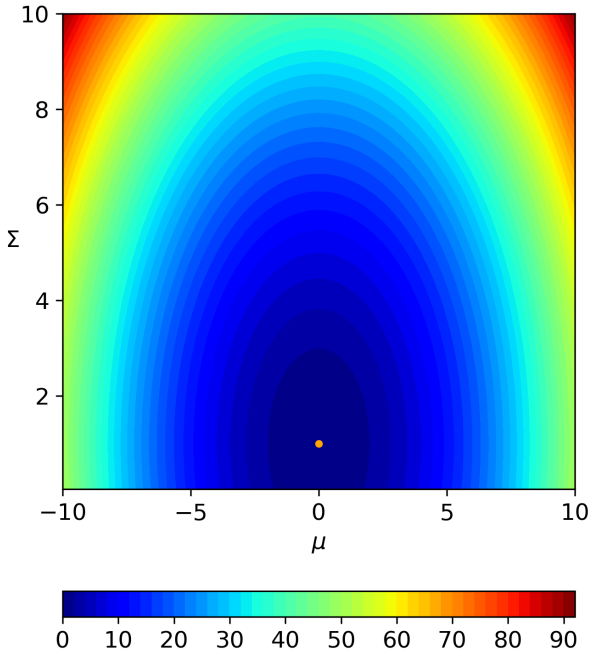
The third option we discuss in this chapter is very similar to the Wasserstein distance from (6.9). The only difference is to consider Σ and $\bar{\Sigma}$ instead of $\Sigma^{1/2}$ and $\bar{\Sigma}^{1/2}$, respectively. Thus, we end up with

$$\ell^\mu_{2F}(\mu, \Sigma, K, c) := \frac{1}{2}\|\mu - \bar{\mu}\|_2^2 + \frac{1}{2}\left\|\Sigma - \bar{\Sigma}\right\|_F^2 + \frac{\gamma}{2}\|K\|_F^2 + \frac{\gamma}{2}\|c\|_2^2. \tag{6.10}$$

This form of the cost function is commonly used in optimization of systems governed by ODE systems. The index used in the notation for this cost, *2F*, indicates the combination of Euclidean and Frobenius norm in the state penalization (and, coincidentally also the control penalization). In the special case $\bar{\Sigma} = I$ we have that $\ell^\mu_{W^2}(\mu, \Sigma^2, K, c) =$



(a) L^2 cost. (b) W^2 cost.



(c) 2F cost.

Figure 6.1: The state cost parts of the three stage costs $\ell^{\mu}_{L^2}(\mu, \Sigma, K, c)$, $\ell^{\mu}_{W^2}(\mu, \Sigma, K, c)$, and $\ell^{\mu}_{2F}(\mu, \Sigma, K, c)$, i.e., (6.8), (6.9), and (6.10) for $\gamma = 0$, respectively. The desired state was set to $(\bar{\mu}, \bar{\Sigma}) = (0, 1)$. The orange dot in the respective plots marks the minimum.

$\ell^{\mu}_{2F}(\mu, \Sigma, K, c)$, i.e., considering the squared covariance matrix Σ^2 instead of Σ in the W^2 cost leads to the 2F cost. All three stage costs—minus the control cost—are illustrated in Figure 6.1.

To prove that (6.7) is strictly dissipative, we need to find a suitable storage function λ

for which the inequality (3.10) in Definition 3.8 holds. In general, it is not easy to find such a function. However, for OCPs with linear discrete-time dynamics

$$z^+ = Az + Bu + c =: f^l(z,u),$$

a convex constraint set, and strictly convex stage cost ℓ, it is known [25] that the linear function

$$\lambda^l(z) := \bar{\lambda}^\top z \tag{6.11}$$

is a suitable storage function; for a proof, see, e.g., [24].[1] Here, $\bar{\lambda} \in \mathbb{R}^d$ is the Lagrange multiplier in the optimization problem consisting of finding the optimal equilibrium (z^e, u^e):

$$\min_{(z,u)} \ell(z,u) \quad \text{s.t. } z = f^l(z,u). \tag{6.12}$$

The reason for this linear storage function is the close connection between the Lagrange function $L(z,u,\lambda)$ associated to (6.12) and the resulting modified cost $\tilde{\ell}$, cf. Definition 3.8:

$$\begin{aligned}\tilde{\ell}(z,u) &= \ell(z,u) - \ell(z^e,u^e) + \lambda^l(z) - \lambda^l(f^l(z,u)) \\ &= \ell(z,u) - \ell(z^e,u^e) + \bar{\lambda}^\top \left(z - f^l(z,u)\right) \\ &= L(z,u,\bar{\lambda}) - \ell(z^e,u^e).\end{aligned} \tag{6.13}$$

In this particular form of dissipativity, also known as strict duality in optimization theory, the (strict) convexity of ℓ carries over to L and therefore to $\tilde{\ell}$, with the global minimum being attained at (z^e,u^e). In the final step, due to $L(z^e,u^e,\bar{\lambda}) = \ell(z^e,u^e)$, we have that $\tilde{\ell}$ is positive definite with respect to the optimal equilibrium (z^e,u^e), which allows to conclude (3.10), i.e., strict dissipativity.

Although this is, in general, not true for nonlinear $f(z,u)$, in the following, we analyze how far the approach of a linear storage function can be successfully extended to bilinear OCPs, such as (6.7) with stage cost ℓ given by (6.8) or (6.9) or (6.10). To this end, in the next section, we state some auxiliary results.

6.2 Auxiliary Results Regarding Dissipativity

In a first step, we characterize equilibria of the one-dimensional (discretized) Ornstein–Uhlenbeck process. We recall that the imposed constraints (6.6) ensure $\bar{K}_\theta = \theta + \bar{K} > 0$.

Lemma 6.2. *The set of equilibria is identical for* (6.4) *and* (6.5) *and is given by*

$$\mathcal{E} := \left\{ (\bar{\mu}, \bar{\Sigma}, \bar{K}, \bar{c}) \mid \bar{\mu} = \frac{\bar{c}}{\bar{K}_\theta}, \bar{\Sigma} = \frac{\varsigma^2}{2\bar{K}_\theta} \right\}. \tag{6.14}$$

The proof is obvious; we merely note that the additional constraint (6.6) holds for $\bar{\Sigma} = \varsigma^2/(2\bar{K}_\theta)$.

Without loss of generality, we assume that $(\bar{\mu}, \bar{\Sigma}) = (0,1)$. Otherwise we introduce a new random variable $Y_t := \bar{\Sigma}^{-1/2}(X_t - \bar{\mu})$ and get a new ODE system similar to (6.4). With this assumption, due to (6.14), we have $\bar{c} = 0$, which allows us to further simplify the dynamics under consideration for the chosen cost criteria.

[1] One can ensure the boundedness from below that is required in Definition 3.8 by state constraints.

Lemma 6.3. *Assume that* $(\bar{\mu}, \bar{\Sigma}) = (0, 1)$. *Then the OCP* (6.7) *with* ℓ *given by* (6.8), *or* (6.9) *or* (6.10) *is strictly dissipative at the equilibrium* $(0, \bar{\Sigma}, \bar{K}, 0)$ *if and only if the OCP*

$$J_N(\mathring{\Sigma}, \mathbf{K}) := \sum_{k=0}^{N-1} \ell((0, \Sigma(k)), (K(k), 0)) \to \min! \quad \text{subject to (6.5b), (6.6),} \tag{6.15}$$

with the same ℓ *is strictly dissipative at the equilibrium* $(\bar{\Sigma}, \bar{K})$.

Proof. First, if $(\bar{\Sigma}, \bar{K})$ is an equilibrium of (6.5b), then $(0, \bar{\Sigma}, \bar{K}, 0)$ is an equilibrium of (6.5) and vice versa.

Assuming strict dissipativity of (6.15) at $(\bar{\Sigma}, \bar{K})$ with ℓ given by (6.8), (6.9), or (6.10) and defining $\tilde{\lambda}(z_1, z_2) := \lambda(z_2)$, we get

$$\begin{aligned} \varrho(|\Sigma|_{\bar{\Sigma}}) &\le \ell(0, \Sigma, K, 0) - \ell(0, \bar{\Sigma}, \bar{K}, 0) + \lambda(\Sigma) - \lambda(\Sigma^+) \\ &\le \ell(\mu, \Sigma, K, c) - \ell(0, \bar{\Sigma}, \bar{K}, 0) + \lambda(\Sigma) - \lambda(\Sigma^+) \\ &= \ell(\mu, \Sigma, K, c) - \ell(0, \bar{\Sigma}, \bar{K}, 0) + \tilde{\lambda}(\mu, \Sigma) - \tilde{\lambda}(\mu^+, \Sigma^+). \end{aligned}$$

Thus, (6.7) is strictly dissipative at $(0, \bar{\Sigma}, \bar{K}, 0)$ with storage function $\tilde{\lambda}$.

Conversely, assuming that (6.7) is strictly dissipative at an equilibrium $(0, \bar{\Sigma}, \bar{K}, 0)$,

$$\varrho(|(\mu, \Sigma)|_{(0,\bar{\Sigma})}) \le \ell(\mu, \Sigma, K, c) - \ell(0, \bar{\Sigma}, \bar{K}, 0) + \lambda(\mu, \Sigma) - \lambda(\mu^+, \Sigma^+)$$

holds for all admissible (μ, Σ, K, c) and some storage function λ. In particular, it holds for $(\mu, c) = (0, 0)$, i.e.,

$$\begin{aligned} &\ell(0, \Sigma, K, 0) - \ell(0, \bar{\Sigma}, \bar{K}, 0) + \lambda(0, \Sigma) - \lambda(f(0, \Sigma, K, 0)) \\ &= \ell(0, \Sigma, K, 0) - \ell(0, \bar{\Sigma}, \bar{K}, 0) + \lambda(0, \Sigma) - \lambda(0, \Sigma^+) \ge \varrho(|(0, \Sigma)|_{(0,\bar{\Sigma})}) = \varrho(|\Sigma|_{\bar{\Sigma}}), \end{aligned}$$

where $f(\mu, \Sigma, K, c)$ is defined by μ^+ and Σ^+ in (6.5). □

Thus, in the following, we only need to examine whether (6.15) is strictly dissipative with the respective stage cost. In this setting the three different stage cost functions under consideration—(6.8), (6.9), and (6.10)—can be simplified to

$$\ell_{L^2}(\Sigma, K) := \ell^{\mu}_{L^2}(0, \Sigma, K, 0) = \frac{1}{4\sqrt{\pi}} \left[\Sigma^{-1/2} + 1 - 2\sqrt{2}(\Sigma + 1)^{-1/2}\right] + \frac{\gamma}{2} K^2, \tag{6.16}$$

$$\ell_{W^2}(\Sigma, K) := \ell^{\mu}_{W^2}(0, \Sigma, K, 0) = \frac{1}{2}\left(\sqrt{\Sigma} - 1\right)^2 + \frac{\gamma}{2} K^2, \tag{6.17}$$

$$\ell_{2F}(\Sigma, K) := \ell^{\mu}_{2F}(0, \Sigma, K, 0) = \frac{1}{2}(\Sigma - 1)^2 + \frac{\gamma}{2} K^2, \tag{6.18}$$

respectively. We conclude this section with some auxiliary statements about optimal equilibria.

Lemma 6.4. *Let* (Σ^e, K^e) *be an optimal equilibrium for one of the stage cost functions* $\ell_{L^2}(\Sigma, K)$, $\ell_{W^2}(\Sigma, K)$, *or* $\ell_{2F}(\Sigma, K)$. *Then*

$$\begin{cases} K^e \in [0, \frac{\varsigma^2}{2} - \theta] \text{ and } \Sigma^e \in [1, \frac{\varsigma^2}{2\theta}], & \text{if } \varsigma^2/2 - \theta > 0, \\ K^e \in [\frac{\varsigma^2}{2} - \theta, 0] \text{ and } \Sigma^e \in [\frac{\varsigma^2}{2\theta}, 1], & \text{if } \varsigma^2/2 - \theta < 0, \\ K^e = 0 \text{ and } \Sigma^e = 1, & \text{if } \varsigma^2/2 - \theta = 0. \end{cases} \tag{6.19}$$

Proof. From (6.14) we know that

$$\Sigma^e = \frac{\varsigma^2}{2(\theta + K^e)}, \tag{6.20}$$

which is monotonically decreasing in K^e. Moreover,

$$\Sigma^e = 1 \quad \Leftrightarrow \quad K^e = \frac{\varsigma^2}{2} - \theta, \tag{6.21}$$

which proves the assertion in the case $\varsigma^2/2 - \theta = 0$. For the remaining two cases, we first note that all three stage costs $\ell_{L^2}(\Sigma, K)$, $\ell_{2F}(\Sigma, K)$, and $\ell_{W^2}(\Sigma, K)$ are minimal with respect to Σ at $\Sigma = 1$ and increase the further away Σ is from the target value 1. While obvious for the two convex functions $\ell_{2F}(\Sigma, K)$ and $\ell_{W^2}(\Sigma, K)$, the same holds for the non-convex stage cost $\ell_{L^2}(\Sigma, K)$:

$$\partial_\Sigma \ell_{L^2}(\Sigma, K) = \frac{(2\Sigma)^{3/2} - (\Sigma + 1)^{3/2}}{2\Sigma^{3/2}(\Sigma + 1)^{3/2}} \begin{cases} > 0, & \text{if } \Sigma > 1, \\ = 0, & \text{if } \Sigma = 1, \\ < 0, & \text{if } \Sigma < 1. \end{cases}$$

Let us now assume that $\frac{\varsigma^2}{2} - \theta > 0$. Then $K^e \geq 0$, since any $K_1 < 0$ is more expensive than $K_2 = 0$ due to $K_1^2 > K_2^2$ and $\Sigma_1 = \frac{\varsigma^2}{2(\theta + K_1)} > \Sigma_2 = \frac{\varsigma^2}{2\theta} > 1$, i.e., Σ_1 induces a higher cost than Σ_2. Moreover, $K^e \leq \frac{\varsigma^2}{2} - \theta$ since some $K_3 > \frac{\varsigma^2}{2} - \theta$ is always more costly than $K_4 := \frac{\varsigma^2}{2} - \theta$ due to $K_3^2 > K_4^2$ and the corresponding state $\Sigma_3 = \frac{\varsigma^2}{2(\theta + K_3)} \neq 1$ inducing additional cost while $\Sigma_4 = 1$ does not.

The case $\frac{\varsigma^2}{2} - \theta < 0$ is analogous. □

We note that for the three stage costs $\ell_{L^2}(\Sigma, K)$, $\ell_{W^2}(\Sigma, K)$, and $\ell_{2F}(\Sigma, K)$ considered here, parameters satisfying $\varsigma^2/2 - \theta = 0$ correspond to stabilizing MPC. In this case, the respective OCPs are strictly dissipative with storage function $\lambda \equiv 0$, cf. Remark 3.9. Hence, that case is excluded in the ensuing analysis.

6.3 Results on Strict Dissipativity

In Section 6.2 we simplified the OCP under consideration, (6.7), by finding an equivalent formulation, (6.15), which is sufficient for analyzing dissipativity. This section is dedicated to the dissipativity analysis of the OCP (6.15) for the L^2 cost (6.16), the W^2 cost (6.17), and the 2F cost (6.18). Throughout this section, the pair (Σ^e, K^e) denotes an *optimal* equilibrium, cf. Definition 3.6, i.e., a solution of

$$\min_{(\Sigma, K)} \ell(\Sigma, K) \quad \text{s.t. } \Sigma - f(\Sigma, K) = 0, \tag{6.22}$$

where $\ell(\Sigma, K)$ is one of the three stage costs $\ell_{L^2}(\Sigma, K)$, $\ell_{W^2}(\Sigma, K)$, or $\ell_{2F}(\Sigma, K)$, depending on the respective subsection.

As mentioned at the end of Section 6.1, we cannot directly apply the dissipativity results from [25, 24] because our dynamics are not linear but bilinear. In particular, convexity of the stage cost ℓ does not necessarily carry over to the modified cost $\tilde{\ell}$, cf. Definition 3.8. Hence, approaches that rely on linearity and/or on convexity in general yield

only local results regarding strict dissipativity in our setting. One such approach that is currently gaining popularity [78, 94, 93, 51, 52] relies on *Pontryagin's Maximum Principle* and the corresponding Hamiltonian optimality system or the corresponding Riccati equation. In these references, instead of showing strict dissipativity, the authors prove certain stability properties of the optimally controlled system, typically in the form of the so-called *turnpike property* that is mentioned at the end of Section 3.3. We recall that the turnpike property and strict dissipativity are closely related, see Figure 3.4. However, again, due to bilinearity and not necessarily convex objective functions, in our setting we would only get this property locally. Yet, for the analysis of MPC schemes we require these properties to hold globally, i.e., global strict dissipativity. Thus, we will perform an ad-hoc analysis for the three different stage costs, which includes a convexity analysis of the respective modified stage costs $\tilde{\ell}$ and—if $\tilde{\ell}$ is not convex—a closer look at stationary points and boundary values of $\tilde{\ell}$. With the structural insight we gain from these computations we are not only able to identify settings where strict dissipativity with a linear storage function can be shown, but also to provide alternative, nonlinear storage functions to prove strict dissipativity in settings where a linear storage function cannot be used for that purpose.

6.3.1 L^2 cost

In this section, we consider the OCP (6.15) to which we have reduced the original problem (6.7). Overall, the optimization problem is given by

$$\begin{aligned} J_N(\mathring{\Sigma}, \mathbf{K}) &:= \sum_{k=0}^{N-1} \ell_{L^2}(\Sigma(k), K(k)) \\ &= \sum_{k=0}^{N-1} \left[\frac{1}{4\sqrt{\pi}} \left[\Sigma(k)^{-1/2} + 1 - 2\sqrt{2}(\Sigma(k)+1)^{-1/2} \right] + \frac{\gamma}{2} K(k)^2 \right] \to \min_{\mathbf{K}}! \qquad (6.23) \\ \text{s.t.} \quad \Sigma^+ &= \Sigma + T_s \left(-2K_\theta \Sigma + \varsigma^2 \right) =: f(\Sigma, K), \\ \Sigma(0) &= \mathring{\Sigma}, \\ 0 &< K_\theta(k) < (\Sigma(k) + T_s\varsigma^2)/(2T_s\Sigma(k)), \quad k \in \{0, ..., N-1\}. \end{aligned}$$

For the linear storage function $\lambda^l(z)$, the modified cost $\tilde{\ell}_{L^2}(\Sigma, K)$, cf. Definition 3.8, reads

$$\begin{aligned} \tilde{\ell}_{L^2}(\Sigma, K) := \frac{1}{4\sqrt{\pi}} \left[\Sigma^{-1/2} + 1 - 2\sqrt{2}(\Sigma+1)^{-1/2} \right] + \frac{\gamma}{2} K^2 \\ - \ell_{L^2}(\Sigma^e, K^e) + \bar{\lambda} \left(-T_s(-2(\theta+K)\Sigma + \varsigma^2) \right), \end{aligned}$$

where we recall that $K + \theta = K_\theta$. The Lagrange function associated to the problem of finding the optimal equilibrium, (6.22), reads

$$\begin{aligned} L_{L^2}(\Sigma, K, \lambda) &:= \ell_{L^2}(\Sigma, K) + \lambda \cdot (\Sigma - f(\Sigma, K)) \qquad (6.24) \\ &= \frac{1}{4\sqrt{\pi}} \left[\Sigma^{-1/2} + 1 - 2\sqrt{2}(\Sigma+1)^{-1/2} \right] + \frac{\gamma}{2} K^2 + \lambda \left(-T_s(-2(\theta+K)\Sigma + \varsigma^2) \right). \end{aligned}$$

In this manner, one obtains the Lagrange multiplier $\bar{\lambda} \in \mathbb{R}$ that corresponds to an optimal equilibrium (Σ^e, K^e). The multiplier $\bar{\lambda} \in \mathbb{R}$ is unique since

$$\nabla \left(\Sigma - f(\Sigma, K) \right) = 2T_s \begin{pmatrix} K_\theta \\ \Sigma \end{pmatrix} \neq 0$$

due to $T_s, K_\theta, \Sigma > 0$. Note the connection between the Lagrange function L_{L^2} and the modified cost $\tilde{\ell}_{L^2}$, cf. (6.13). In particular, for the uniquely defined $\bar{\lambda} \in \mathbb{R}$, (Σ^e, K^e) is a stationary point of $\tilde{\ell}_{L^2}$. However, $\tilde{\ell}_{L^2}$ might exhibit additional stationary points, which will be of interest in the following, since a necessary condition for strict dissipativity at (Σ^e, K^e) is that this equilibrium is the unique global minimum of the modified cost $\tilde{\ell}_{L^2}(\Sigma, K)$. These additional stationary points need to be checked for admissibility, i.e., whether they satisfy the state and control constraints, while for optimal equilibria (Σ^e, K^e), these constraints are always automatically satisfied, see Lemma 6.4. We count the stationary points of $\tilde{\ell}_{L^2}$ for a fixed $\bar{\lambda}$ next. To this end, we introduce the notation

$$Z := 2\bar{\lambda} T_s,$$

which we will use throughout this section. The gradient $\nabla \tilde{\ell}_{L^2}(\Sigma, K)$ is then given by

$$\nabla \tilde{\ell}_{L^2}(\Sigma, K) = \begin{pmatrix} \left(-\Sigma^{-3/2} + 2\sqrt{2}(\Sigma+1)^{-3/2}\right)/(8\sqrt{\pi}) \\ \gamma K \end{pmatrix} + Z \begin{pmatrix} \theta + K \\ \Sigma \end{pmatrix}. \tag{6.25}$$

Proposition 6.5. *For a fixed $\bar{\lambda}$ and thus fixed Z the modified cost $\tilde{\ell}_{L^2}(\Sigma, K)$ has at most two admissible stationary points. If $Z = 0$, then only one admissible stationary point of $\tilde{\ell}_{L^2}(\Sigma, K)$ exists and it is given by $(\Sigma^e, K^e) = (1, 0)$.*

Proof. From $\nabla \tilde{\ell}_{L^2}(\Sigma, K) = 0$, cf. (6.25), we get $K = -Z\Sigma/\gamma$. Thus,

$$0 = \frac{1}{8\sqrt{\pi}} \left(-\frac{1}{\Sigma^{3/2}} + \frac{2\sqrt{2}}{(\Sigma+1)^{3/2}} \right) + Z \left(\theta - \frac{Z\Sigma}{\gamma} \right) =: h(\Sigma).$$

If $Z = 0$, then $K = 0 = K^e$ and hence $h(\Sigma) = 0 \Leftrightarrow \Sigma = 1 = \Sigma^e$, i.e., $(\Sigma^e, K^e) = (1, 0)$ is the unique admissible stationary point of $\tilde{\ell}_{L^2}(\Sigma, K)$.

Let $Z \neq 0$. If $h(\Sigma)$ has a unique admissible stationary point, then only up to two admissible solutions for $h(\Sigma) = 0$ can exist, i.e., the assertion follows. To this end, we look at the first two derivatives of h:

$$\begin{aligned} h'(\Sigma) &= 3/\left(16\sqrt{\pi}\right) \left(\Sigma^{-5/2} - 2\sqrt{2}(\Sigma+1)^{-5/2} \right) - Z^2/\gamma, \\ h''(\Sigma) &= 15/\left(32\sqrt{\pi}\right) \left(-\Sigma^{-7/2} + 2\sqrt{2}(\Sigma+1)^{-7/2} \right). \end{aligned}$$

It is easily seen that

$$h''(\Sigma) \begin{cases} < 0, & \Sigma < \Sigma^{**} \\ = 0, & \Sigma = \Sigma^{**} \\ > 0, & \Sigma > \Sigma^{**} \end{cases} \quad \text{and} \quad h'(\Sigma) \begin{cases} > -Z^2/\gamma, & \Sigma < \Sigma^{*} \\ = -Z^2/\gamma, & \Sigma = \Sigma^{*} \\ < -Z^2/\gamma, & \Sigma > \Sigma^{*} \end{cases},$$

where $\Sigma^{**} := \frac{2^{4/7}}{2-2^{4/7}} \approx 2.89$ and $\Sigma^{*} := \frac{2^{2/5}}{2-2^{2/5}} \approx 1.94$. In particular, $h'(\Sigma) < 0$ for $\Sigma > \Sigma^*$. Therefore, stationary points of $h(\Sigma)$ can only exist for $\Sigma \in]0, \Sigma^*[$. Since $h''(\Sigma) < 0$ for $\Sigma \leq \Sigma^* < \Sigma^{**}$, at most one stationary point of $h(\Sigma)$ can exist (and it is a local maximum). Due to $h'(\Sigma) \to \infty$ for $\Sigma \searrow 0$, $h'(\Sigma) < 0$ for $\Sigma > \Sigma^*$, and the intermediate value theorem, a stationary point does exist. Thus, there always exists a unique stationary point of $h(\Sigma)$, concluding the proof. □

From the Hessian

$$\nabla^2 \tilde{\ell}_{L^2}(\Sigma, K) = \begin{pmatrix} \frac{3}{16\sqrt{\pi}} \left(\frac{1}{\Sigma^{5/2}} - \frac{2\sqrt{2}}{(\Sigma+1)^{5/2}} \right) & Z \\ Z & \gamma \end{pmatrix} \tag{6.26}$$

it is obvious that for any fixed $Z \neq 0$, $\tilde{\ell}_{L^2}$ is not convex for sufficiently large Σ. This prevents us from easily deducing strict dissipativity, as opposed to the case of linear dynamics, in which case the Hessian $\nabla^2 \tilde{\ell}$ is constant. The Hessian (6.26), however, is not constant. Hence, even if $\nabla^2 \tilde{\ell}(\Sigma^e, K^e)$ is positive definite, then we can only conclude local convexity near (Σ^e, K^e), which implies strict dissipativity if state and control are constrained to a neighborhood of (Σ^e, K^e). To make statements about global strict dissipativity, we take a closer look at the structure of $\tilde{\ell}_{L^2}$. Based on Lemma 6.4 we consider the two cases $\varsigma^2/2 - \theta > 0$ and $\varsigma^2/2 - \theta < 0$ separately.

The case $\varsigma^2/2 - \theta > 0$

For a large set of parameters, (strict) dissipativity does not hold with a linear storage function, see the following proposition.

Proposition 6.6. *If $\varsigma^2/2 - \theta > 0$, then for large enough Σ the OCP (6.23) cannot be dissipative with a linear storage function.*

Proof. The idea of the proof is to show that the modified cost $\tilde{\ell}_{L^2}$ can assume negative values, which violates (3.10).

As $\Sigma \to \infty$, $\tilde{\ell}_{L^2}(\Sigma, K) \to \operatorname{sgn}(Z(K+\theta)) \cdot \infty$. Hence, if $\operatorname{sgn}(Z(K+\theta)) < 0$, then (Σ^e, K^e) cannot be a global minimum, contradicting dissipativity. Since $K + \theta > 0$, only the sign of Z is of importance. Thus, in the rest of the proof, we show that $Z < 0$. From

$$\partial_K L_{L^2}(\Sigma, K, \bar{\lambda}) = \partial_K \tilde{\ell}_{L^2}(\Sigma, K) = \gamma K + Z\Sigma$$

we deduce that

$$\partial_K L_{L^2}(\Sigma, K, \bar{\lambda}) = 0 \quad \Leftrightarrow \quad \begin{cases} \Sigma = -\gamma K/Z, & Z \neq 0 \\ K = 0, & Z = 0 \end{cases}.$$

Due to $\partial_K L_{L^2}(\Sigma^e, K^e, \bar{\lambda}) = 0$, we can exclude $Z = 0$: If $Z = 0$, then $K^e = 0$ and thus $\Sigma^e = 1$ because of $\partial_\Sigma L_{L^2}(\Sigma^e, K^e, \bar{\lambda}) = \partial_\Sigma \tilde{\ell}_{L^2}(\Sigma^e, K^e) = 0$, cf. (6.25). But this contradicts (6.21) since $\varsigma^2/2 - \theta > 0$. Thus, $Z \neq 0$ and we have $\Sigma^e = -\gamma K^e/Z$ and $K^e \neq 0$, which, together with Lemma 6.4, results in $K^e > 0$. Then due to $\gamma > 0$ and $\Sigma^e > 0$ we arrive at $Z < 0$, concluding the proof. □

One might conjecture that strict dissipativity can be recovered by restricting the set of admissible states $\Sigma > 0$. This seems like a promising direction, as we need to restrict the state domain, anyway, to obtain boundedness from below for λ^l. Yet, if $\Sigma^e > 2^{2/5}/\left(2 - 2^{2/5}\right) \approx 1.94$, then from $\nabla^2 \tilde{\ell}_{L^2}(\Sigma^e, K^e)_{11} < 0$ and $\gamma > 0$ we infer that (strict) dissipativity does not hold since the optimal equilibrium (Σ^e, K^e) is not a (local) minimum of $\tilde{\ell}_{L^2}$. Instead, a descent direction exists in (Σ^e, K^e), i.e., $\tilde{\ell}_{L^2}$ can attain negative values since $\tilde{\ell}(\Sigma^e, K^e) = 0$ always holds. Thus, for a large parameter set, this problem persists.

The case $\varsigma^2/2 - \theta < 0$

For $\varsigma^2/2 - \theta < 0$, the asymptotic behavior of $\tilde{\ell}_{L^2}(\Sigma, K)$ for $\Sigma \to \infty$ is not a problem since $Z > 0$.[2] However, in addition to a potential second stationary state, cf. Proposition 6.5, one needs to check the boundaries $\Sigma \searrow 0$ and $K \searrow -\theta$ (the asymptotic behavior of $\tilde{\ell}_{L^2}(\Sigma, K)$ for $K \to \infty$ is never a problem), see the following example.

Example 6.7. *Consider* (6.23) *with the parameters*

$$\varsigma = 9/20, \quad \theta = 13/20, \quad \gamma = 3/5, \quad \textit{and} \quad T_s = 1/10.$$

The optimal equilibrium and the corresponding Lagrangian multiplier are calculated numerically, yielding $\Sigma^e \approx 0.42117895$, $K^e \approx -0.40960337$ *and* $Z \approx 0.5835097$. *The Hessian* $\nabla^2 \tilde{\ell}_{L^2}$ *evaluated at* (Σ^e, K^e),

$$\nabla^2 \tilde{\ell}_{L^2}(\Sigma^e, K^e) \approx \begin{pmatrix} 0.7946167 & Z \\ Z & \gamma \end{pmatrix},$$

is positive definite since $|\nabla^2 \tilde{\ell}_{L^2}(\Sigma^e, K^e)| \approx 0.136 > 0$. *However, at the boundary we find that* $\tilde{\ell}_{L^2}(1, -\theta) \approx -0.00640024 < 0$. *Thus, due to continuity of* $\tilde{\ell}_{L^2}$, *strict dissipativity with a linear storage function does not hold.*

Based on this structural insight, we can identify situations in which a linear storage function does work, cf. Example 6.8.

Example 6.8. *Consider* (6.23) *with the parameters*

$$\varsigma = 1/3, \quad \theta = 7/2, \quad \gamma = 1/4, \quad \textit{and} \quad T_s = 1/10.$$

Then numerical computations yield $\Sigma^e \approx 0.0199205$, $K^e \approx -0.7111341$, *and* $Z \approx 8.9246597$. *The second stationary point of* $\tilde{\ell}_{L^2}$ *is found at approximately*

$$(0.0904564, -3.2291691) =: (\Sigma^s, K^s),$$

with $\tilde{\ell}_{L^2}(\Sigma^s, K^s) \approx 0.45 > 0$. *At the boundary, since* $Z > 0$, $\tilde{\ell}_{L^2}(\Sigma, K) \to \infty$ *for* $\Sigma \to \infty$ *as well as for* $K \to \infty$. *Also,* $\tilde{\ell}_{L^2}(\Sigma, K) \to \infty$ *as* $\Sigma \searrow 0$ *for any fixed admissible* K. *At the remaining boundary* $K = -\theta$ *we have*

$$\tilde{\ell}_{L^2}(\Sigma, -\theta) = \left(\Sigma^{-1/2} + 1 - 2\sqrt{2}(\Sigma + 1)^{-1/2}\right)/(4\sqrt{\pi}) + \frac{\gamma}{2}\theta^2 - \ell_{L^2}(\Sigma^e, K^e) - Z\frac{\varsigma^2}{2},$$

which is minimal at $\Sigma = 1$ *with*

$$\tilde{\ell}_{L^2}(1, -\theta) = \frac{\gamma}{2}\theta^2 - \ell_{L^2}(\Sigma^e, K^e) - Z\frac{\varsigma^2}{2}.$$

For the parameters in this example, this results in $\tilde{\ell}_{L^2}(1, -\theta) \approx 0.2268570 > 0$. *Thus, we can find a function* $\varrho \in \mathcal{K}_\infty$ *such that the dissipativity inequality* (3.10) *holds.*

Examples 6.7 and 6.8 reveal that a case-by-case analysis is needed in order to decide whether strict dissipativity can be established using a linear storage function in the case $\varsigma^2/2 - \theta < 0$.

[2] The proof is analogous to the one of Proposition 6.6.

Modifications to the stage cost ℓ_{L^2}

In this part we propose two modifications to the stage cost ℓ_{L^2} and discuss whether they facilitate the analysis. The first proposal is a scaling of the stage cost.

Remark 6.9 (Scaling of the stage cost)**.** *One could argue that, due to the forward Euler approximation* (6.5)*, the dynamics are effectively scaled by the sampling time T_s, and this scaling should also be applied to the stage cost, i.e., use $T_s \cdot \ell_{L^2}$ instead of ℓ_{L^2}. In that case, instead of* (6.24)*, the Lagrange function reads*

$$L_{L^2}(\Sigma, K, \lambda) = T_s \ell_{L^2}(\Sigma, K) + \lambda \cdot (\Sigma - f(\Sigma, K)) = T_s L^c_{L^2}(\Sigma, K, \lambda_c),$$

where

$$L^c_{L^2}(\Sigma, K, \lambda_c) := \frac{1}{4\sqrt{\pi}} \left[\Sigma^{-1/2} + 1 - 2\sqrt{2}(\Sigma + 1)^{-1/2}\right] + \frac{\gamma}{2} K^2 - \lambda_c \left(-2(\theta + K)\Sigma + \varsigma^2\right)$$

is the Lagrange function associated to the problem of finding the optimal equilibrium for the (original unscaled) stage cost $\ell_{L^2}(\Sigma, K)$ and continuous dynamics (6.4b)*. The Lagrange multiplier $\bar{\lambda}_c$ is unique and independent of the sampling time T_s. The connection to* (6.24) *is easily established via*

$$\bar{\lambda}_c = \bar{\lambda} T_s.$$

Thus, while the Lagrange multiplier $\bar{\lambda}$ from (6.24) *changes with T_s, the product $\bar{\lambda} T_s$ and the optimal equilibria (Σ^e, K^e) are independent of T_s. Since in this subsection only the product $\bar{\lambda} T_s$ is of relevance, scaling the stage cost ℓ_{L^2} yields no benefit.*

The second proposal concerns the control cost $\frac{\gamma}{2} K^2$ in the stage cost ℓ_{L^2}.

Remark 6.10 (Penalizing $(\theta + K)^2$ instead of K^2 in ℓ_{L^2})**.** *When switching from linear to bilinear systems, it appears reasonable to replace the term penalizing the control effort, K^2, with $(\theta + K)^2$ in $\ell_{L^2}(\Sigma, K)$ because this removes the discrepancy between the control term K^2 in the stage cost and the bilinear term $(\theta + K)\Sigma$ in the dynamics. However, the new cost yields the same optimal equilibria as formally setting θ to zero in the original stage cost. In particular, the case $\varsigma^2/2 - \theta < 0$ does not occur anymore and only the problematic case $\varsigma^2/2 - \theta > 0$ remains.*

A nonlinear storage function

Even though the OCP in Example 6.7 is not strictly dissipative with a linear storage function, numerical simulations indicate that the turnpike property holds for these parameters, see Figure 6.2. Due to the close connection of the turnpike property to dissipativity, see the end of Section 3.3, this strongly suggests that the OCP is indeed strictly dissipative, but with a nonlinear storage function. Thus, in the remainder of this subsection, we propose the nonlinear storage function

$$\lambda^s(z) := \alpha(z + 1)^{-1/2},$$

where $\alpha \in \mathbb{R}$ is chosen such that the optimal equilibrium (Σ^e, K^e) is a stationary point of the new modified cost

$$\tilde{\ell}^s_{L^2}(\Sigma, K) := \ell_{L^2}(\Sigma, K) - \ell_{L^2}(\Sigma^e, K^e) + \lambda^s(\Sigma) - \lambda^s(\Sigma^+).$$

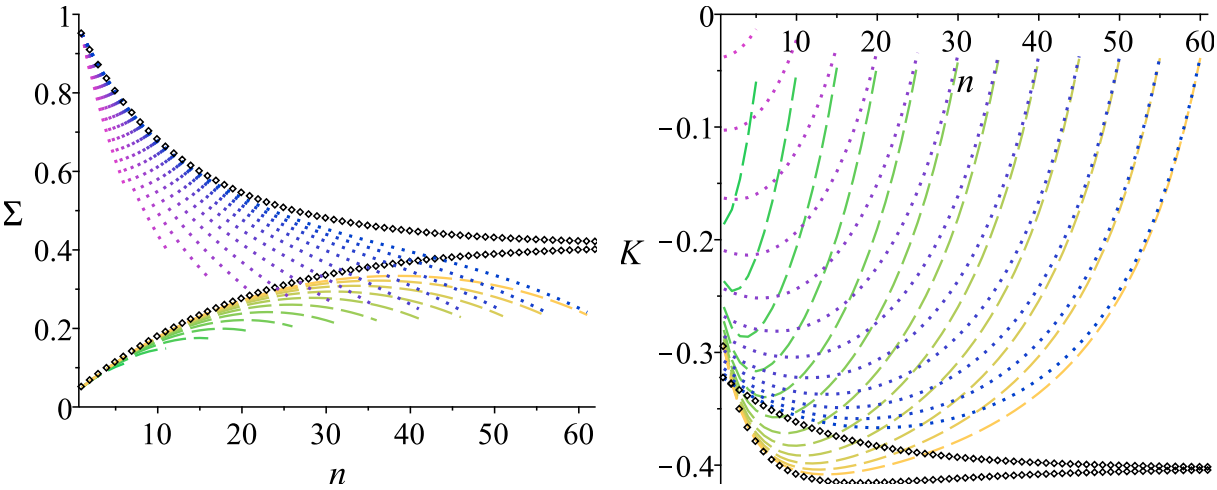


Figure 6.2: Open-loop optimal trajectories for various horizons N between 2 and 61 and MPC closed-loop trajectories for two initial conditions, indicating turnpike behavior in Example 6.7; Σ (left) and K (right).

Note that $\lambda^s(\Sigma^+)$ is well-defined since $\Sigma^+ > 0$, cf. (6.6).

In the case of Example 6.7, the level sets in Figure 6.3 (right) illustrate that the lowest value is attained at the optimal equilibrium (Σ^e, K^e), suggesting that strict dissipativity holds with the new storage function λ^s. In contrast, the white area in Figure 6.3 (left) shows that with a linear storage function, $\tilde{\ell}_{L^2}$ attains negative values.

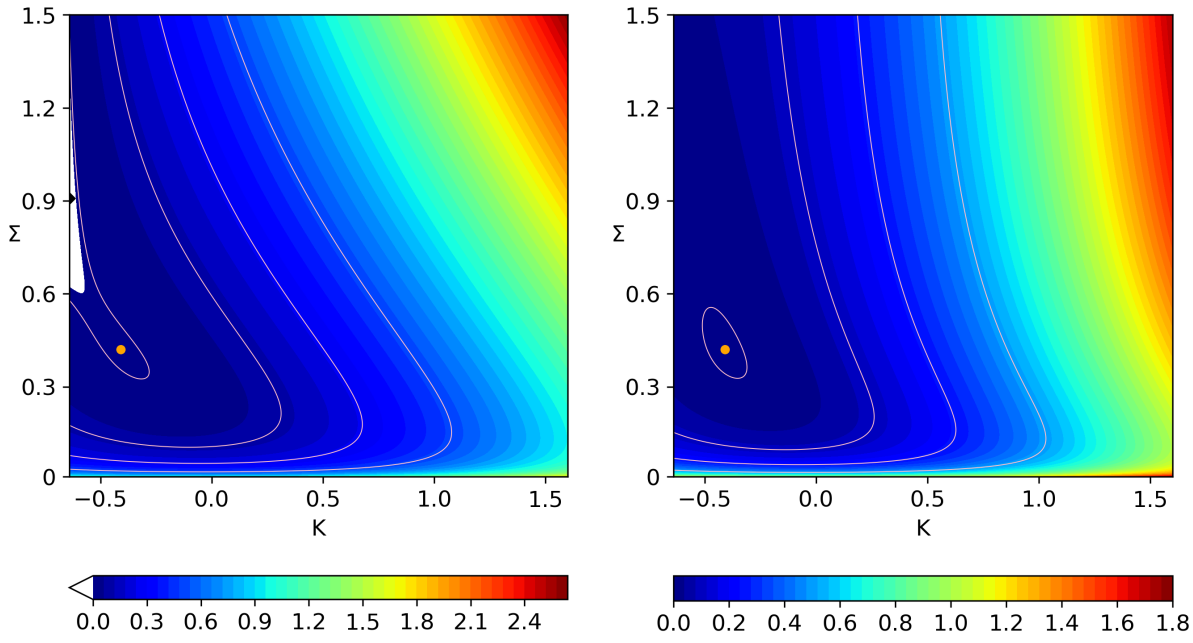


Figure 6.3: Modified costs $\tilde{\ell}_{L^2}(\Sigma, K)$ (left) and $\tilde{\ell}^s_{L^2}(\Sigma, K)$ (right) for Example 6.7. The optimal equilibrium (Σ^e, K^e) is illustrated by the orange circle. In the left plot, the white area on the left represents negative values; the black diamond at the left boundary marks the minimum of the depicted area.

Our final example suggests that λ^s also works for parameter values for which Proposition 6.6 rules out strict dissipativity with a linear storage function.

Example 6.11. *Consider* (6.23) *with the parameters*

$$\varsigma = 10, \quad \theta = 2, \quad \gamma = 1/4, \quad \text{and} \quad T_s = 1/10.$$

The optimal equilibrium (Σ^e, K^e) *is given by* $\Sigma^e \approx 24.4333301$ *and* $K^e \approx 0.04638499$*; with* $Z \approx -0.00237304$*. Figure 6.4 and the level sets therein indicate that strict dissipativity holds with* λ^s*, however not with* λ^l*.*

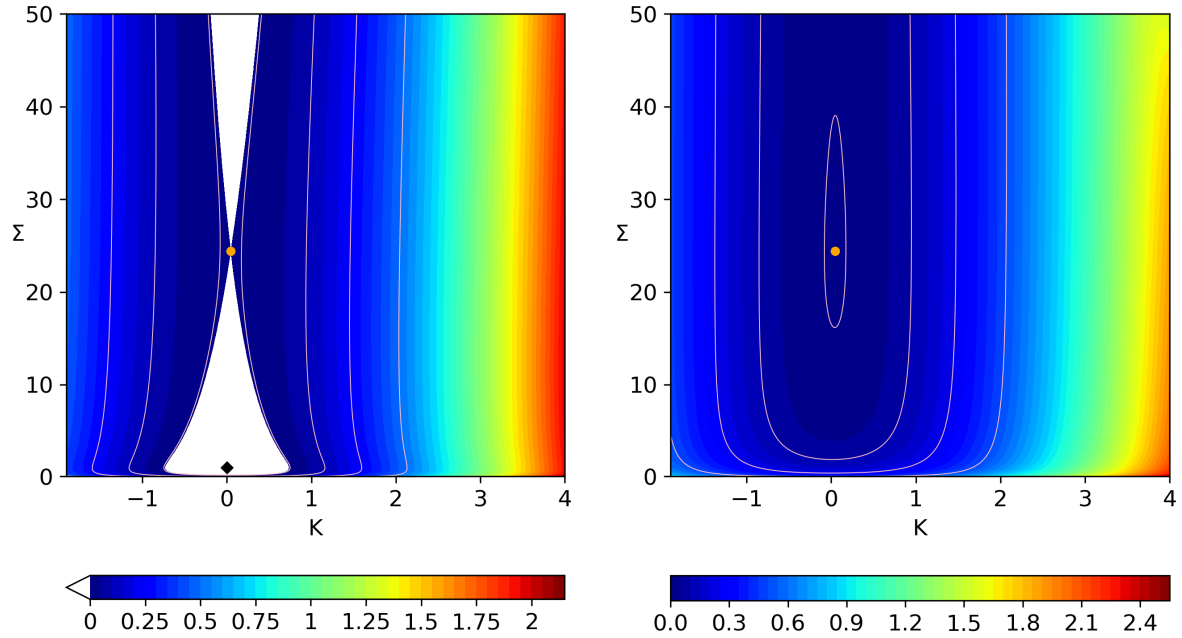


Figure 6.4: Modified costs $\tilde{\ell}_{L^2}(\Sigma, K)$ (left) and $\tilde{\ell}^s_{L^2}(\Sigma, K)$ (right) for Example 6.11. The optimal equilibrium (Σ^e, K^e) is illustrated by the orange circle. In the left plot, the white area on the left represents negative values; the black diamond at the bottom marks the minimum of the depicted area.

The storage function $\lambda^s(z)$ "fixes" the asymptotic behavior of the modified cost $\tilde{\ell}_{L^2}(\Sigma, K)$ for $\Sigma \to \infty$ and admissible controls, at least to a certain extent:

$$\lim_{\Sigma\to\infty} \tilde{\ell}^s_{L^2}(\Sigma, K) = \frac{1}{4\sqrt{\pi}} + \frac{\gamma}{2}K^2 - \ell_{L^2}(\Sigma^e, K^e) \geq \frac{1}{4\sqrt{\pi}} - \ell_{L^2}(\Sigma^e, K^e).$$

This storage function is unsuitable if $\frac{1}{4\sqrt{\pi}} - \ell_{L^2}(\Sigma^e, K^e) < 0$ since $K = 0$ is admissible. One possible extension to overcome this problem is to add a linear term βz with $\beta > 0$ to $\lambda^s(z)$, i.e.,

$$\lambda^{s_2}(z) := \alpha(z+1)^{-1/2} + \beta z.$$

Then for the corresponding modified cost $\tilde{\ell}^{s_2}_{L^2}(\Sigma, K)$ we get $\tilde{\ell}^{s_2}_{L^2}(\Sigma, K) \to \infty$ as $\Sigma \to \infty$ for admissible controls. Moreover, the additional degree of freedom makes it easier to deal with boundary values $\Sigma \searrow 0$ and $K \searrow -\theta$. However, the parameters $\alpha \in \mathbb{R}$ and $\beta > 0$ still need to be chosen such that (Σ^e, K^e) is a stationary state of the modified cost. Furthermore, the problem of possibly multiple stationary states remains. Ideally, the storage function should be chosen such that the stationary states of the modified cost can be calculated analytically, or at least allow a statement about the maximum number of stationary states, such as Proposition 6.5.

Summary

The different cases in Lemma 6.4 were decisive for the analysis:

- For $\varsigma^2/2 - \theta > 0$, strict dissipativity cannot hold with a linear storage function.
- In the case $\varsigma^2/2 - \theta < 0$, strict dissipativity with a linear storage function has to be checked on a case-by-case basis.
- For both $\varsigma^2/2 - \theta > 0$ and $\varsigma^2/2 - \theta < 0$, we have constructed a nonlinear storage function for which numerical evidence strongly suggests that strict dissipativity holds for certain values of θ and ς.
- Numerical verification of the turnpike property suggests that strict dissipativity holds for many parameters for which the analytical verification is not (yet) possible.

6.3.2 2F cost

In this subsection we consider (6.15) with the 2F stage cost (6.18). In the one-dimensional case, this amounts to penalizing the quadratic deviation of the variance in addition to the control effort. Overall, the optimization problem in this section is given by

$$\begin{aligned} J_N(\mathring{\Sigma}, \mathbf{K}) \;&:=\; \sum_{k=0}^{N-1} \ell_{2F}(\Sigma(k), K(k)) \;=\; \sum_{k=0}^{N-1} \left[\frac{1}{2}(\Sigma(k)-1)^2 + \frac{\gamma}{2}K(k)^2\right] \;\to\; \min_{\mathbf{K}}! \\ \text{s.t.}\quad & \Sigma^+ = \Sigma + T_s\left(-2K_\theta\Sigma + \varsigma^2\right) = f(\Sigma, K), \\ & \Sigma(0) = \mathring{\Sigma}, \\ & 0 < K_\theta(k) < (\Sigma(k) + T_s\varsigma^2)/(2T_s\Sigma(k)), \quad k \in \{0, ..., N-1\}. \end{aligned} \tag{6.27}$$

For the linear storage function $\lambda^l(z)$, the corresponding modified cost $\tilde{\ell}_{2F}(\Sigma, K)$ reads

$$\tilde{\ell}_{2F}(\Sigma, K) := \frac{1}{2}(\Sigma - 1)^2 + \frac{\gamma}{2}K^2 - \ell_{2F}(\Sigma^e, K^e) - \bar{\lambda}T_s\left(-2(\theta + K)\Sigma + \varsigma^2\right). \tag{6.28}$$

Analogous to Subsection 6.3.1, the unique Lagrange multiplier $\bar{\lambda} \in \mathbb{R}$ is obtained from the Lagrange function

$$L_{2F}(\Sigma, K, \lambda) := \frac{1}{2}(\Sigma - 1)^2 + \frac{\gamma}{2}K^2 + \lambda\left[-T_s\left(-2(\theta + K)\Sigma + \varsigma^2\right)\right], \tag{6.29}$$

which is closely connected to the modified cost $\tilde{\ell}_{2F}$. As in Subsection 6.3.1, we first characterize the stationary points of $\tilde{\ell}_{2F}$ for a fixed $\bar{\lambda}$. Using the notation $Z = 2\bar{\lambda}T_s$, the gradient of $\tilde{\ell}_{2F}$ is given by

$$\nabla\tilde{\ell}_{2F}(\Sigma, K) = \begin{pmatrix} \Sigma - 1 \\ \gamma K \end{pmatrix} + Z\begin{pmatrix} \theta + K \\ \Sigma \end{pmatrix}, \tag{6.30}$$

and it holds

$$\nabla\tilde{\ell}_{2F}(\Sigma, K) = \nabla_{\Sigma,K} L_{2F}(\Sigma, K, \bar{\lambda}). \tag{6.31}$$

Lemma 6.12. *For a fixed $\bar{\lambda} \in \mathbb{R}$ and thus fixed Z, the stationary points of $\tilde{\ell}_{2F}(\Sigma, K)$ are given by either*

$$\Sigma = -\frac{\gamma(Z\theta - 1)}{\gamma - Z^2}, \quad K = \frac{Z(Z\theta - 1)}{\gamma - Z^2} \tag{6.32}$$

if $\gamma - Z^2 \neq 0$ or by

$$\Sigma = -\frac{K}{\theta} \tag{6.33}$$

for arbitrary K in case $\gamma - Z^2 = 0$.

Proof. Solving $\partial_\Sigma \tilde{\ell}_{2F}(\Sigma, K) = 0$ for Σ yields

$$\Sigma = 1 - Z(\theta + K), \tag{6.34}$$

cf. (6.30). Plugging this into $\partial_K \tilde{\ell}_{2F}(\Sigma, K) = 0$ results in

$$0 = \gamma K + Z\left(1 - Z(\theta + K)\right) = \left(\gamma - Z^2\right) K + Z(1 - Z\theta). \tag{6.35}$$

Assuming that $\gamma - Z^2 \neq 0$, one can solve for K, which results in the equation for K in (6.32). Plugging this K into (6.34) gives the equation for Σ in (6.32).

If $\gamma - Z^2 = 0$, then $Z \neq 0$ since $\gamma > 0$. Since an optimal equilibrium (Σ^e, K^e) is always a stationary point of $\tilde{\ell}_{2F}(\Sigma, K)$ due to $\nabla \tilde{\ell}_{2F}(\Sigma^e, K^e) = \nabla_{\Sigma,K} L_{2F}(\Sigma^e, K^e, \bar{\lambda}) = 0$, cf. (6.31), we infer from (6.35) that $1 - Z\theta = 0$, i.e., $Z = 1/\theta$. In this case, from (6.34) we get (6.33) for arbitrary K. □

Remark 6.13. *(a) The set of possible stationary points in Lemma 6.12 is restricted by the constraints $K > -\theta$ and $\Sigma > 0$. More importantly, however, in the case of $\gamma - Z^2 \neq 0$, the stationary point is unique and coincides with (Σ^e, K^e), whereas there can be infinitely many stationary points if $\gamma - Z^2 = 0$.*

(b) From the proof of Lemma 6.12 we see that $\gamma - Z^2 = 0$ can only occur if $\gamma = 1/\theta^2$.

As indicated in the proof of Lemma 6.12, the sign of $\gamma - Z^2$ is indeed crucial for the rest of this subsection: Since the Hessian

$$\nabla^2 \tilde{\ell}_{2F}(\Sigma, K) = \begin{pmatrix} 1 & Z \\ Z & \gamma \end{pmatrix} \tag{6.36}$$

is constant, the necessary condition for strict dissipativity that the optimal equilibrium is a (strict) global minimum of the modified cost $\tilde{\ell}_{2F}$ is indeed sufficient, thus equivalent. This requirement is met if and only if $\gamma - Z^2 > 0$, i.e., if the modified cost $\tilde{\ell}_{2F}$ is strongly convex. Hence, strict dissipativity with the linear storage function $\lambda^l(z)$ is equivalent to strong convexity of the modified cost $\tilde{\ell}_{2F}$.

This is in contrast to the L^2 cost from Subsection 6.3.1, where the modified cost $\tilde{\ell}_{L^2}$ is not convex for sufficiently large Σ and strong convexity (and also strict convexity) of the modified cost is only sufficient for strict dissipativity. Instead, the 2F cost is more similar to the linear setting considered in [24], where strict convexity of the stage cost ℓ is sufficient for strict dissipativity. The difference, of course, lies in the bilinear terms in the dynamics f, which cause non-zero entries on the off-diagonal of the Hessian (6.36). Because of this, convexity of the stage cost ℓ_{2F} does not necessarily carry over to the modified cost $\tilde{\ell}_{2F}$. Hence, we check the convexity of $\tilde{\ell}_{2F}$ directly. To this end, the decisive factor is the sign of $\gamma - Z^2$. Thus, in the following, we focus on finding sets of parameters for which a certain sign of $\gamma - Z^2$ can be guaranteed. As in Subsection 6.3.1, we consider the two cases $\varsigma^2/2 - \theta > 0$ and $\varsigma^2/2 - \theta < 0$ separately.

The case $\varsigma^2/2 - \theta > 0$

In contrast to the L^2 cost, where for $\varsigma^2/2-\theta > 0$ and large enough Σ (strict) dissipativity does not hold with a linear storage function (see Proposition 6.6), with the 2F cost (strict) dissipativity does hold with a linear storage function, see the following proposition.

Proposition 6.14. *If $\varsigma^2/2 - \theta > 0$, then* (6.27) *is strictly dissipative with the linear storage function $\lambda^l(z)$ from* (6.11)*.*

Proof. The assertion follows from the fact that for $\varsigma^2/2 - \theta > 0$ the Hessian $\nabla^2 \tilde{\ell}_{2F}(\Sigma, K)$ is positive definite. Indeed, in this case the modified cost $\tilde{\ell}_{2F}$ is strongly convex, which immediately implies the existence of a quadratic lower bound $\varrho \in \mathcal{K}_\infty$ in the dissipativity inequality (3.10).

It is thus sufficient to prove that the Hessian is positive definite, which holds if and only if $\gamma - Z^2 > 0$. To prove this, we need some information about the Lagrange multiplier $\bar{\lambda}$, which we get by taking a closer look at the Lagrange function (6.29). Since (Σ^e, K^e) is an optimal equilibrium, $\nabla L_{2F}(\Sigma^e, K^e, \bar{\lambda}) = 0$. In particular, we can use the results of Lemma 6.12 due to (6.31).

First, we show that $\gamma - Z^2 \neq 0$: If we assume the opposite, then from Lemma 6.12 we see that the optimal equilibrium (Σ^e, K^e) satisfies (6.33), i.e., $\Sigma^e = -K^e/\theta$ for some K^e. However, from Lemma 6.4 we know that $K^e \in [0, \frac{\varsigma^2}{2} - \theta]$ and $\Sigma^e \in [1, \frac{\varsigma^2}{2\theta}]$. In particular, $\Sigma^e \geq 1$ and $K^e \geq 0$, which contradicts (6.33).

Knowing that $\gamma - Z^2 \neq 0$, we now show that $\gamma - Z^2 > 0$: Since $\Sigma^e > 0$, (6.32) can only be satisfied in the following two cases:

$$\begin{aligned} &\text{Case 1: } Z\theta - 1 < 0 \quad \wedge \quad \gamma - Z^2 > 0, \\ &\text{Case 2: } Z\theta - 1 > 0 \quad \wedge \quad \gamma - Z^2 < 0. \end{aligned}$$

Since $K^e + \theta > 0$ and $\Sigma^e \in [1, \frac{\varsigma^2}{2\theta}]$, from (6.34) we conclude that $Z \leq 0$. Then due to $\theta > 0$ case 2 can be excluded, which concludes the proof. □

The case $\varsigma^2/2 - \theta = 0$ is of no particular interest as it corresponds to the case of stabilizing MPC, cf. Lemma 6.4. Therefore, the natural follow-up question is what happens in case of $\varsigma^2/2 - \theta < 0$.

The case $\varsigma^2/2 - \theta < 0$

Analogously to the proof of Proposition 6.14 one can show that $Z \geq 0$ if $\varsigma^2/2-\theta < 0$. Still, we can prove strong convexity of $\tilde{\ell}_{2F}$ also for $\varsigma^2/2-\theta < 0$, by adjusting the regularization parameter γ.

Proposition 6.15. *Let $\varsigma^2/2 - \theta < 0$ and $\gamma > 1/(4\varsigma^4)$. Then* (6.27) *is strictly dissipative with the linear storage function $\lambda^l(z)$.*

Proof. From (6.19) we know that $\Sigma^e \leq 1$. Then from (6.34) and $\theta + K^e > 0$ we conclude that $Z \geq 0$. If $Z = 0$ then the assertion follows ($\gamma - Z^2 = \gamma > 0$). Thus, we consider $Z > 0$. It holds that

$$\Sigma^e = 1 - Z(\theta + K^e) \overset{!}{=} \frac{\varsigma^2}{2(\theta + K^e)} \quad \Leftrightarrow \quad K^e + \theta = \frac{1}{2Z}(1 \pm \sqrt{1 - 2Z\varsigma^2}).$$

In particular, $1 - 2Z\varsigma^2 \geq 0$, which, due to $Z, \varsigma^2 > 0$, is equivalent to $Z^2 \leq \frac{1}{4\varsigma^4}$. Thus, for $\gamma > \frac{1}{4\varsigma^4}$, we have

$$Z^2 \leq \frac{1}{4\varsigma^4} < \gamma,$$

i.e., $\gamma - Z^2 > 0$, which concludes the proof. □

Without the restriction on γ there is one problematic case, in which we indeed lose strict dissipativity due to $\gamma - Z^2 = 0$. According to Remark 6.13(b), for this to happen it is necessary that $\gamma = 1/\theta^2$. The following proposition deals with this special case.

Proposition 6.16. *Let $\gamma = 1/\theta^2$.*

(a) If $2\varsigma^2 - \theta < 0$, then the optimal equilibrium (Σ^e, K^e) is not unique. In particular, (6.27) is not strictly dissipative (irrespective of the storage function λ). However, it is dissipative with the linear storage function $\lambda^l(z)$.

(b) If $2\varsigma^2 - \theta = 0$, then (6.27) is dissipative with $\lambda^l(z)$ but not strictly dissipative.

(c) If $2\varsigma^2 - \theta > 0$, then (6.27) is strictly dissipative with $\lambda^l(z)$.

Proof. We first calculate the stationary points that are equilibria. To this end, we use

$$0 = \Sigma - f(\Sigma, K) \quad \Leftrightarrow \quad \Sigma = \frac{\varsigma^2}{2(\theta + K)} \tag{6.37}$$

and plug this state into the cost function ℓ_{2F}, i.e.,

$$\ell_{2F}\left(\frac{\varsigma^2}{2(\theta+K)}, K\right) = \frac{1}{2}\left[\left(\frac{\varsigma^2}{2(\theta+K)} - 1\right)^2 + \gamma K^2\right] =: \hat{\ell}_{2F}(K). \tag{6.38}$$

Then we compute the stationary points of the *reduced cost function* $\hat{\ell}_{2F}(K)$ in the special case $\gamma = \frac{1}{\theta^2}$:

$$\hat{\ell}'_{2F}(K) = -\frac{2\varsigma^2}{(2(\theta+K))^2}\left(\frac{\varsigma^2}{2(\theta+K)} - 1\right) + \frac{K}{\theta^2} = 0 \quad \Leftrightarrow \quad K = K_i, \quad i = 1, \ldots, 4$$

with

$$K_{1/2} := -\frac{\theta}{2} \pm \frac{\sqrt{\theta}\sqrt{\theta - 2\varsigma^2}}{2} \quad \text{and} \quad K_{3/4} := -\theta \pm \frac{\sqrt{2\theta\varsigma^2}}{2}.$$

Since $K_4 = -\theta - \frac{\sqrt{2\theta\varsigma^2}}{2}$ violates the constraint $K > -\theta$, we ignore this solution. Moreover, we only care about real solutions. Therefore, we have three distinct solutions if and only if $2\varsigma^2 - \theta < 0$.[3] Now we consider the three different cases in the Proposition.

(a) Let $2\varsigma^2 - \theta < 0$. Then the controls K_1, K_2, and K_3 satisfy (6.6) with Σ as in (6.37). The respective cost is given by

$$\hat{\ell}_{2F}(K_1) = \frac{(\varsigma^2 - \theta)\left(\varsigma^2 - \theta - \sqrt{-2\varsigma^2\theta + \theta^2}\right)}{\left(\theta + \sqrt{-2\varsigma^2\theta + \theta^2}\right)^2} = \frac{\theta - \varsigma^2}{2\theta} = \hat{\ell}_{2F}(K_2)$$

[3] For $2\varsigma^2 - \theta = 0$ we have that $K_1 = K_2 = K_3$, i.e., only one stationary point.

and

$$\hat{\ell}_{2F}(K_3) = \frac{\varsigma^2 - 2\sqrt{2\varsigma^2\theta} + 2\theta}{2\theta}.$$

We can exclude a minimum of $\hat{\ell}_{2F}(K)$ on the boundary since $\hat{\ell}_{2F}(K) \to \infty$ for $K \searrow -\theta$ and for $K \to \infty$. Since

$$\hat{\ell}_{2F}(K_3) - \hat{\ell}_{2F}(K_1) = \frac{2\varsigma^2 - 2\sqrt{2\varsigma^2\theta} + \theta}{2\theta} = \frac{(\sqrt{2\varsigma^2} - \sqrt{\theta})^2}{2\theta} > 0,$$

there are two optimal equilibria, characterized by K_1 and K_2. Thus, strict dissipativity is out of the question. However, we argue that dissipativity with $\lambda^l(z)$ does hold. For this, we show that $\gamma - Z^2 = 0$, i.e., that $\tilde{\ell}_{2F}(\Sigma, K)$ is convex but not strictly (and thus not strongly) convex. With the corresponding states

$$\Sigma_1 = \frac{\varsigma^2}{\theta + \sqrt{\theta}\sqrt{\theta - 2\varsigma^2}} \quad \text{and} \quad \Sigma_2 = \frac{\varsigma^2}{\theta - \sqrt{\theta}\sqrt{\theta - 2\varsigma^2}},$$

a short calculation using

$$0 = \partial_K L_{2F}(\Sigma^e, K^e, \bar{\lambda}) = \gamma K^e + Z\Sigma^e$$

yields the associated Lagrange multipliers $Z_1 = \frac{1}{\theta} = Z_2$. In particular, we have

$$\gamma - Z_1^2 = 0 = \gamma - Z_2^2.$$

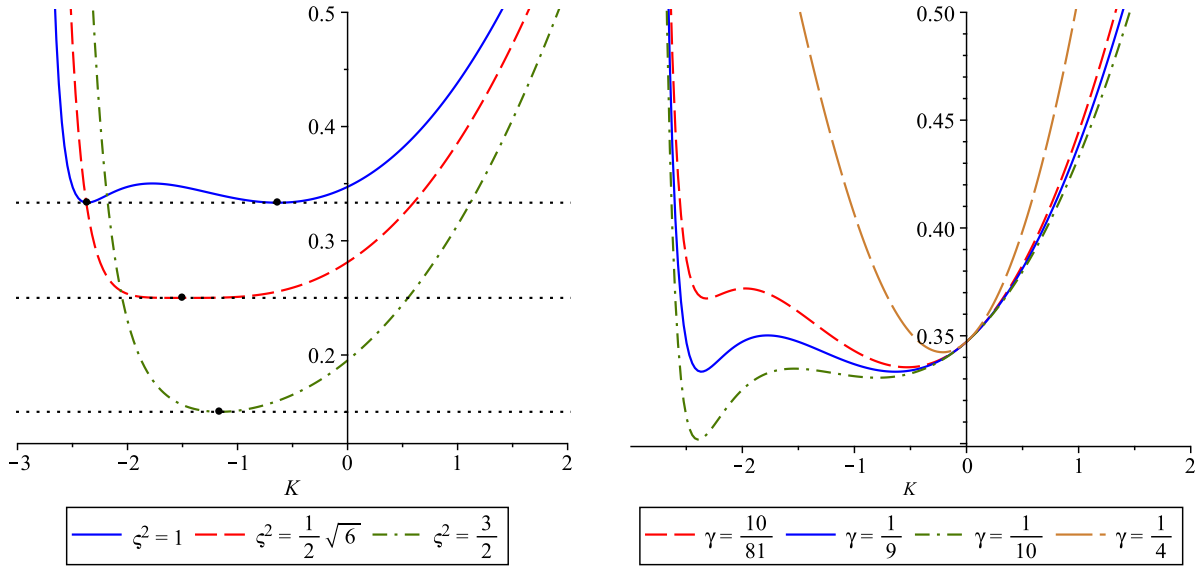
(b) For $2\varsigma^2 - \theta = 0$, we get the same result, i.e., dissipativity but not strict dissipativity.

(c) Lastly, if $2\varsigma^2 - \theta > 0$, then (Σ_3, K_3) with $\Sigma_3 = \sqrt{\frac{\varsigma^2}{2\theta}}$ is the unique optimal equilibrium and an analogous calculation reveals that $\gamma - Z_3^2 > 0$, i.e., strong convexity of $\tilde{\ell}_{2F}$ and thus strict dissipativity. □

The three cases of Proposition 6.16 are exemplarily illustrated in Figure 6.5a.

Remark 6.17. *Coinciding with the requirement on γ in Proposition 6.15, the reduced cost $\hat{\ell}_{2F}(K)$ from (6.38) is convex if and only if $\gamma \geq 1/(4\varsigma^4)$, cf. Figure 6.5b. However, in general, convexity of the reduced cost $\hat{\ell}$ does not transfer to the modified cost $\tilde{\ell}$.*

We briefly summarize the case $\varsigma^2/2 - \theta < 0$. Instead of a case-by-case analysis that was required for the L^2 cost in Subsection 6.3.1, we have shown strict dissipativity provided that $\gamma > 1/(4\varsigma^4)$. Furthermore, we have identified cases in which strict dissipativity does not hold due to the existence of two optimal equilibria, which can only happen if $\gamma = 1/\theta^2$. Even for $\varsigma^2/2 - \theta < 0$ and $\gamma \leq 1/(4\varsigma^4)$, as long as $\gamma \neq 1/\theta^2$, our numerous simulations indicate that $\gamma - Z^2 > 0$. Thus, we conjecture that strict dissipativity (with a linear storage function) holds for the 2F cost provided that $\gamma \neq 1/\theta^2$. To prove this rigorously, one could solve $\nabla L_{2F}(\Sigma, K, \bar{\lambda}) = 0$ for arbitrary $\gamma > 0$. Ultimately, as (6.38) indicates, this requires finding the roots of a fourth-order polynomial. We avoid from carrying out this computation here for the sake of brevity.



(a) $\hat{\ell}_{2F}(K)$ for $\theta = 3$, $\gamma = 1/\theta^2$ and various values of ς^2 with the respective minima.

(b) $\hat{\ell}_{2F}(K)$ for $\varsigma = 1$, $\theta = 3$, and various values of γ.

Figure 6.5: (Non-)Convexity of the reduced cost $\hat{\ell}_{2F}(\Sigma, K)$ depending on ς^2 (left) and on γ (right).

Modifications to the stage cost ℓ_{2F}

In this part we discuss two modifications to the stage cost ℓ_{2F}, both of which have been considered for the L^2 cost, see Remarks 6.9 and 6.10.

Remark 6.18 (Scaling of the stage cost)**.** *Analogously to the L^2 cost, cf. Remark 6.9, we do not scale the stage cost ℓ_{2F} since, throughout this subsection, only the product $\bar{\lambda}T_s$ is of relevance.*

Proposition 6.19 (Penalizing $(\theta + K)^2$ instead of K^2 in ℓ_{2F})**.** *If, instead of $\ell_{2F}(\Sigma, K)$, the stage cost* (6.18) *in the OCP* (6.27) *is defined by*

$$\ell_{2F,\theta}(\Sigma, K) := \frac{1}{2}(\Sigma - 1)^2 + \frac{\gamma}{2}(K + \theta)^2, \tag{6.39}$$

then (6.27) *is strictly dissipative with the linear storage function $\lambda^l(z)$.*

Proof. To conclude strict dissipativity, we prove that $\tilde{\ell}_{2F,\theta}(\Sigma, K)$, defined analogously to (6.28), is strongly convex. To this end, we define $L_{2F,\theta}(\Sigma, K, \lambda)$ analogously to (6.29). Then

$$\partial_\Sigma L_{2F,\theta}(\Sigma, K, \bar{\lambda}) = \Sigma - 1 + 2\bar{\lambda}T_s(\theta + K) \tag{6.40}$$

and

$$\partial_K L_{2F,\theta}(\Sigma, K, \bar{\lambda}) = \gamma(\theta + K) + 2\bar{\lambda}T_s\Sigma. \tag{6.41}$$

With $Z = 2\bar{\lambda}T_s$, solving $\partial_\Sigma L_{2F,\theta}(\Sigma, K, \bar{\lambda}) = 0$ for Σ yields

$$\Sigma = 1 - Z(\theta + K). \tag{6.42}$$

Plugging this into $\partial_K L_{2F,\theta}(\Sigma, K, \bar{\lambda}) = 0$ results in

$$0 = \gamma(\theta + K) + Z\left(1 - Z(\theta + K)\right) = \left(\gamma - Z^2\right)(\theta + K) + Z. \tag{6.43}$$

From (6.43) we can exclude the case $\gamma - Z^2 = 0$ since $\gamma > 0$ and we know that at least one optimal equilibrium exists, i.e., (6.43) has at least one admissible solution. Thus, $\gamma - Z^2 \neq 0$, in which case

$$\theta + K = -\frac{Z}{\gamma - Z^2} \tag{6.44}$$

and therefore, according to (6.42),

$$\Sigma = 1 + \frac{Z^2}{\gamma - Z^2} = \frac{\gamma}{\gamma - Z^2}. \tag{6.45}$$

Since $\Sigma > 0$ and $\gamma > 0$, from (6.45) we infer that $\gamma - Z^2 > 0$, i.e., $\tilde{\ell}_{2F,\theta}(\Sigma, K)$ is strongly convex. □

Note that (6.44)-(6.45) coincides with (6.32) in the case $\theta = 0$. For $\theta = 0$ the requirements of Proposition 6.14 are met and thus the result of Proposition 6.19 is not surprising. Although the stage cost $\ell_{2F,\theta}(\Sigma, K)$ is much easier to handle, the price to pay is the loss of optimal equilibria with $\Sigma^e \in [0, 1]$: we can see from (6.45) that $\Sigma^e = 1 + \frac{Z^2}{\gamma - Z^2} > 1$ since $\gamma - Z^2 > 0$.

Summary

We summarize our results for the 2F cost in a similar form as for the L^2 cost:

- For $\varsigma^2/2 - \theta > 0$, strict dissipativity holds with a linear storage function.
- For $\varsigma^2/2 - \theta < 0$ and $\gamma > 1/(4\varsigma^2)$, strict dissipativity holds with a linear storage function.
- For $\varsigma^2/2 - \theta < 0$ and $\gamma \leq 1/(4\varsigma^2)$, strict dissipativity fails to hold for some parameter values if $\gamma = 1/\theta^2$. Numerical evidence suggests that strict dissipativity always holds if $\gamma \neq 1/\theta^2$.
- If ℓ_{2F} is replaced by $\ell_{2F,\theta}$ from (6.39), then strict dissipativity holds for all parameter values.

We emphasize once more that for the 2F stage cost considered in this subsection, proving strict dissipativity with a linear storage function is equivalent to proving strong convexity of $\tilde{\ell}_{2F}(\Sigma, K)$. This is in contrast to the L^2 cost (6.8), where the modified cost was never convex, but for some parameters the OCP was nevertheless strictly dissipative with a linear storage function, cf. Example 6.8. In this sense, the W^2 cost considered in the following subsection is more similar to the L^2 cost than to the 2F cost.

6.3.3 W^2 cost

The W^2 cost is designed to measure the distance between two PDFs. In our case, it differs only slightly from the 2F cost in Subsection 6.3.2: Instead of $(\Sigma-1)^2$, the square root of the current and the desired state is taken and a quadratic cost is inflicted on the distance thereof, i.e., $(\sqrt{\Sigma}-1)^2$. In this one-dimensional case, this amounts to penalizing the difference in the standard deviation instead of in the variance. Surprisingly, this small difference changes the dissipativity analysis considerably.

Overall, the optimization problem in this section is given by

$$\begin{aligned} J_N(\mathring{\Sigma},\mathbf{K}) &:= \sum_{k=0}^{N-1} \ell_{W^2}(\Sigma(k),K(k)) = \sum_{k=0}^{N-1}\left[\frac{1}{2}\left(\sqrt{\Sigma(k)}-1\right)^2+\frac{\gamma}{2}K(k)^2\right] \to \min_{\mathbf{K}}! \\ \text{s.t.}\quad \Sigma^+ &= \Sigma + T_s\left(-2K_\theta\Sigma+\varsigma^2\right) = f(\Sigma,K), \\ \Sigma(0) &= \mathring{\Sigma}, \\ 0 &< K_\theta(k) < (\Sigma(k)+T_s\varsigma^2)/(2T_s\Sigma(k)), \quad k\in\{0,...,N-1\}. \end{aligned} \tag{6.46}$$

For the linear storage function $\lambda^l(z)$, the corresponding modified cost $\tilde{\ell}_{W^2}(\Sigma,K)$ reads

$$\tilde{\ell}_{W^2}(\Sigma,K) := \frac{1}{2}\left(\sqrt{\Sigma}-1\right)^2+\frac{\gamma}{2}K^2-\ell_{W^2}(\Sigma^e,K^e)+\bar{\lambda}\left(-T_s(-2(\theta+K)\Sigma+\varsigma^2)\right).$$

Analogous to Subsections 6.3.1 and 6.3.2, the Lagrange multiplier $\bar{\lambda}\in\mathbb{R}$ obtained from the Lagrange function

$$L_{W^2}(\Sigma,K,\lambda) := \frac{1}{2}\left(\sqrt{\Sigma}-1\right)^2+\frac{\gamma}{2}K^2+\lambda\left[-T_s\left(-2(\theta+K)\Sigma+\varsigma^2\right)\right] \tag{6.47}$$

is unique. We begin as in Subsections 6.3.1 and 6.3.2, i.e., by counting the stationary points of $\tilde{\ell}_{W^2}$. With $Z=2\bar{\lambda}T_s$, the gradient reads

$$\nabla\tilde{\ell}_{W^2}(\Sigma,K) = \begin{pmatrix} \frac{\sqrt{\Sigma}-1}{2\sqrt{\Sigma}} \\ \gamma K \end{pmatrix} + Z\begin{pmatrix} \theta+K \\ \Sigma \end{pmatrix} \tag{6.48}$$

and we arrive at the same result as for the L^2 cost, cf. Proposition 6.5.

Proposition 6.20. *For a fixed $\bar{\lambda}$ and thus fixed Z the modified cost $\tilde{\ell}_{W^2}(\Sigma,K)$ has at most two admissible stationary points. If $Z=0$, then only one admissible stationary point of $\tilde{\ell}_{W^2}(\Sigma,K)$ exists and it is given by $(\Sigma^e,K^e)=(1,0)$.*

Proof. From the gradient (6.48) we infer that for stationary points $K=-Z\Sigma/\gamma$ and therefore,

$$0=\frac{\sqrt{\Sigma}-1}{2\sqrt{\Sigma}}+Z(\theta+K)=\frac{\sqrt{\Sigma}-1}{2\sqrt{\Sigma}}+Z\theta-\frac{Z^2\Sigma}{\gamma} =: h(\Sigma). \tag{6.49}$$

If $Z=0$, then $K=0=K^e$ and hence $h(\Sigma)=0\Leftrightarrow\Sigma=1=\Sigma^e$, i.e., $(\Sigma^e,K^e)=(1,0)$ is the unique admissible stationary point of $\tilde{\ell}_{W^2}(\Sigma,K)$.

Let $Z\neq 0$. If $h(\Sigma)$ has a unique admissible stationary point, then at most two admissible solutions for (6.49) can exist. To this end, we look at $h'(\Sigma)$:

$$h'(\Sigma)=\frac{1}{4\Sigma^{3/2}}-\frac{Z^2}{\gamma}=0 \quad\Leftrightarrow\quad \Sigma=\left(\frac{\gamma}{4Z^2}\right)^{2/3} =: \Sigma^s_{W^2}.$$

Since $\Sigma^s_{W^2}$ is admissible due to $Z\neq 0$, the assertion follows. □

The result of Proposition 6.20 is in contrast to the 2F cost: Apart from the degenerate case $\gamma = 1/\theta^2$, in which infinitely many stationary points of $\tilde{\ell}_{2F}$ exist, $\tilde{\ell}_{2F}$ exhibits a unique stationary point for a fixed Z, cf. Lemma 6.12. Hence, concerning stationary points of the modified cost, the W^2 cost is more similar to the L^2 cost than to the 2F cost.

The similarity of the W^2 cost to the L^2 cost appears in the Hessian as well: For any fixed $Z \neq 0$, it is obvious from the Hessian

$$\nabla^2 \tilde{\ell}_{W^2}(\Sigma, K) = \begin{pmatrix} \frac{1}{4\Sigma^{3/2}} & Z \\ Z & \gamma \end{pmatrix} \tag{6.50}$$

that $\tilde{\ell}_{W^2}$ is not convex for sufficiently large Σ. This is in contrast to Subsection 6.3.2, where the constant Hessian considerably simplified the analysis. Of course strong convexity of $\tilde{\ell}_{W^2}$ is only a sufficient condition for strict dissipativity. A requirement, however, is that the optimal equilibrium (Σ^e, K^e) is the unique global minimum of the modified cost $\tilde{\ell}_{W^2}$. Hence, in the following, we will take a closer look at the structure of $\tilde{\ell}_{W^2}$. As before, cf. Subsections 6.3.1 and 6.3.2, we separate the two cases $\varsigma^2/2 - \theta > 0$ and $\varsigma^2/2 - \theta < 0$.

The case $\varsigma^2/2 - \theta > 0$

Similar to the L^2 cost and in contrast to the 2F cost, cf. Propositions 6.6 and 6.14, in case of the W^2 cost, for a large set of parameters (strict) dissipativity does not hold with a linear storage function.

Proposition 6.21. *If $\varsigma^2/2 - \theta > 0$, then for sufficiently small sampling times $T_s > 0$,* (6.46) *is not dissipative with a linear storage function $\lambda^l(z) = \bar{\lambda} z$.*

Proof. The idea of the proof is the same as in the proof of Proposition 6.6, i.e., to show that the modified cost $\tilde{\ell}_{W^2}$ can assume negative values, which violates (3.10). To this end, we first note that

$$\lim_{\Sigma \to \infty} \tilde{\ell}_{W^2}(\Sigma, K) = \operatorname{sgn}\left((K+\theta)Z + \frac{1}{2}\right) \cdot \infty. \tag{6.51}$$

The next step is to show that $Z < 0$, which is completely analogous to the proof of Proposition 6.6 and will thus not be repeated here.

Due to $Z < 0$, the term $(K+\theta)Z$ from (6.51) decreases as K increases. Taking into account the control constraint (6.6), we consider the limiting case of

$$K \nearrow \frac{\Sigma + T_s \varsigma^2}{2T_s \Sigma} - \theta,$$

which, for $\Sigma \to \infty$, results in

$$K \nearrow \frac{1}{2T_s} - \theta.$$

Hence,

$$(K+\theta)Z + \frac{1}{2} \searrow \frac{Z}{2T_s} + \frac{1}{2} \quad \text{as} \quad K \to \frac{1}{2T_s} - \theta.$$

Thus, if $\frac{Z}{2T_s} + \frac{1}{2} < 0$, then $\operatorname{sgn}\left((K+\theta)Z + \frac{1}{2}\right) = -1$ for large enough admissible K. In this case, (Σ^e, K^e) cannot be a global minimum, contradicting dissipativity. Analogously to the L^2 and the 2F cost, cf. Remarks 6.9 and 6.18, the product $\bar{\lambda} T_s$ and thus Z is constant in T_s. Hence, due to $Z < 0$, one can always achieve $\frac{Z}{2T_s} + \frac{1}{2} < 0$ for small enough $T_s > 0$. □

The result of Proposition 6.21 is very similar to the L^2 case, see Proposition 6.6. For the W^2 cost, however, the statement depends on the sampling time $T_s > 0$. For instance, one can verify that the OCP (6.46) with parameters

$$\varsigma = 5, \quad \theta = 2, \quad \gamma = 1/4, \quad \text{and} \quad T_s = 1$$

is indeed strictly dissipative with $\lambda^l(z)$.

Of course this does not mean that increasing the sampling time always helps. Consider the following example.

Example 6.22. *Consider* (6.46) *with the parameters*

$$\varsigma = 10, \quad \theta = 2, \quad \gamma = 1/4, \quad \textit{and} \quad T_s = 1/100.$$

We want to construct the modified cost $\tilde{\ell}_{W^2}(\Sigma, K)$*. First, we determine the optimal equilibrium* (Σ^e, K^e) *and the corresponding Lagrange multiplier* $\bar{\lambda}$*. We formulate the Lagrange function associated to* (6.22) *with the* W^2 *stage cost* $\ell_{W^2}(\Sigma, K)$ *and solve the problem numerically. Note from* (6.48) *and* (6.50) *that the interest is in* $Z = 2\bar{\lambda}T_s$ *rather than in* $\bar{\lambda}$*. In particular, the optimal equilibrium is independent of the sampling time* T_s*. We get:*

$$\Sigma^e \approx 10.2393012, \quad K^e \approx 2.8831457, \quad Z \approx -0.070394104.$$

With this, we can construct the modified cost $\tilde{\ell}_{W^2}(\Sigma, K)$*, which is depicted in Figure 6.6. All pairs* (Σ, K) *illustrated in this figure satisfy the constraints* (6.6)*. The white area depicts negative values, i.e., pairs* (Σ, K) *in which* (3.10) *is violated. Thus, (strict) dissipativity does not hold with a linear storage function.*

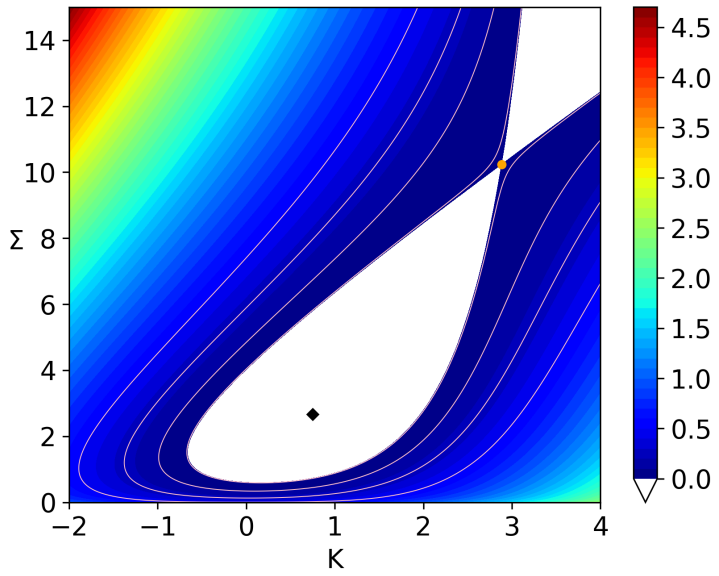


Figure 6.6: Modified cost $\tilde{\ell}_{L^2}(\Sigma, K)$ for Example 6.22. The optimal equilibrium (Σ^e, K^e) is illustrated by the orange circle. The white area represents negative values; the black diamond marks the minimum of the depicted area.

Example 6.22 and the corresponding Figure 6.6 illustrate two reasons why strict dissipativity with $\lambda^l(z)$ does not hold in this example. The first is the asymptotic behavior for $\Sigma \to \infty$, which might be fixed for large enough sampling times T_s. The second reason is the second stationary point of $\tilde{\ell}_{W^2}$, cf. Proposition 6.20. In Example 6.22, it is given by

$$(\Sigma^s, K^s) \approx (2.6621866, 0.749609).$$

It is important to see that the stationary points of $\tilde{\ell}_{W^2}(\Sigma, K)$ depend not on T_s but on Z (see (6.48)), and Z is unaffected by a change in T_s (since $\bar{\lambda}$ also changes accordingly). Likewise, the modified cost $\tilde{\ell}_{W^2}(\Sigma, K)$ itself is unaffected by a change in T_s. Hence, the problem of a second stationary point attaining negative values persists independently of T_s. Moreover, note that in Example 6.22, γ is such that the reduced cost

$$\hat{\ell}_{W^2}(K) := \ell_{W^2}\left(\frac{\varsigma^2}{2(\theta+K)}, K\right) = \frac{1}{2}\left[\left(\sqrt{\frac{\varsigma^2}{2(\theta+K)}} - 1\right)^2 + \gamma K^2\right] \tag{6.52}$$

is strictly convex.[4]

In short, the properties that were used for the 2F cost in Subsection 6.3.2 (see Propositions 6.14 and 6.15 and Remark 6.17) to guarantee strict dissipativity of (6.27) are not appropriate to prove strict dissipativity of (6.46). Instead, a case-by-case analysis is required if $\varsigma^2/2 - \theta > 0$.

The case $\varsigma^2/2 - \theta < 0$

If $\varsigma^2/2 - \theta < 0$, then as in the previous subsections one can show that $Z > 0$. Hence, $\lim_{\Sigma\to\infty} \tilde{\ell}_{W^2}(\Sigma, K) = \infty$, cf. (6.51). Moreover, $\lim_{K\to\infty} \tilde{\ell}_{W^2}(\Sigma, K) = \infty$. However, the two boundaries $\Sigma \searrow 0$ and $K \searrow -\theta$ and the potential second stationary state from Proposition 6.20 need to be checked in order to verify strict dissipativity with a linear storage function. This is the same procedure as for the L^2 cost, cf. Examples 6.7 and 6.8, and requires a case-by-case analysis, as the following two examples demonstrate.

Example 6.23. *Consider* (6.46) *with the parameters*

$$\varsigma = 3/4, \quad \theta = 3/2, \quad \gamma = 1/5, \quad \textit{and} \quad T_s = 1/10.$$

As in Example 6.22, we determine the optimal equilibrium (Σ^e, K^e) *and the associated* Z *numerically:*

$$\Sigma^e \approx 0.4679159, \quad K^e \approx -0.8989304, \quad Z \approx 0.3842274.$$

The reduced cost $\hat{\ell}_{W^2}$*, cf.* (6.52)*, is convex, since* $\frac{5^5}{2^{16}\varsigma^4} = \frac{5^5}{2^8 3^4} < \frac{1}{5} = \gamma$*. Furthermore, the Hessian of the modified cost* $\tilde{\ell}_{W^2}$ *evaluated at* (Σ^e, K^e) *is positive definite:*

$$\nabla^2\tilde{\ell}_{W^2}(\Sigma^e, K^e) \approx \begin{pmatrix} 0.7810671 & Z \\ Z & \gamma \end{pmatrix} \quad \Rightarrow \quad \left|\nabla^2\tilde{\ell}_{W^2}(\Sigma^e, K^e)\right| \approx 0.00858275 > 0.$$

[4] One can show that $\hat{\ell}_{W^2}$ is strictly convex for $\gamma > \frac{5^5}{2^{16}\varsigma^4}$. However, as this fact is not crucial for the subsequent statements we refrain from giving a rigorous proof.

Moreover, the second stationary point of $\tilde{\ell}_{W^2}$ *at approximately*

$$(0.5044150447, -0.9690503190) =: (\Sigma^s, K^s)$$

is not an issue, since $\tilde{\ell}_{W^2}(\Sigma^s, K^s) \approx 9.2315 \cdot 10^{-6} > 0$. *However, we face problems when looking at the boundary* $K = -\theta$ *respective* $\Sigma = 0$*:*

$$\tilde{\ell}_{W^2}(0, K) = \frac{1}{2} + \frac{K^2}{2} - \ell_{W^2}(\Sigma^e, K^e) - \frac{Z\varsigma^2}{2},$$

which is minimal at $K = 0$ *with*

$$\tilde{\ell}_{W^2}(0, 0) = \frac{1}{2} - \ell_{W^2}(\Sigma^e, K^e) - \frac{Z\varsigma^2}{2}.$$

Analogously, at the boundary $K = -\theta$*, we have:*

$$\tilde{\ell}_{W^2}(\Sigma, -\theta) = \frac{1}{2}\left(\sqrt{\Sigma} - 1\right)^2 + \frac{\gamma}{2}\theta^2 - \ell_{W^2}(\Sigma^e, K^e) - \frac{Z\varsigma^2}{2},$$

which is minimal at $\Sigma = 1$ *with*

$$\tilde{\ell}_{W^2}(1, -\theta) = \frac{\gamma}{2}\theta^2 - \ell_{W^2}(\Sigma^e, K^e) - \frac{Z\varsigma^2}{2}.$$

In total, we require that

$$\min\left\{\frac{1}{2}, \frac{\gamma}{2}\theta^2\right\} - \ell_{W^2}(\Sigma^e, K^e) - \frac{Z\varsigma^2}{2} \geq 0. \tag{6.53}$$

Otherwise, due to continuity of $\tilde{\ell}_{W^2}$*, strict dissipativity with this storage function does not hold. Indeed, in this example, we have*

$$\min\left\{\frac{1}{2}, \frac{\gamma}{2}\theta^2\right\} - \ell_{W^2}(\Sigma^e, K^e) - \frac{Z\varsigma^2}{2} \approx -0.0137857 < 0,$$

see Figure 6.7, and thus, no strict dissipativity with $\lambda^l(z)$.

Example 6.24. *Consider* (6.46) *with the parameters*

$$\varsigma = 2/3, \quad \theta = 3/2, \quad \gamma = 1/3, \quad \text{and} \quad T_s = 1/10.$$

We identify the optimal equilibrium and the corresponding value for Z *numerically:*

$$\Sigma^e \approx 0.1865912, \quad K^e \approx -0.3090422, \quad Z \approx 0.5520844.$$

We also determine the second stationary point of $\tilde{\ell}_{W^2}$ *numerically:*

$$(0.8642951, -1.4314914) =: (\Sigma^s, K^s).$$

We get $\tilde{\ell}_{W^2}(\Sigma^s, K^s) \approx 0.07675 > 0$ *and* $\min\left\{\frac{1}{2}, \frac{\gamma}{2}\theta^2\right\} - \ell_{W^2}(\Sigma^e, K^e) - \frac{Z\varsigma^2}{2} \approx 0.075063 > 0$*, cf.* (6.53)*. Hence, both the second stationary point and the boundary yield positive values. Due to* $Z > 0$*,* $\tilde{\ell}_{W^2}(\Sigma, K) \to \infty$ *for* $\Sigma \to \infty$ *or* $K \to \infty$*. Since no other stationary point exists, we can find a function* $\varrho \in \mathcal{K}_\infty$ *such that the dissipativity inequality* (3.10) *holds with* $\lambda^l(z)$*. Figure 6.8 depicts the corresponding modified cost* $\tilde{\ell}_{W^2}(\Sigma, K)$.

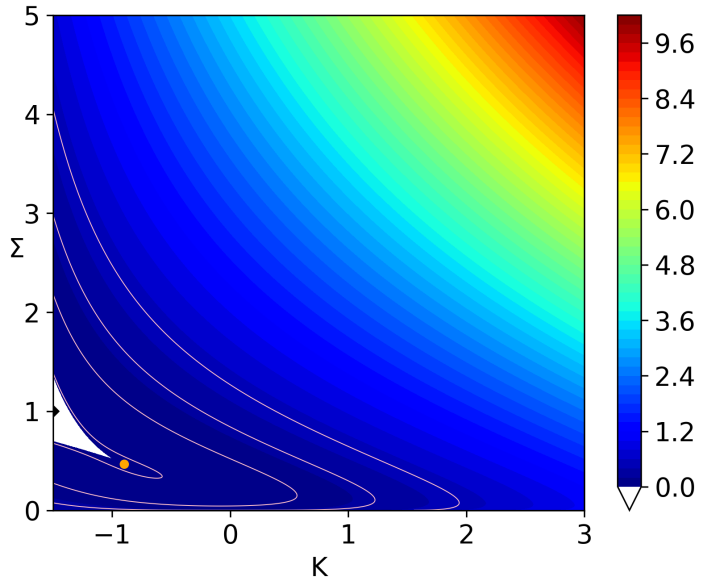


Figure 6.7: Modified cost $\tilde{\ell}_{L^2}(\Sigma, K)$ for Example 6.23. The optimal equilibrium (Σ^e, K^e) is illustrated by the orange circle. The white area represents negative values; the black diamond marks the minimum of the depicted area.

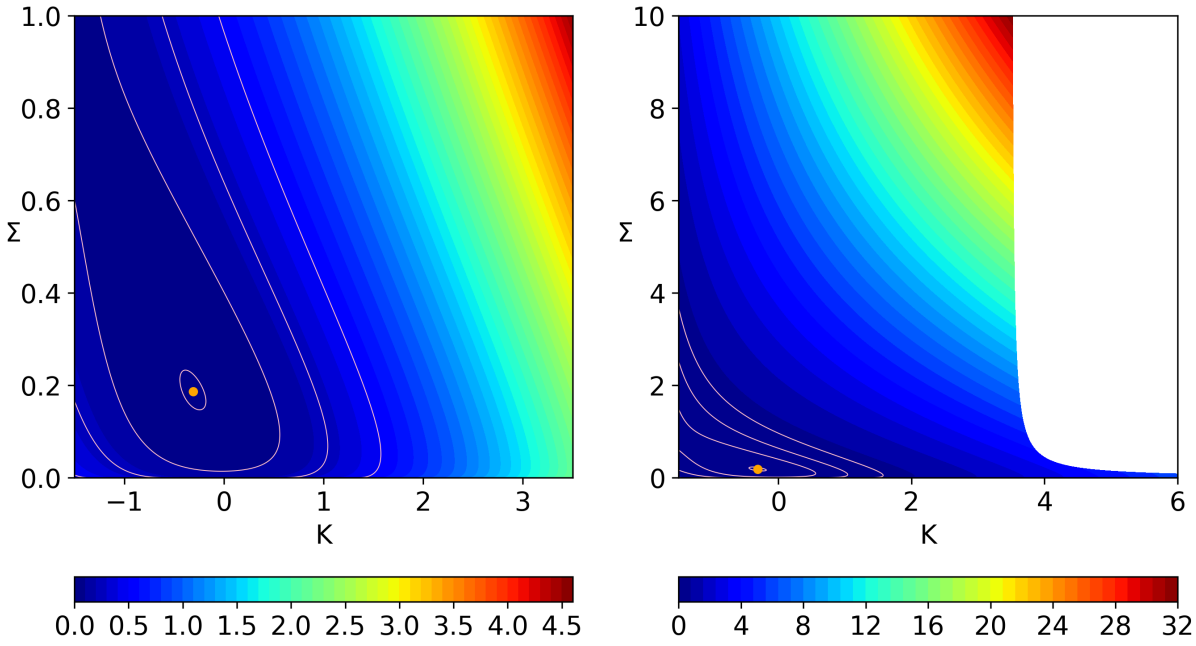


Figure 6.8: Modified cost $\tilde{\ell}_{W^2}(\Sigma, K)$ for Example 6.24 zoomed in (left) and zoomed out (right). The optimal equilibrium (Σ^e, K^e) is illustrated by the orange circle. The white area on the right plot is due to control constraints (6.6).

Modifications to the stage cost ℓ_{W^2}

In this part we discuss the two modifications to the stage cost ℓ_{W^2} that were discussed in the previous two Subsections 6.3.1 and 6.3.2.

Remark 6.25 (Scaling of the stage cost)**.** *For the W^2 cost, scaling the stage cost by a factor T_s as mentioned in Remarks 6.9 and 6.18 could help in verifying strict dissipativity*

using linear storage functions in the case of $\frac{\varsigma^2}{2} - \theta > 0$*, at least for some parameters. Yet, this scaling does not help if a stationary point with a negative function value exists, since it exists independently of* T_s*, requiring a case-by-case analysis still. Hence, analogously to the* L^2 *and the 2F cost, we do not scale the stage cost* $\ell_{W^2}(\Sigma, K)$ *in this subsection.*

Remark 6.26 (Penalizing $(\theta + K)^2$ instead of K^2 in ℓ_{W^2})**.** *Modifying the cost function* ℓ_{W^2} *by penalizing* $(\theta + K)^2$ *instead of* K^2 *does not guarantee strict dissipativity with a linear storage function: Since the modified cost function yields the same optimal equilibria as considering* $\theta = 0$*, in particular,* $\varsigma^2/2 - \theta > 0$ *holds. However, this property does neither guarantee strict dissipativity*[5] *(in contrast to the 2F cost), cf. Example 6.22, nor does it rule out strict dissipativity (in contrast to the* L^2 *cost).*

A nonlinear storage function

Despite the similarity of the two cost functions ℓ_{W^2} and ℓ_{2F}, the results are very different. In fact, regarding dissipativity with the linear storage function $\lambda^l(z)$, the Wasserstein cost ℓ_{W^2} has more in common with the L^2 cost considered in Subsection 6.3.1. This includes that, when running numerical simulations, the MPC closed loop converges to the optimal equilibrium (Σ^e, K^e)—even for the parameters in Examples 6.22 and 6.23, see Figures 6.9 and 6.10. These figures indicate that the turnpike property holds even in cases where the linear storage function fails.

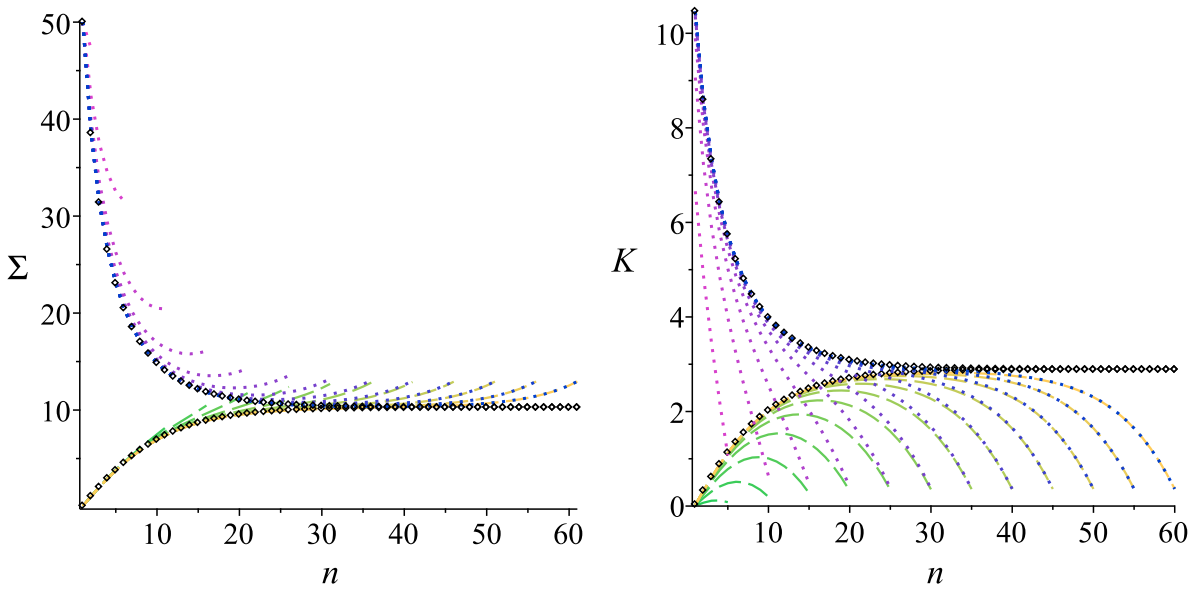


Figure 6.9: Open-loop optimal trajectories for various horizons N between 2 and 61 and MPC closed-loop trajectories for two different initial conditions, indicating turnpike behavior in Example 6.22; state Σ (left) and control K (right).

Due to the close relationship between dissipativity and the turnpike property, see the end of Section 3.3, this strongly suggests that strict dissipativity does indeed hold, but with a nonlinear storage function. Thus, in the rest of this section, we revisit these

[5] Neither does the condition $\varsigma^2/2 - \theta < 0$, see Example 6.23.

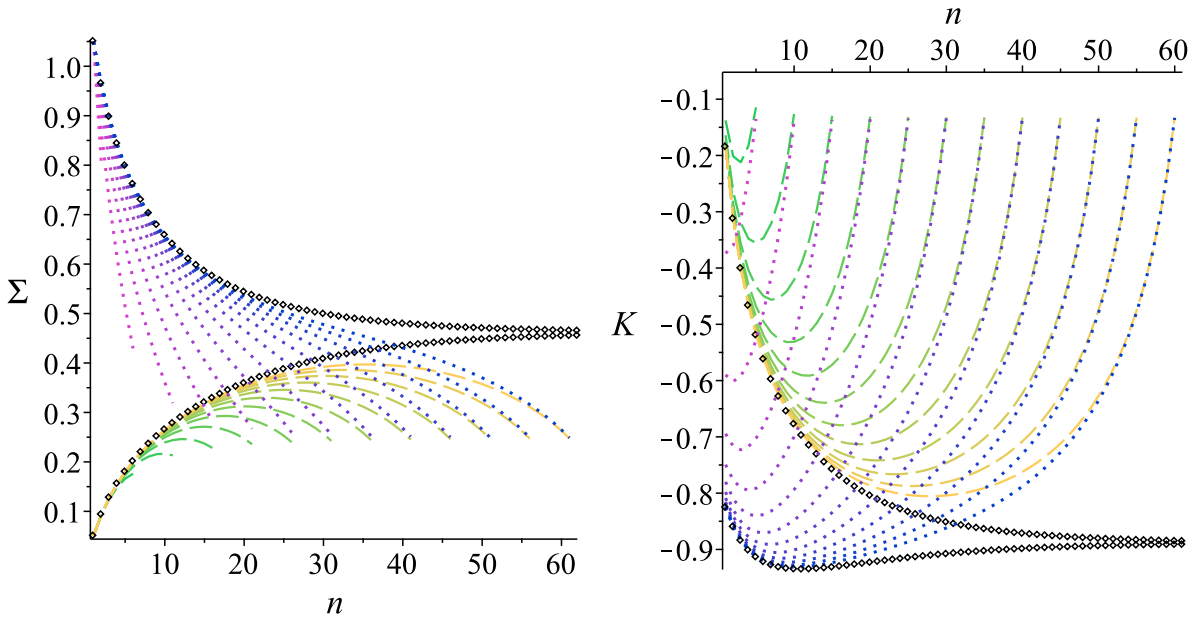


Figure 6.10: Open-loop optimal trajectories for various horizons N between 2 and 61 and MPC closed-loop trajectories for two different initial conditions, indicating turnpike behavior in Example 6.23; state Σ (left) and control K (right).

examples with the nonlinear storage function

$$\lambda^s(z) := \alpha\sqrt{z+1}. \tag{6.54}$$

The parameter $\alpha \in \mathbb{R}$ is chosen such that the optimal equilibrium (Σ^e, K^e) is a stationary point of the new modified cost

$$\tilde{\ell}^s_{W^2}(\Sigma, K) := \ell_{W^2}(\Sigma, K) - \ell_{W^2}(\Sigma^e, K^e) + \lambda^s(\Sigma) - \lambda^s(\Sigma^+).$$

One notable advantage of $\lambda^s(z)$ over $\lambda^l(z)$ is the asymptotic behavior of the modified cost: While

$$\lim_{\Sigma\to\infty} \tilde{\ell}_{W^2}(\Sigma, K) = \operatorname{sgn}\left((K+\theta)Z + \frac{1}{2}\right)\cdot\infty$$

depends on the value of Z, the nonlinear storage function $\lambda^s(z)$ yields $\tilde{\ell}^s_{W^2}(\Sigma, K) \to \infty$ as $\Sigma \to \infty$ or $K \to \infty$ irrespective of the value of α for admissible controls. Thus, when looking for a suitable/promising storage function $\lambda(z)$, the asymptotic behavior of $\lambda(z)$ should be compared to that of the cost $\ell(\Sigma, K)$.

Ideally, the storage function should be chosen such that the Hessian $\nabla^2\tilde{\ell}(\Sigma, K)$ is constant. Then one can avoid checking everything by foot, i.e., the boundary values and the stationary points of the modified cost function. Unfortunately, the Hessian $\nabla^2\tilde{\ell}^s_{W^2}(\Sigma, K)$ is not constant. However, the level sets in Figure 6.11 clearly suggest that strict dissipativity holds for both Examples 6.22 and 6.23. We take a closer look at these.

Let us first consider Example 6.22. Our numerical calculations yield $\alpha \approx -23.5996705$ and three stationary states of $\tilde{\ell}^s_{W^2}(\Sigma, K)$, of which one violates $K > -\theta$. The remaining two are (Σ^e, K^e) and $(\Sigma^s, K^s) \approx (265.4413283, 41.51437144)$. The second one is admissible but not a problem since $\tilde{\ell}^s_{W^2}(\Sigma^s, K^s) \approx 86.1249768 > 0$. At the boundary $\Sigma = 0$, the

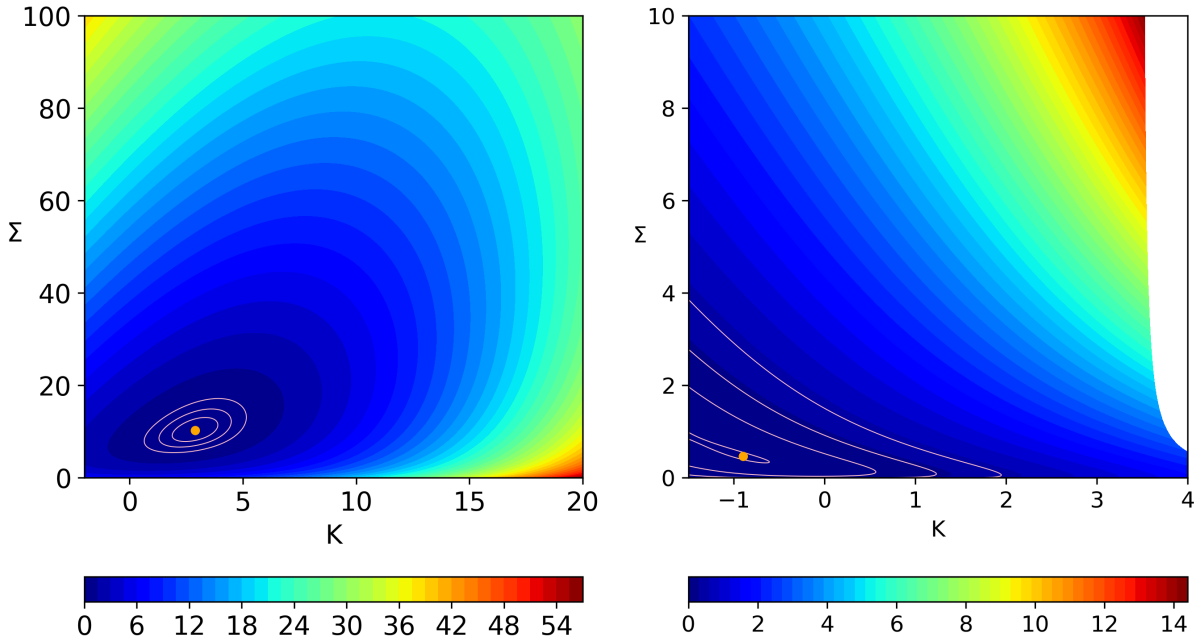


Figure 6.11: New modified cost $\tilde{\ell}^s_{W^2}(\Sigma, K)$ for Examples 6.22 (left) and 6.23 (right). The optimal equilibrium (Σ^e, K^e) is illustrated by the orange circle. The white area on the right plot is due to the control constraints (6.6).

minimum is attained at $K = 0$, with a value of $\tilde{\ell}^s_{W^2}(0, 0) \approx 6.816477628 > 0$. For $K = -\theta$, a minimum of approximately $2.40236824 > 0$ is attained at $\Sigma \approx 5.897079388$. Thus, we can find a function $\varrho \in \mathcal{K}_\infty$ such that the dissipativity inequality (3.10) holds with $\lambda^s(z)$.

Next, we look at Example 6.23. In this case, from numerical calculations we get $\alpha \approx 4.6552057$. In addition to (Σ^e, K^e), the new modified cost $\tilde{\ell}^s_{W^2}(\Sigma, K)$ exhibits a second admissible stationary point at $(\Sigma^s, K^s) \approx (0.8398851754, -1.424465947)$ with $\tilde{\ell}^s_{W^2}(\Sigma^s, K^s) \approx 0.00136419 > 0$. A third one exists but violates the constraint $K > -\theta$. At the boundary $\Sigma = 0$, the minimum is attained at $K = 0$, with a value of $\tilde{\ell}^s_{W^2}(0, 0) \approx 0.2401417173 > 0$. Lastly, for $K = -\theta$, the minimum value is attained at approximately $\Sigma_\theta := 0.9095436678$, with a value of $\tilde{\ell}^s_{W^2}(\Sigma_\theta, -\theta) \approx 0.005474005283 > 0$. Thus, again, we can find a function $\varrho \in \mathcal{K}_\infty$ such that the dissipativity inequality (3.10) holds with $\lambda^s(z)$.

Summary

As in the previous two Subsections 6.3.1 and 6.3.2 we end our analysis by summarizing our main results in short form:

- For $\varsigma^2/2 - \theta > 0$ and small enough sampling times $T_s > 0$, strict dissipativity cannot hold with a linear storage function. For large enough $T_s > 0$ strict dissipativity may hold, but has to be checked on a case-by-case basis.
- In the case $\varsigma^2/2 - \theta < 0$, strict dissipativity with a linear storage function is independent of the sampling time T_s, but has to be checked on a case-by-case basis.
- For various values of θ and ς strict dissipativity holds with the nonlinear storage function (6.54). However, the verification is tedious and must be done on a case-by-case basis.

- Numerical verification of the turnpike property suggests that strict dissipativity holds for many parameters for which the analytical verification is not (yet) possible.

The above examples show that the W^2 cost is more difficult to manage than the 2F stage cost. The most striking difference is that positive definiteness of the Hessian $\nabla^2\tilde{\ell}_{W^2}(\Sigma^e, K^e)$ is not sufficient for strict dissipativity with a linear storage function since $\nabla^2\tilde{\ell}_{W^2}(\Sigma, K)$ is not constant. Although this property can be used to conclude local convexity in a neighborhood of (Σ^e, K^e) (which implies strict dissipativity if state and control are constrained to that region), in general it will not yield global convexity. Another difficulty arises due to the second stationary state of $\tilde{\ell}_{W^2}$ (see Proposition 6.20), for which, on top of that, there is no analytic formula, as opposed to the 2F cost, cf. Lemma 6.12. Moreover, one needs to take into account the boundary, which was unnecessary for the 2F cost due to the constant Hessian $\nabla^2\tilde{\ell}_{2F}(\Sigma, K)$, cf. (6.36). All in all, even though the W^2 cost $\ell_{W^2}(\Sigma, K)$ from (6.17) looks more similar to the 2F cost $\ell_{2F}(\Sigma, K)$ from (6.18) than to the L^2 cost $\ell_{L^2}(\Sigma, K)$ from (6.16), the W^2 cost behaves more like the L^2 cost than like the 2F cost when it comes to analyzing strict dissipativity.

Concluding, the Wasserstein metric, which is in many aspects very suitable for measuring distances of PDFs, does not allow for a simple analysis of strict dissipativity, although our results give strong indication that strict dissipativity holds for many parameter values.

6.3.4 Quick Comparison of L^2, 2F, and W^2 Stage Costs

We summarize and compare the results of the three stage costs from the three previous subsections in Table 6.1.

6.4 Conclusion

In this chapter we have analyzed whether a particular optimal control problem with bilinear dynamics connected to the Fokker–Planck equation is strictly dissipative. We have compared three different cost functions, the L^2 and the W^2 cost, which are suited for being used in a nonlinear setting for general PDFs, and the 2F cost, which is based on quadratic cost functions commonly used in tracking objectives but adjusted to the Gaussian setting.

We have found that for the 2F cost, a linear storage function can be used to prove strict dissipativity for a large parameter set. The linear storage function is convenient due to its close connection to the Lagrange function associated with the problem of finding optimal equilibria. We have also demonstrated that for large sets of parameters, a linear storage function is unsuitable if the L^2 or the W^2 cost are used. To show that the optimal control problems are strictly dissipative in these situations, we have introduced appropriate classes of nonlinear storage functions.

Unfortunately, the 2F cost is the only one that is not derived from a metric for general PDFs, and thus it is only applicable to the Gaussian setting. It will be an interesting question for further research to see whether it is possible to extend this cost and the associated strict dissipativity results beyond the Gaussian case, or whether a class of nonlinear storage functions that captures a large parameter set exists.

	L^2 cost	2F cost	W^2 cost
State cost $(\ell - \frac{\gamma}{2}K^2)$	$\frac{1}{4\sqrt{\pi}}\left[\Sigma^{-1/2}+1 \right.$ $\left.-2\sqrt{2}(\Sigma+1)^{-1/2}\right]$	$\frac{1}{2}(\Sigma-1)^2$	$\frac{1}{2}(\sqrt{\Sigma}-1)^2$
Hessian $\nabla^2\tilde{\ell}$	not constant; indefinite for $\Sigma > 2^{2/5}/(2-2^{2/5})$	constant; positive definite in all tested non-degenerate cases; positive semi-definite otherwise	not constant; indefinite for sufficiently large Σ
The case $\varsigma^2/2-\theta > 0$	not strictly dissipative with $\lambda^l(z)$	strictly dissipative with $\lambda^l(z)$	not strictly dissipative with $\lambda^l(z)$ for sufficiently small T_s; otherwise requires case-by-case analysis
The case $\varsigma^2/2-\theta < 0$	requires case-by-case analysis	strictly dissipative with $\lambda^l(z)$ for $\gamma > 1/(4\varsigma^4)$; presumably strictly dissipative in general if $\gamma \neq 1/\theta^2$	requires case-by-case analysis
Stationary points of $\tilde{\ell}$	no explicit formula; up to two	explicit formula; unique if $Z \neq 0$, infinitely many otherwise (degenerate case)	no explicit formula; up to two
Scaling of ℓ with T_s	does not affect results	does not affect results	may affect results in the case $\varsigma^2/2 - \theta > 0$
Penalizing $(\theta + K)^2$ instead of K^2 in ℓ	leads to the case $\varsigma^2/2 - \theta > 0$	leads to the case $\varsigma^2/2 - \theta > 0$	leads to the case $\varsigma^2/2 - \theta > 0$
Nonlinear storage function $\lambda^s(z)$	can be used to prove strict dissipativity for certain parameters in cases where $\lambda^l(z)$ fails; $\lambda^s(z) = (z+1)^{-1/2}\alpha$ or $\lambda^s(z) = (z+1)^{-1/2}\alpha + \beta z$	presumably not required due to preliminary supporting evidence	can be used to prove strict dissipativity for certain parameters in cases where $\lambda^l(z)$ fails; $\lambda^s(z) = (z+1)^{1/2}\alpha$

Table 6.1: Summary and comparison of the results of Subsections 6.3.1, 6.3.2, and 6.3.3.

Numerical Implementation and Simulations

7

The numerical implementation is written mainly in `C++` [91] and consists of the three parts, `PDE-MPC`, `OU-MPC`, and `SDEControl`, which are described in their respective sections below. In addition, `Python` [97] scripts were used to create plots (e.g., in Chapter 6) and `Maple` [70] scripts were used for plots and symbolic computation. At the end of this chapter, in Section 7.4, we present additional numerical examples that might be of interest, but were not discussed in the previous chapters.

7.1 PDE-MPC

The `C++` program `PDE-MPC` was written to numerically solve PDE-constrained optimal control problems via MPC. It is able to deal with the Fokker–Planck equation (1.2) including zero-flux boundary conditions. The discretization of the Fokker–Planck equation is written in a way that does not limit the spatial dimension d, i.e., an arbitrarily large system of stochastic processes can be considered from the macroscopic point of view via the associated PDF—the limiting factor is the computer memory and the computation time. This program does not rely on any external libraries. Moreover, its modular structure allows to easily consider other PDEs and different solvers within the MPC framework.

In the following, the emphasis is on the structure of `PDE-MPC` and on the program flow. More low-level details can be found in the documentation provided together with the source code. An overview of the program structure is shown in Figure 7.1. All currently existing classes and their subclasses in the program are listed on the left, with their main purpose(s) on the right.

At the top is the `MPC` class, which implements the MPC Algorithm 3.1. In every MPC step we have to solve (OCP_N), an optimal control problem subject to (discrete-time) dynamics, on a certain time interval. This is the main task of the `Optimizer` class. It being an abstract class allows the user to choose any desired nonlinear programming (NLP) solver without changing the MPC code, i.e, at run time. Currently, the Projected Gradient descent algorithm [95] as well as the BFGS algorithm [73] are implemented in `ProjectedGradient` and `BFGS`, both subclasses of `Optimizer`. Both methods require information about the gradient of the (reduced) cost functional, which, in the PDE setting, is usually obtained by solving the associated adjoint PDE (see the derivation of the first order necessary optimality conditions in Section 2.5).

One possible approach to solve these PDEs is to first discretize the spatial domain, e.g., by Finite Differences, while the time is kept continuous. This is the approach we follow in the implementation. For the spatial discretization of the Fokker–Planck equation with zero-flux boundary conditions we implemented the Chang–Cooper scheme [21], a Finite

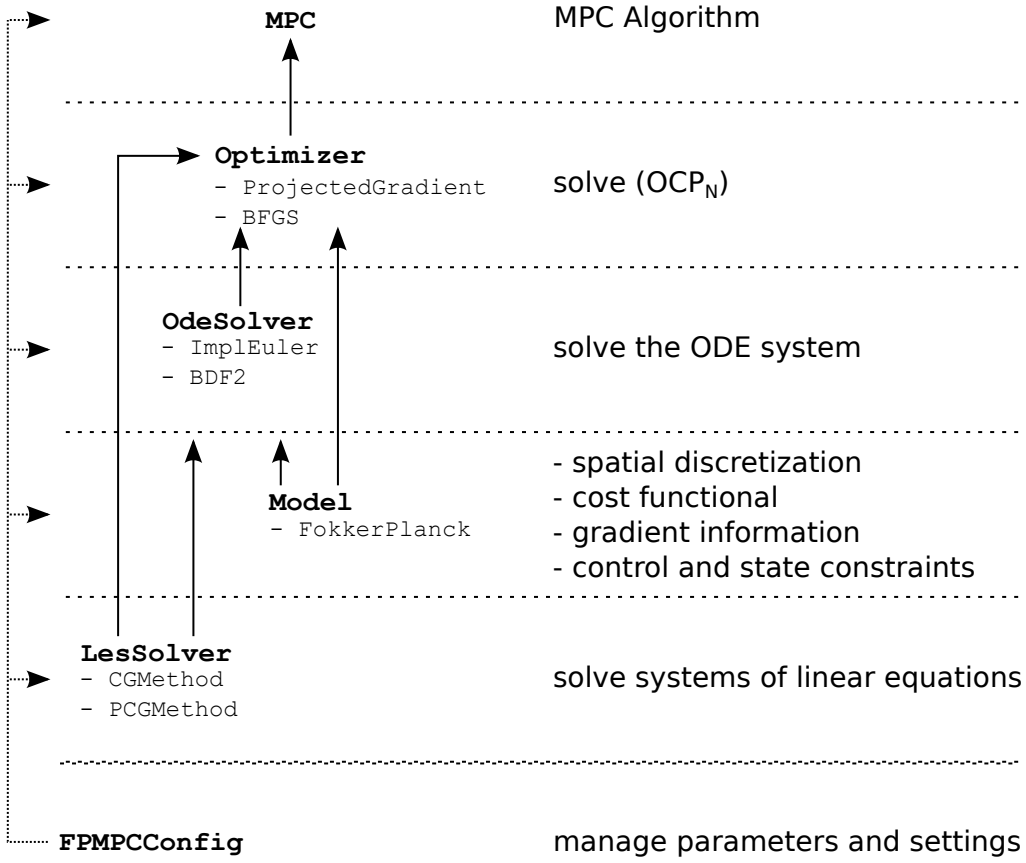


Figure 7.1: Program structure.

Difference scheme that, by design, guarantees non-negativity and conservation of mass, the two crucial properties of a PDF. This is done in the `FokkerPlanck` class, a subclass of the abstract `Model` class, and results in a (large) system of ordinary differential equations.

The `OdeSolver` class is responsible for solving these ODE systems. Like `Optimizer` and `Model` it is an abstract class. Currently, the implicit (backward) Euler method and the BDF2 algorithm [87] are implemented in `ImplEuler` and `BDF2`, respectively. In particular, we can combine the Chang–Cooper scheme and the BDF2 scheme to discretize the Fokker–Planck equation in space and time, respectively, which yields an approximation of second order, cf. [72]. The adjoint PDE is discretized in time with the same method as the forward Fokker–Planck equation. For the discretization in space we follow the discretize-before-optimize approach from [5, p. 497f], i.e., the matrix that represents the spatial discretization of the forward Fokker–Planck equation is being transposed for the backward (adjoint) equation.

The BDF2, the implicit Euler, and the BFGS algorithms require solving systems of linear equations, for which the `LesSolver` abstract class exists. Currently, the Conjugate Gradient (CG) method [57] is implemented in `CGMethod`. Additionally, `PCGMethod`, a preconditioned version of the CG method, is added for testing purposes, where, given a linear system $Ax = b$ with $A \in \mathbb{R}^{d\times d}$, the preconditioner matrix equals the diagonal matrix D^{-1} with $D_{ii} = A_{ii}$, $i = 1, ..., d$.

In addition to the spatial discretization, the `Model` class respective the `FokkerPlanck` class encompasses the evaluation of the cost functional, associated gradient information,

and control and state constraints. All occurring L^2 norms and all integrals are computed by either trapezoidal or rectangle formulas.

Lastly, slightly separated from the other classes is the `FPMPCConfig` class at the bottom since, strictly speaking, it is not necessary for the computations. Yet, it is provided for convenience and acts as a central access point to set the 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. Only few "expert" parameters, e.g., the step size in the optimization algorithm, need to be adjusted directly in the respective class.

The program generates several files, in which it stores the results, e.g., the state y, the control u, the adjoint state p, and the costs. For more details we refer to Table 7.1.

Filename	Description
adjTrajectory.txt/.vtr	Contains the values of the adjoint state p on the mesh. For $d \geq 2$ a sequence of `vtr` files is generated, one file per MPC time step.
config-detailed.txt	Stores the relevant configuration parameters in a more readable form.
config-multidim.txt	Stores the relevant configuration parameters for further use.
control.txt/.vtr	Contains the control values u on the mesh. For $d \geq 2$ a sequence of `vtr` files is generated, one file per control and per MPC time step.
controlBounds.txt	Contains the lower and upper control bounds.
controlcosts.txt	Stores the control costs, which are a part of the stage costs, for each implemented MPC step.
desiredTrajectory.txt/.vtr	Contains the values of the desired state on the mesh. For $d \geq 2$ a sequence of `vtr` files is generated, one file per MPC time 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/.vtr	Contains the values of the state y on the mesh. For $d \geq 2$ a sequence of `vtr` files is generated, one file per MPC time step.
xCoords.txt	This file is only created in the one-dimensional case and contains the spatial discretization of $\Omega \subset \mathbb{R}$. For $d \geq 2$ this information is stored in the respective `vtr` files.

Table 7.1: Files generated by `PDE-MPC`.

7.2 OU-MPC

A recurrent stochastic process in this thesis is the Ornstein–Uhlenbeck process. If the control is space-independent (Chapter 4) or linear in space (Chapters 5 and 6) and assuming the initial PDF is Gaussian, then the associated PDF stays Gaussian and its evolution is entirely prescribed by the evolution of the mean and the covariance matrix, i.e., by the ODE system (5.5). Moreover, assuming the control is piecewise constant in time, the solution of the Fokker–Planck equation associated with the Ornstein–Uhlenbeck process exists in closed form, cf. Section 4.1 and Example 5.1.

The `C++` program `OU-MPC` takes advantage of this, i.e., numerical errors in the discretization are eliminated by using the closed form solution. In addition, the user can switch to the discrete-time dynamics (6.5), a forward Euler approximation of the ODE system (5.5), which is implemented to numerically solve the OCPs considered in Chapter 6. Both the closed form solution and the Euler approximation are implemented to handle arbitrary dimensions $d \in \mathbb{N}$. A boolean variable `euler` is used to switch between the two dynamics. Additional boolean switches, e.g., `economicCost` and `wasserstein`, allow the user to choose which stage cost function to use, i.e., switch between the stabilizing MPC case (Chapters 4 and 5) and the economic MPC case (Chapter 6), and switch between the L^2 and the W^2 cost in the state penalization terms, cf. (6.8) and (6.9).

The program relies on the header-only library `CppOptimizationLibrary` [100], which in turn relies on `Eigen`, a "`C++` template library for linear algebra" [54]. The optimization library `CppOptimizationLibrary` was slightly modified to incorporate the projected gradient descent solver. The gradient of the objective function can be approximated numerically by the optimization library. However, for the L^2 stage cost and the shortest possible MPC horizon we implemented the exact gradient to get more accurate solutions.

In addition to optimization via MPC, it is possible to run simulations with a given control, e.g., with the equilibrium control $\bar{u}$ in Examples 5.14, 5.16 and 5.19, or to compute and store open-loop optimal trajectories, e.g., to illustrate the turnpike property in Figures 6.2, 6.9, and 6.10. The settings for the aforementioned examples and for the numerical simulations in Section 4.3 are provided in the source code for convenience. The results are stored in several files; for more details we refer to Table 7.2.

7.3 SDEControl

The `C++` program `SDEControl` was written to return from the macroscopic perspective to the underlying stochastic process at hand. It numerically solves stochastic ODEs with a given control/input u using the Euler–Maruyama method [62]. The main purpose is to verify the results obtained by the Fokker–Planck approach on the microscopic level. This program does not rely on any external libraries.

The control and the configuration details (such as the stochastic process at hand, the parameters for the initial distribution, and the control dimension) are read from the corresponding files generated by `PDE-MPC`, cf. Table 7.1. The output is a sequence of `csv` files, one per time step, that contains the state (the position) of all simulated "particles" subject to the stochastic process at hand.

Filename	Description
<u>_outControl.txt	Contains the control value $u = (K, c)$ for each implemented MPC step.
<u>_outCosts.txt	Contains the stage costs for each implemented MPC step.
<u>_outMeanVarianceDiff.txt	Contains the differences $\|\mu - \bar{\mu}\|_2^2$ and $\left\|\Sigma - \bar{\Sigma}\right\|_F^2$ for each implemented MPC step. These are used, e.g., in Figures 5.1 and 5.2.
<u>_outMu0.txt	Contains the state μ (mean) for each implemented MPC step.
<u>_outSigmaSq0.txt	Contains the state Σ (covariance matrix) for each implemented MPC step.
outParams.txt	Stores the relevant configuration parameters in a more readable form.
outParamsMaple.txt	Stores the relevant configuration parameters for further use in Maple.

Table 7.2: Files generated by `OU-MPC`. The prefix <u> is used to distinguish between the optimal control u^* (`uOpt`) and the equilibrium control $\bar{u}$ (`uTarget`).

7.4 Additional Numerical Examples

In this section we present some additional numerical examples where, to our knowledge, there is no known closed form solution. The Fokker–Planck optimal control problems were solved in `PDE-MPC`. We verified the results of the Fokker–Planck approach on the microscopic level by solving the SDE numerically in `SDEControl` with the (optimal) controls calculated by `PDE-MPC`. The plots in this section were created in ParaView [11].

Example 7.1 (Shallow Water)**.** *The following two-dimensional stochastic process models the dispersion of substance in shallow water [56]: Consider* (1.1) *with*

$$\tilde{a}(x,t) := \begin{pmatrix} \sqrt{2h(x)} & 0 \\ 0 & \sqrt{2h(x)} \end{pmatrix} \quad \text{and} \quad b(x,t;u) := \begin{pmatrix} u_1(x,t) - x_1/32 + 1/40 \\ u_2(x,t) - x_2/32 + 1/40 \end{pmatrix}$$

and the associated Fokker–Planck equation on $Q := \Omega \times [0,5]$ *with* $\Omega :=]0,8[^2$*, where* $h(x) := -\frac{1}{64}\left((x_1-4)^2 + (x_2-4)^2\right) + \frac{3}{5} \geq 0$ *in* Ω*. The spatial domain* Ω *is discretized using a* 321×321 *grid, which results in* $2 \cdot 102720$ *control variables in the case of space-dependent control.*

To experiment with non-Gaussian PDFs and how well they can be attained with space-dependent controls in various settings, the initial PDF is a (smoothed) delta-Dirac located at the center $(4,4)$ *and we choose the target PDF*

$$\bar{\rho}(x) := m \left[\frac{1}{x_1} \exp\left(\frac{2C_1}{\sigma^2}\log(x_1)\right) - \frac{2}{\sigma^2}(x_1 - 1)\right] \left[\frac{1}{x_2} \exp\left(\frac{2C_2}{\sigma^2}\log(x_2)\right) - \frac{2}{\sigma^2}(x_2 - 1)\right]$$

with $C_1 = 2.625$*,* $C_2 = 2.125$*,* $\sigma = 0.5$*, and* $m \approx 0.00004591595108$*, which is an equilibrium PDF of a stochastic Lotka-Volterra two-species prey-predator model [103].*

We employ control constraints $u_1, u_2 \in [-10, 10]$ *and solve the optimal control problem using MPC with the shortest possible horizon* $N = 2$*, a sampling time of* $T_s = 0.5$*, and the* L^2 *stage cost*

$$\ell(\rho(k), u(k)) = \frac{1}{2} \left\| \rho(k) - \bar{\rho} \right\|^2_{L^2(\mathbb{R}^2)} + \frac{\gamma}{2} \left\| u(k) \right\|^2_{L^2(\mathbb{R}^2;\mathbb{R}^2)},$$

where $\gamma = 0.001$*. The solution and the evolution of the stochastic process on a microscopic level (100000 paths) are depicted in Figure 7.2, with the corresponding controls in Figure 7.3.*

In the case of space-independent control $(u_1(t), u_2(t))$*, such as illustrated in Figure 2.1, the stage cost is given by*

$$\ell(\rho(k), u(k)) = \frac{1}{2} \left\| \rho(k) - \bar{\rho} \right\|^2_{L^2(\mathbb{R}^2)} + \frac{\gamma}{2} \left| u(k) \right|^2.$$

Example 7.2 (Bimodal Target)**.** *Consider a 2D Ornstein–Uhlenbeck process, i.e.,* (1.1) *with*

$$\tilde{a}(x,t) := \begin{pmatrix} 1/2 & 0 \\ 0 & 1/2 \end{pmatrix} \quad \text{and} \quad b(x,t;u) := \begin{pmatrix} u_1(x,t) - \nu x_1 \\ u_2(x,t) - \nu x_2 \end{pmatrix} \text{ with } \nu := \frac{3}{4},$$

and the associated Fokker–Planck equation on $Q := \Omega \times [0,5]$ *with* $\Omega := \left]-5,5\right[^2$*. The spatial domain* Ω *is discretized using a* 301×301 *grid, which results in* $2 \cdot 90300$ *control variables.*

We consider bimodal PDFs of the form

$$\tilde{\rho}(x;\eta) := \frac{1}{2} \frac{\exp\left(-\frac{(x_1+2\eta)^2}{2\bar{\sigma}_1^2} - \frac{(x_2-2\eta)^2}{2\bar{\sigma}_2^2} \right)}{2\pi\bar{\sigma}_1\bar{\sigma}_2} + \frac{1}{2} \frac{\exp\left(-\frac{(x_1-2\eta)^2}{2\bar{\sigma}_3^2} - \frac{(x_2+2\eta)^2}{2\bar{\sigma}_4^2} \right)}{2\pi\bar{\sigma}_3\bar{\sigma}_4},$$

where $\bar{\sigma} := (0.4, 0.4, 0.6, 0.6)$*. Starting from the initial PDF* $\mathring{\rho}(x) := \tilde{\rho}(x;0)$*, we want the PDF* $\rho(x,t)$ *to attain the target PDF* $\bar{\rho}(x) := \tilde{\rho}(x;1)$*.*

We employ control constraints $u_1, u_2 \in [-10, 10]$ *and we solve the optimal control problem using MPC with the shortest possible horizon* $N = 2$*, a sampling time of* $T_s = 0.5$*, and the* L^2 *stage cost*

$$\ell(\rho(k), u(k)) = \frac{1}{2} \left\| \rho(k) - \bar{\rho} \right\|^2_{L^2(\mathbb{R}^2)} + \frac{\gamma}{2} \left\| u(k) \right\|^2_{L^2(\mathbb{R}^2;\mathbb{R}^2)},$$

where $\gamma = 0.001$*. The solution and the evolution of the stochastic process on a microscopic level (100000 paths) are depicted in Figure 7.4, with the corresponding controls in Figure 7.5.*

Example 7.3 (Moving Bimodal Target)**.** *Consider Example 7.2 with the only change being a moving bimodal target PDF*

$$\begin{aligned} \bar{\rho}(x,t) := {} & \frac{1}{2} \frac{\exp\left(-\frac{[x_1+2\sin(\pi t/5)]^2}{2\bar{\sigma}_1^2} - \frac{[x_2-2\sin(\pi t/5)]^2}{2\bar{\sigma}_2^2} \right)}{2\pi\bar{\sigma}_1\bar{\sigma}_2} \\ & + \frac{1}{2} \frac{\exp\left(-\frac{[x_1-2\sin(\pi t/5)]^2}{2\bar{\sigma}_3^2} - \frac{[x_2+2\sin(\pi t/5)]^2}{2\bar{\sigma}_4^2} \right)}{2\pi\bar{\sigma}_3\bar{\sigma}_4}, \end{aligned}$$

where $\bar{\sigma} := (0.4, 0.4, 0.6, 0.6)$.

The solution is depicted in Figure 7.6, with the corresponding controls in Figure 7.7. Figure 7.8 illustrates the evolution of the stochastic process on a microscopic level (100000 paths).

Example 7.4 (Bimodal Uniform Target)**.** *Consider the 2D Ornstein–Uhlenbeck process from Example 7.2, i.e.,*

$$\tilde{a}(x,t) := \begin{pmatrix} 1/2 & 0 \\ 0 & 1/2 \end{pmatrix} \quad \text{and} \quad b(x,t;u) := \begin{pmatrix} u_1(x,t) - \nu x_1 \\ u_2(x,t) - \nu x_2 \end{pmatrix} \text{ with } \nu := \frac{3}{4},$$

and the associated Fokker–Planck equation on $Q := \Omega \times [0,1]$ *with* $\Omega :=]-3,3[^2$. *The spatial domain* Ω *is discretized using a* 121×121 *grid, which results in* $2 \cdot 14520$ *control variables.*

Starting from the initial Gaussian PDF

$$\mathring{\rho}(x) := \left((2\pi)^2 \prod_{i=1}^{2} \mathring{\sigma}_i^2 \right)^{-1/2} \exp\left(-\sum_{i=1}^{2} \frac{x_i^2}{2\mathring{\sigma}_i^2} \right)$$

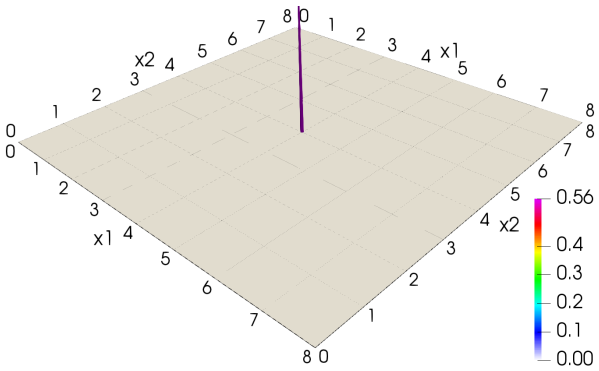
with $\mathring{\sigma}_1^2 = \mathring{\sigma}_2^2 = 1/6$, *we want the PDF* $\rho(x,t)$ *to attain the uniform target PDF*

$$\bar{\rho}(x) := \begin{cases} 1, & x \in [-1,0]^2 \cup [0,1]^2, \\ 0, & \text{otherwise.} \end{cases}$$

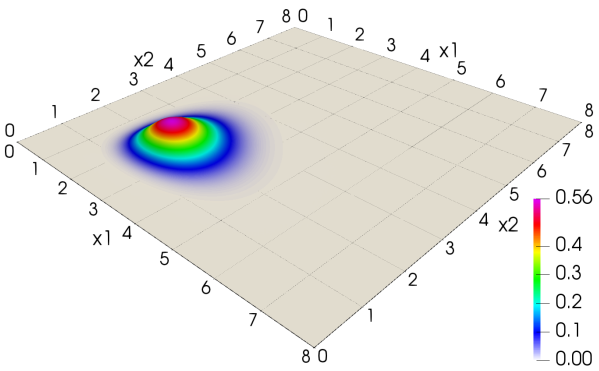
We solve the optimal control problem using MPC with the shortest possible horizon $N = 2$, *a sampling time of* $T_s = 0.1$, *and the* L^2 *stage cost*

$$\ell(\rho(k), u(k)) = \frac{\alpha}{2} \left\| \rho(k) - \bar{\rho} \right\|_{L^2(\mathbb{R}^2)}^2 + \frac{\gamma}{2} \left\| u(k) \right\|_{L^2(\mathbb{R}^2;\mathbb{R}^2)}^2,$$

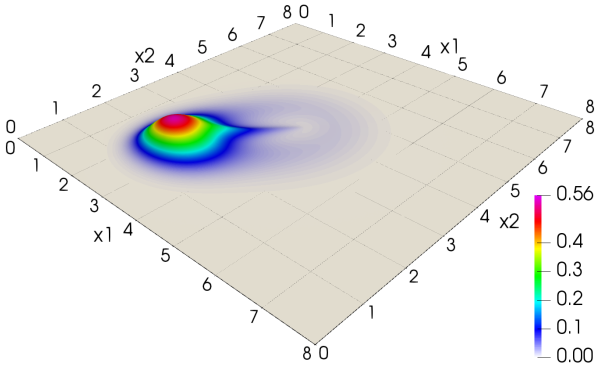
where $\alpha = 2$ *and* γ *is either* 0.01 *or* 0.001. *The solutions are depicted in Figure 7.9, with the corresponding controls in Figure 7.10. Figure 7.11 illustrates the evolution of the stochastic process on a microscopic level.*



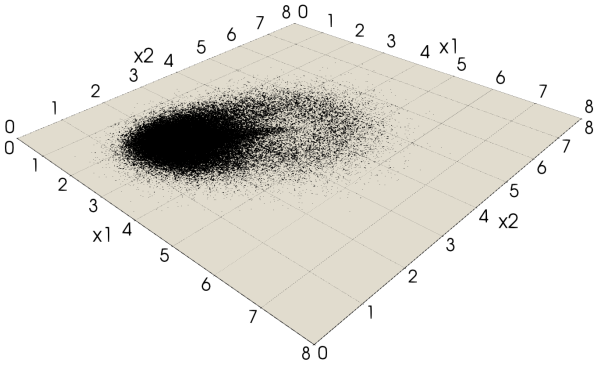
(a) Initial PDF (concentration at $(4, 4)$).



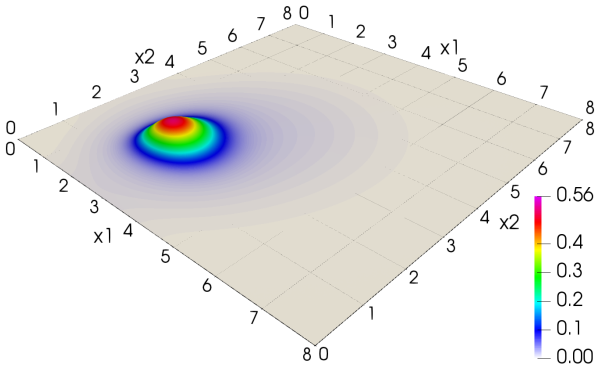
(b) Desired PDF.



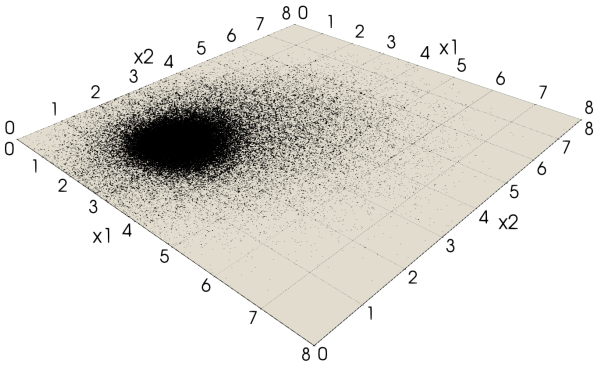
(c) Controlled PDF at $t = 0.5$.



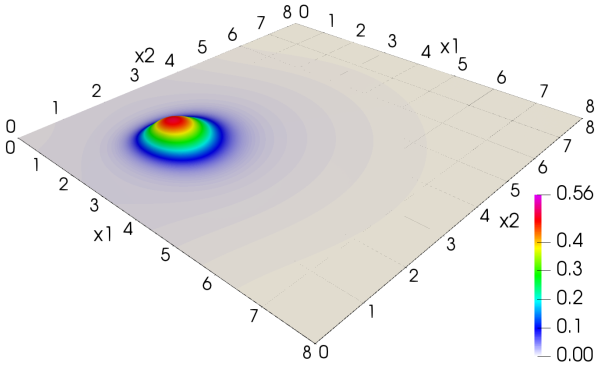
(d) Controlled SDE at $t = 0.5$.



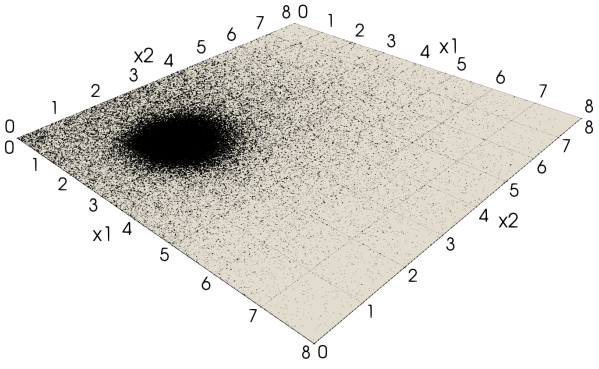
(e) Controlled PDF at $t = 1.5$.



(f) Controlled SDE at $t = 1.5$.

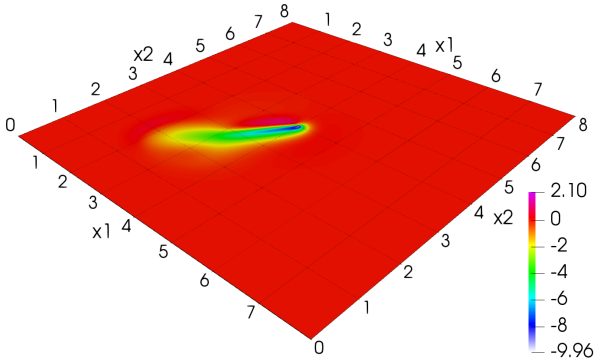


(g) Controlled PDF at $t = 5$.

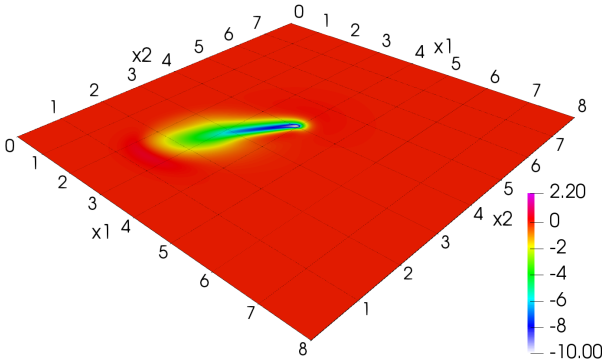


(h) Controlled SDE at $t = 5$.

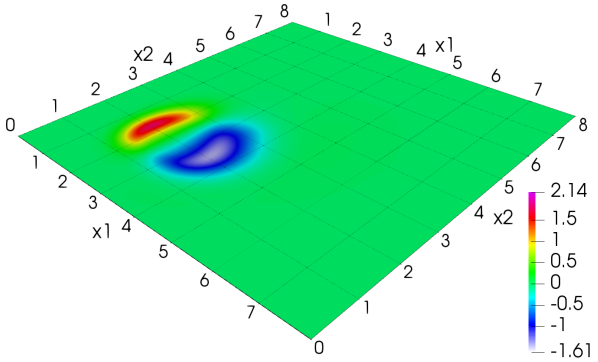
Figure 7.2: Desired and controlled PDF (using space-dependent control) for Example 7.1 (Shallow Water) and the evolution of the stochastic process on a microscopic level.



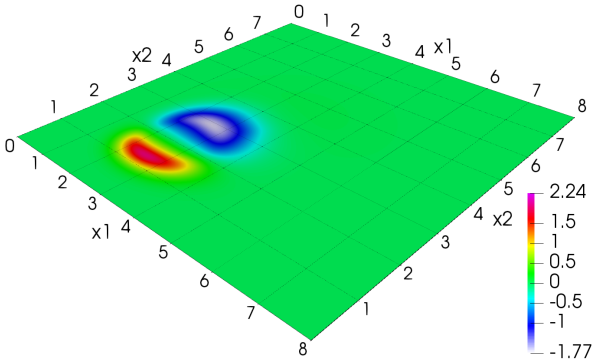
(a) Control $u_1(x, 0)$.



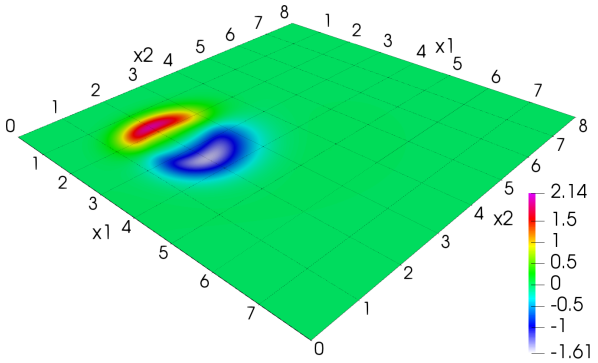
(b) Control $u_2(x, 0)$.



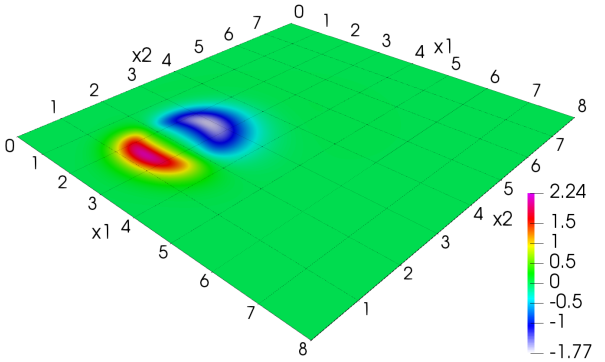
(c) Control $u_1(x, 0.5)$.



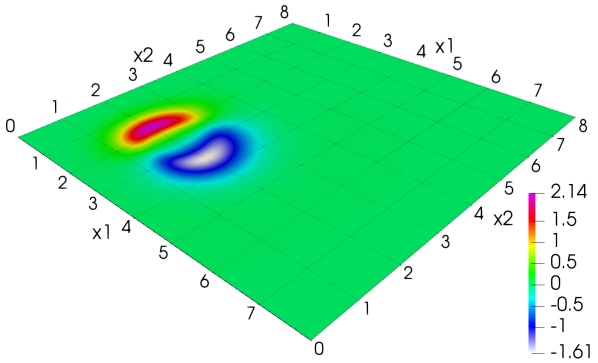
(d) Control $u_2(x, 0.5)$.



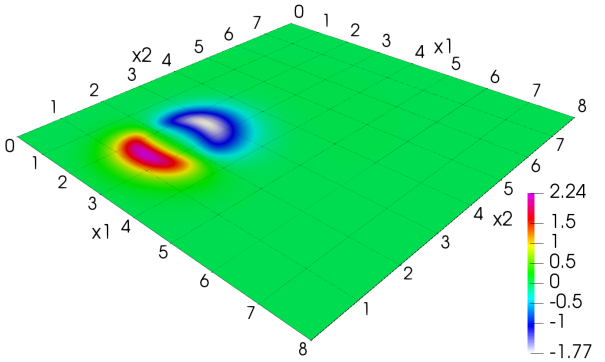
(e) Control $u_1(x, 1)$.



(f) Control $u_2(x, 1)$.

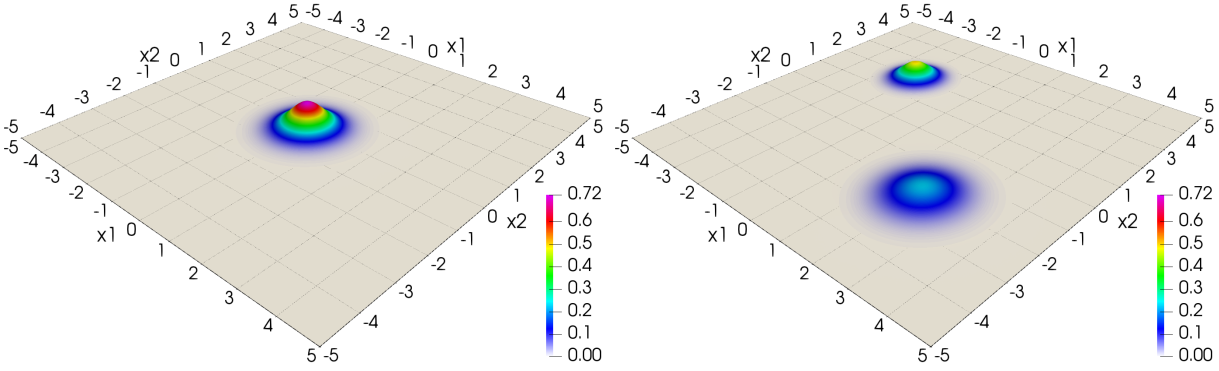


(g) Control $u_1(x, 4.5)$.



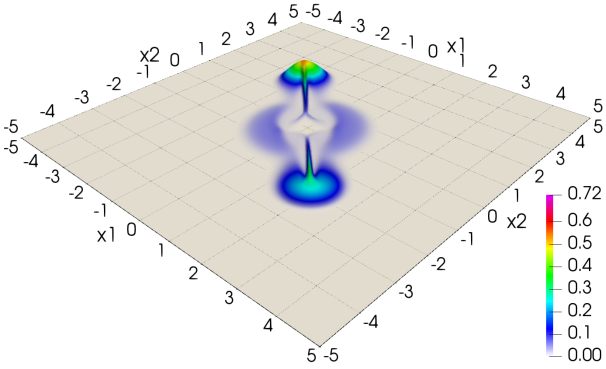
(h) Control $u_2(x, 4.5)$.

Figure 7.3: Controls $u_1(x, t)$ and $u_2(x, t)$ for Example 7.1 (Shallow Water). Note the different scales at $t = 0$.

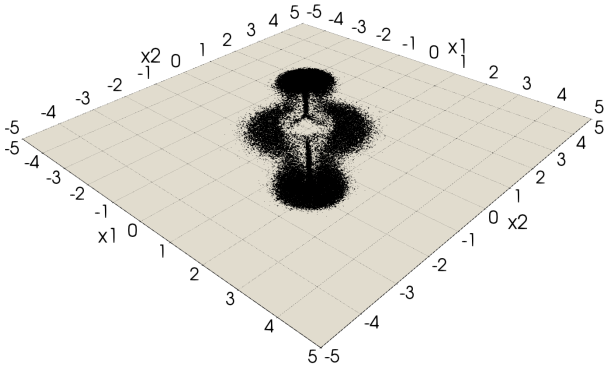


(a) Initial PDF.

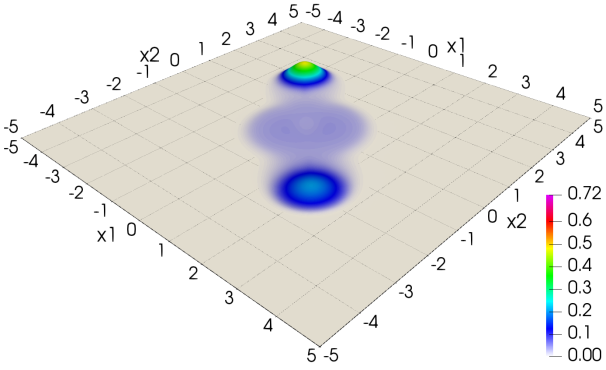
(b) Desired PDF.



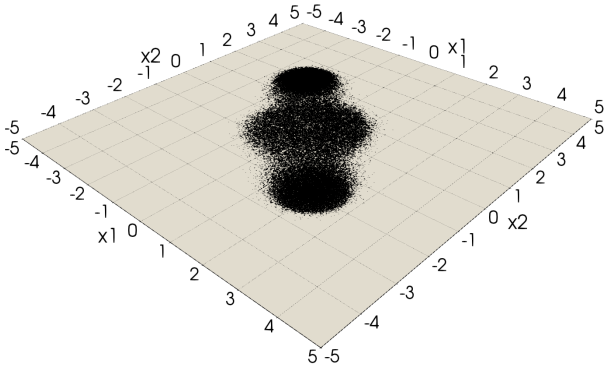
(c) Controlled PDF at $t = 0.5$.



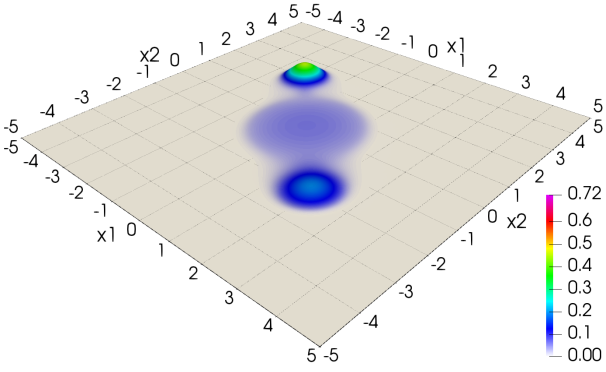
(d) Controlled SDE at $t = 0.5$.



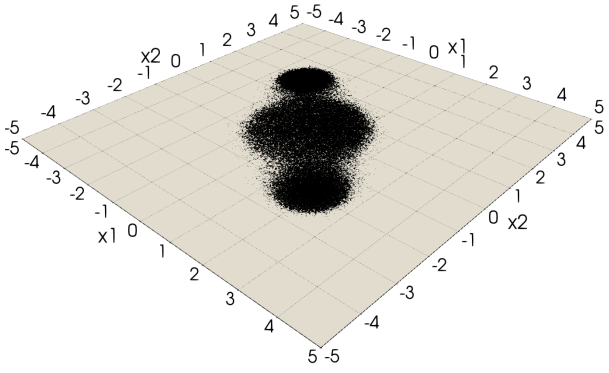
(e) Controlled PDF at $t = 1$.



(f) Controlled SDE at $t = 1$.

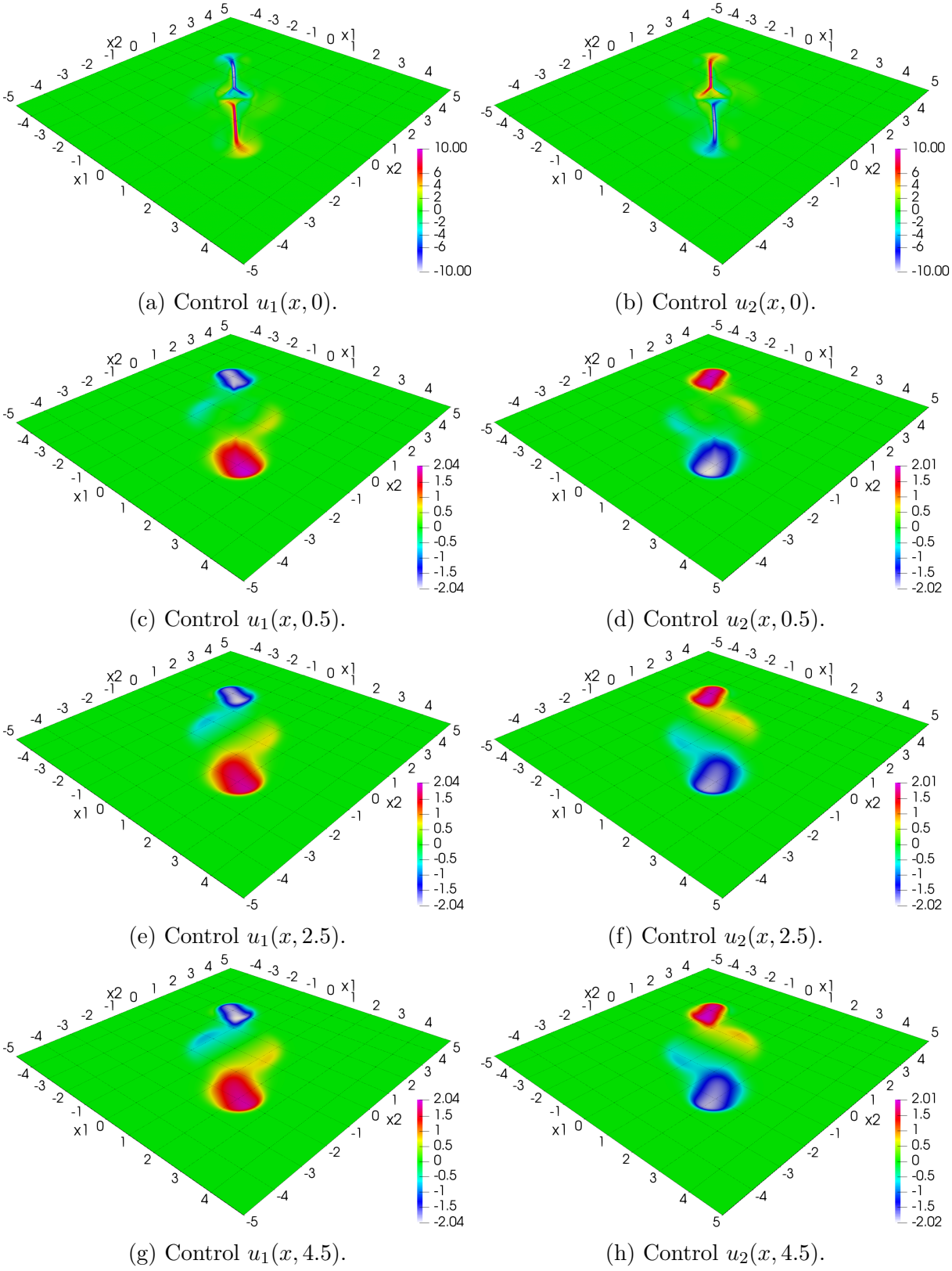


(g) Controlled PDF at $t = 5$.



(h) Controlled SDE at $t = 5$.

Figure 7.4: Desired and controlled PDF for Example 7.2 (Bimodal Target) and the evolution of the stochastic process on a microscopic level.



(a) Control $u_1(x, 0)$.

(b) Control $u_2(x, 0)$.

(c) Control $u_1(x, 0.5)$.

(d) Control $u_2(x, 0.5)$.

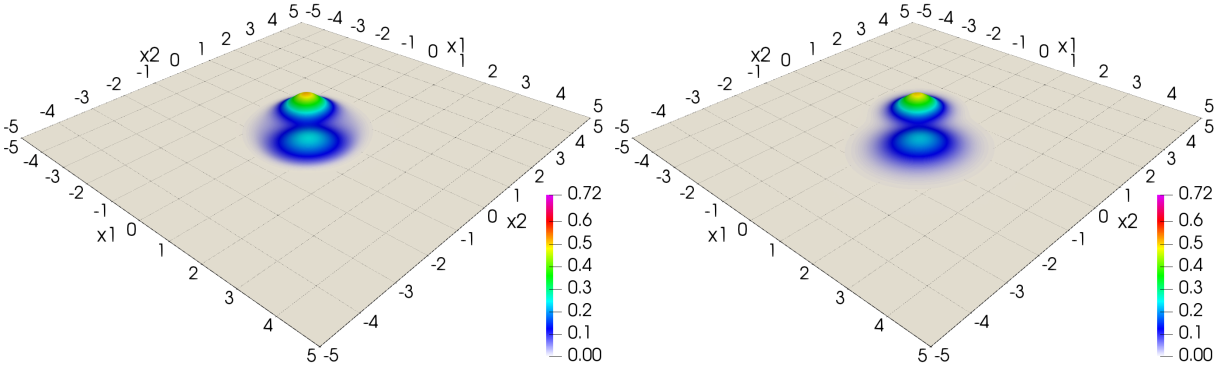
(e) Control $u_1(x, 2.5)$.

(f) Control $u_2(x, 2.5)$.

(g) Control $u_1(x, 4.5)$.

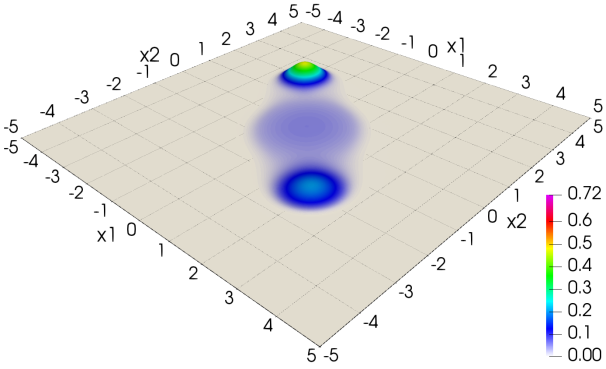
(h) Control $u_2(x, 4.5)$.

Figure 7.5: Controls $u_1(x, t)$ and $u_2(x, t)$ for Example 7.2 (Bimodal Target). Note the different scales at $t = 0$.

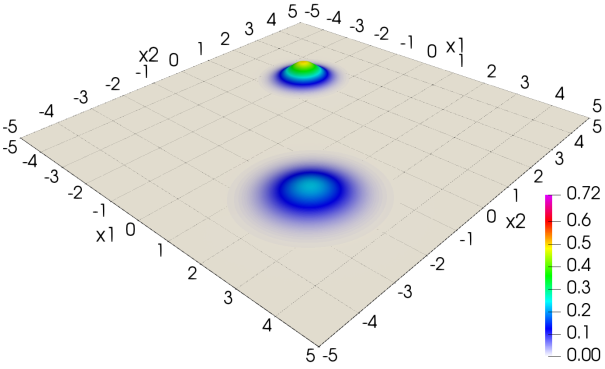


(a) Controlled PDF at $t = 0.5$.

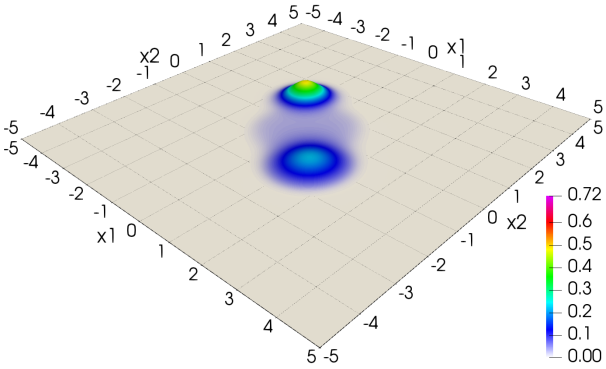
(b) Desired PDF at $t = 0.5$.



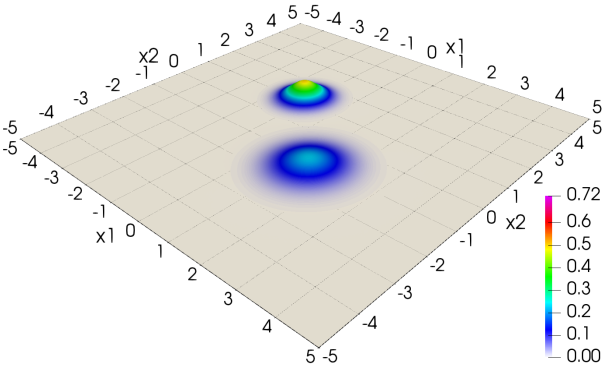
(c) Controlled PDF at $t = 2.5$.



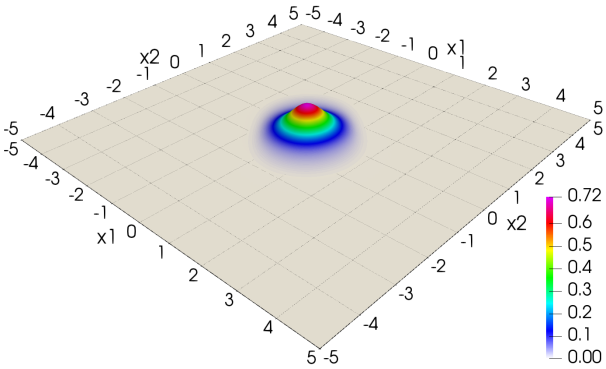
(d) Desired PDF at $t = 2.5$.



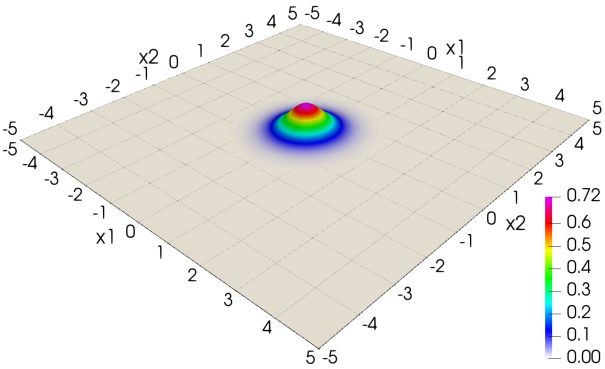
(e) Controlled PDF at $t = 4$.



(f) Desired PDF at $t = 4$.

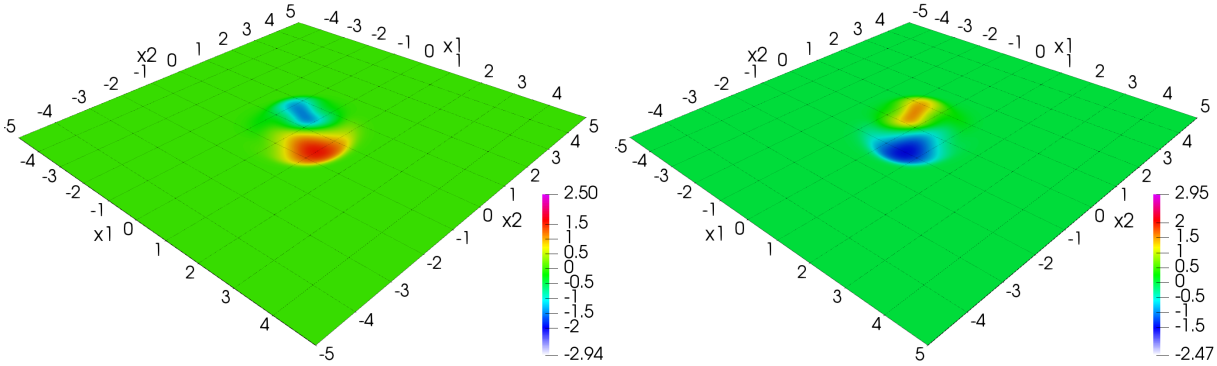


(g) Controlled PDF at $t = 5$.



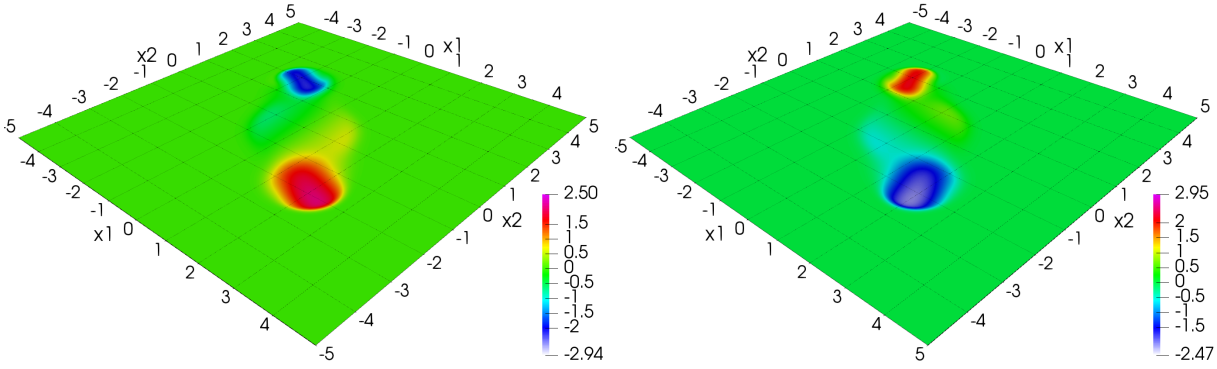
(h) Desired PDF at $t = 5$.

Figure 7.6: Desired and controlled PDF for Example 7.3 (Moving Bimodal Target).



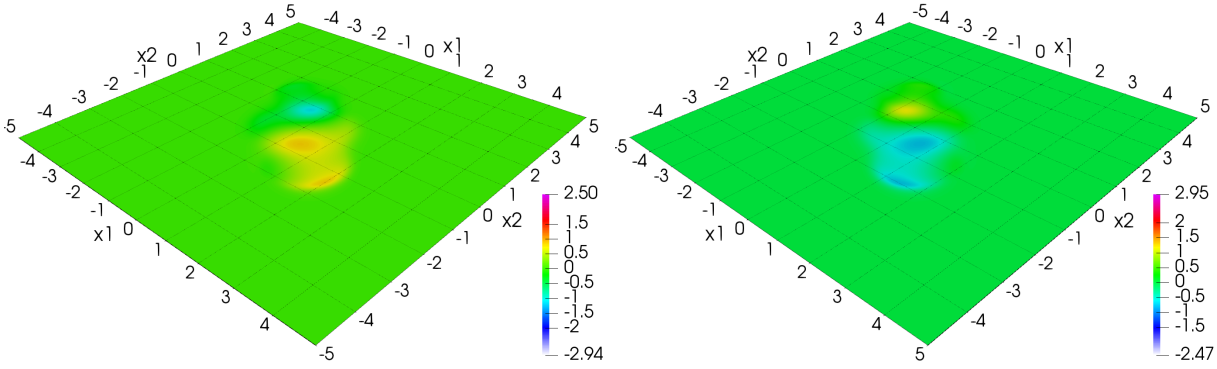
(a) Control $u_1(x, 0)$.

(b) Control $u_2(x, 0)$.



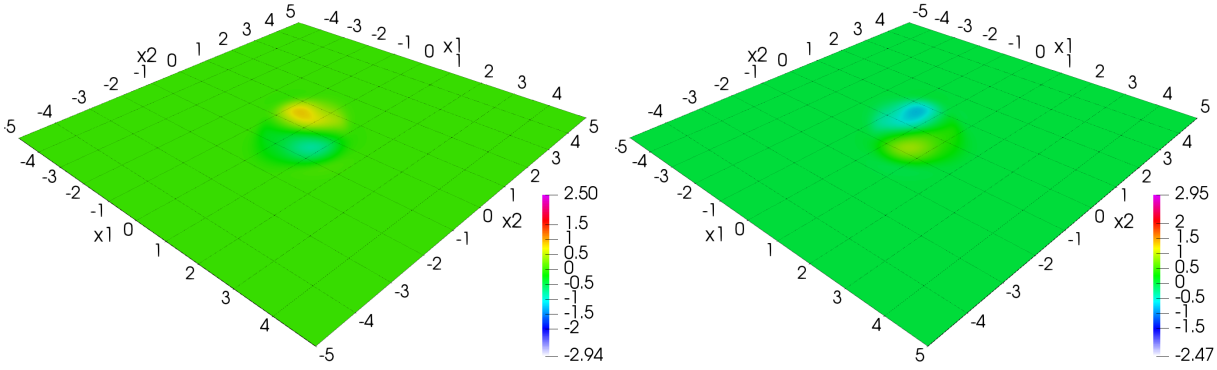
(c) Control $u_1(x, 2)$.

(d) Control $u_2(x, 2)$.



(e) Control $u_1(x, 3.5)$.

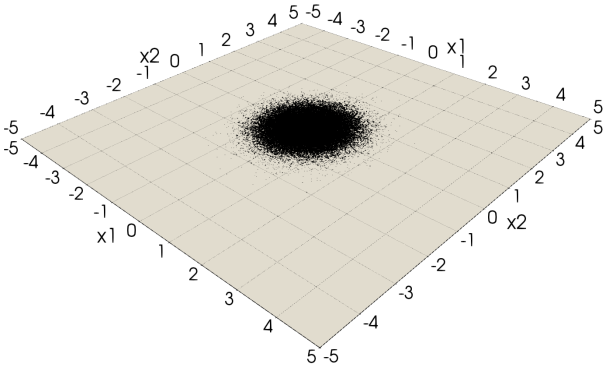
(f) Control $u_2(x, 3.5)$.



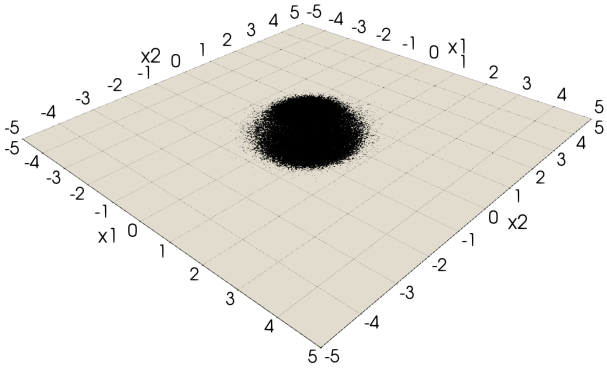
(g) Control $u_1(x, 4.5)$.

(h) Control $u_2(x, 4.5)$.

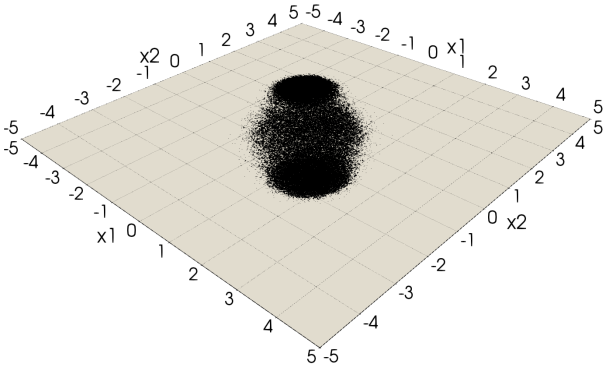
Figure 7.7: Controls $u_1(x, t)$ and $u_2(x, t)$ for Example 7.3 (Moving Bimodal Target).



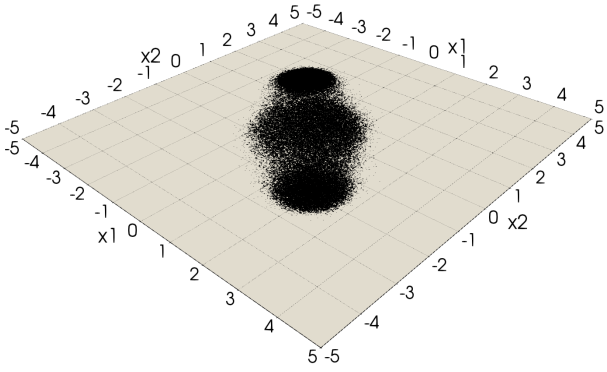
(a) Controlled SDE at $t = 0$.



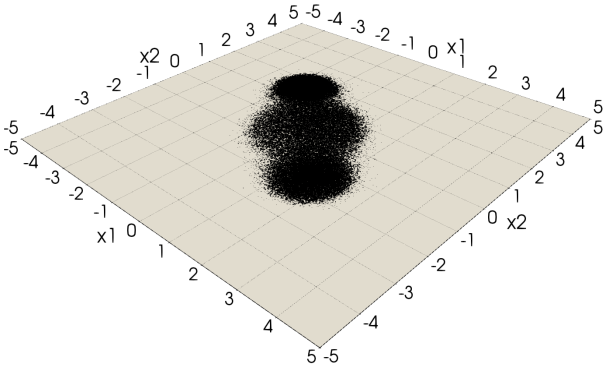
(b) Controlled SDE at $t = 0.5$.



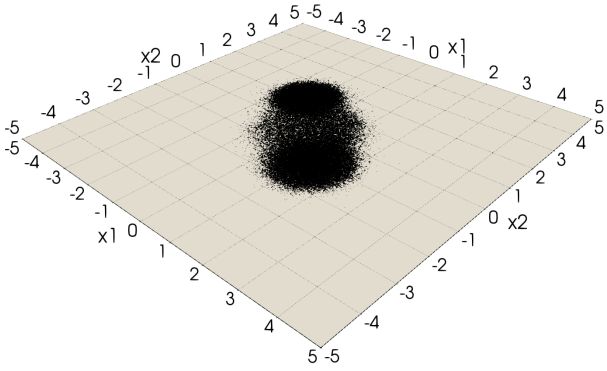
(c) Controlled SDE at $t = 1.5$.



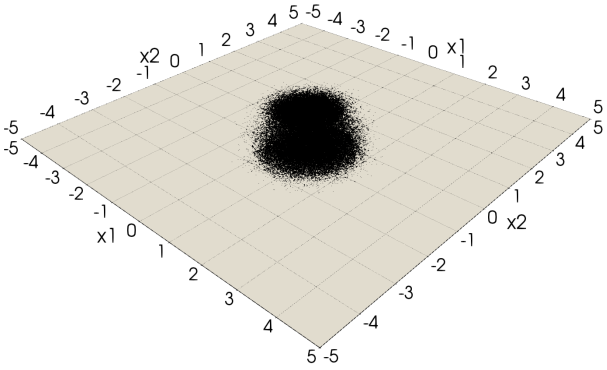
(d) Controlled SDE at $t = 2.5$.



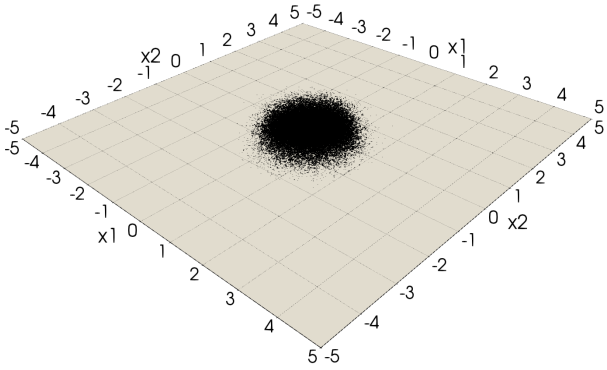
(e) Controlled SDE at $t = 3.5$.



(f) Controlled SDE at $t = 4$.

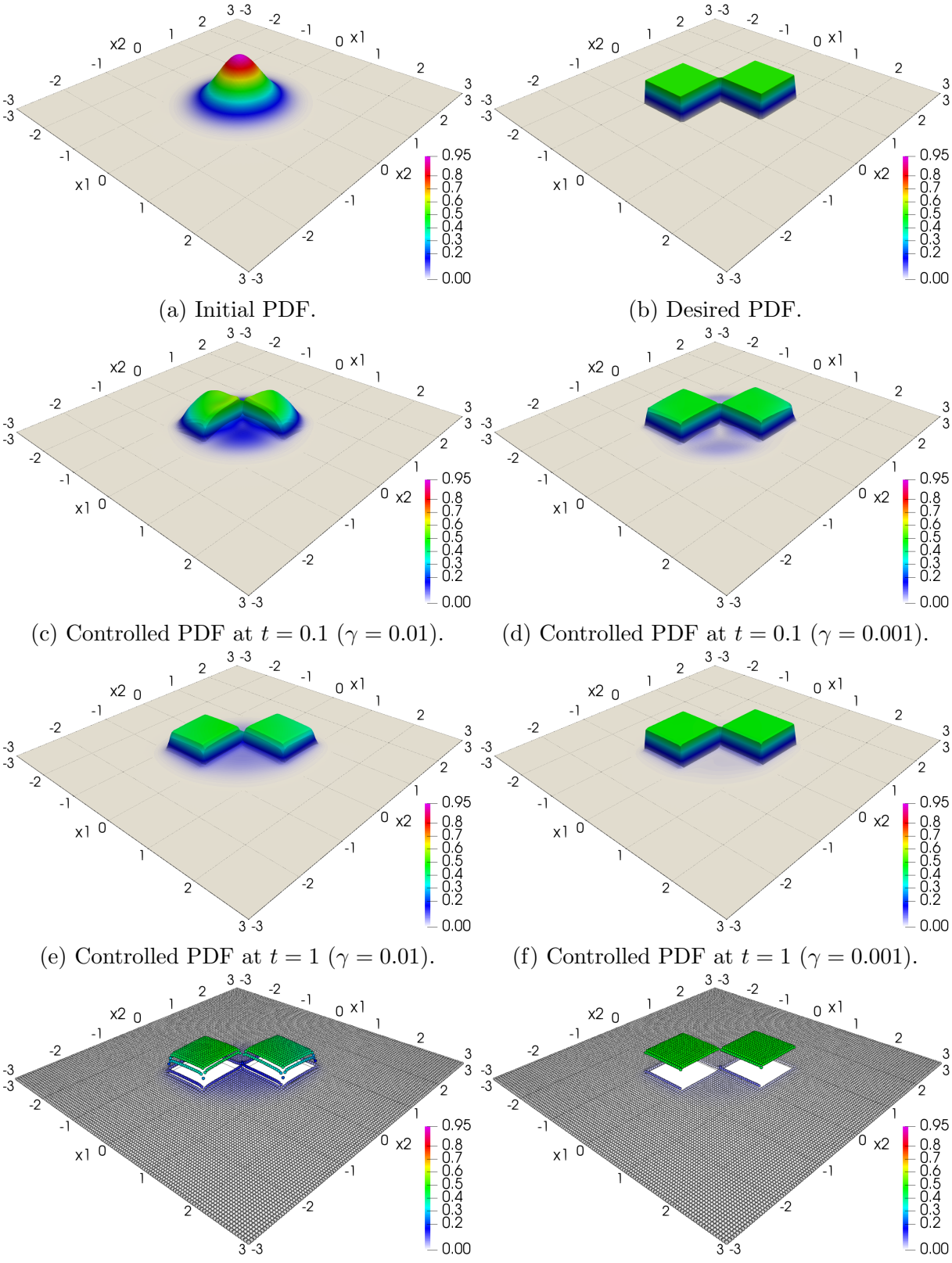


(g) Controlled SDE at $t = 4.5$.



(h) Controlled SDE at $t = 5$.

Figure 7.8: Controlled SDE for Example 7.3 (Moving Bimodal Target); 100000 paths.



(a) Initial PDF.

(b) Desired PDF.

(c) Controlled PDF at $t = 0.1$ ($\gamma = 0.01$).

(d) Controlled PDF at $t = 0.1$ ($\gamma = 0.001$).

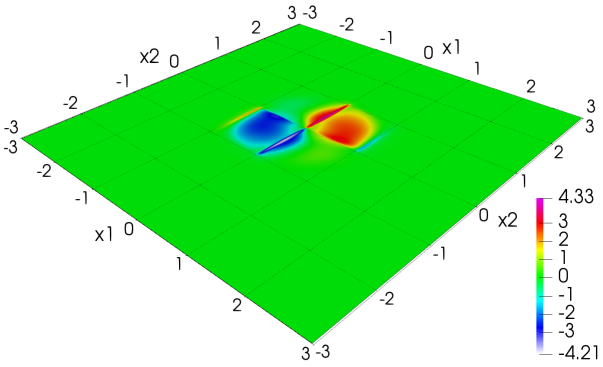
(e) Controlled PDF at $t = 1$ ($\gamma = 0.01$).

(f) Controlled PDF at $t = 1$ ($\gamma = 0.001$).

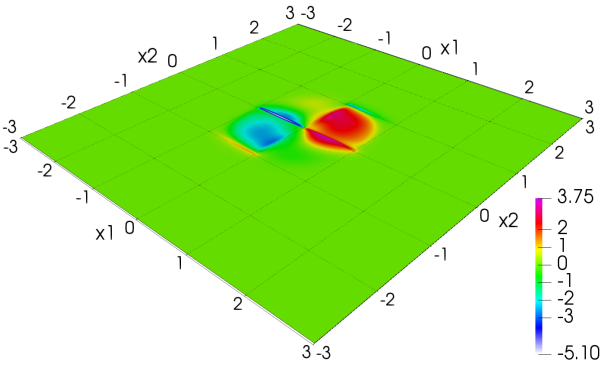
(g) Controlled PDF at $t = 1$ ($\gamma = 0.01$; grid points).

(h) Controlled PDF at $t = 1$ ($\gamma = 0.001$; grid points).

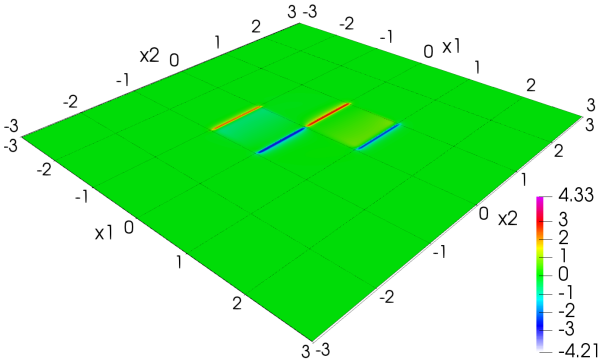
Figure 7.9: Desired and controlled PDF for Example 7.4 (Bimodal Uniform Target) and various regularization parameters γ.



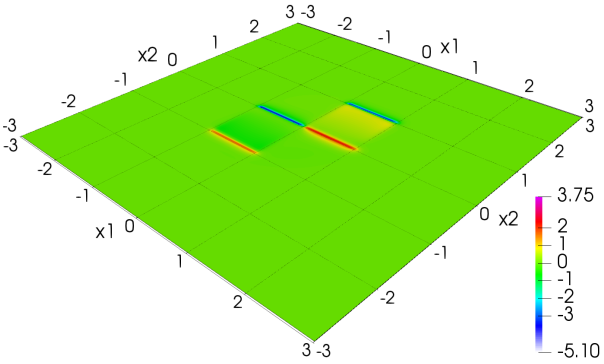
(a) Control $u_1(x, 0)$ for $\gamma = 0.01$.



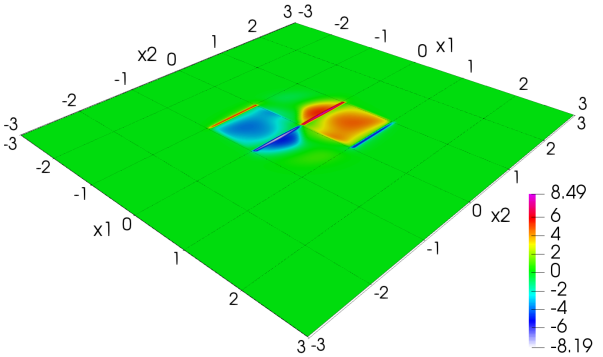
(b) Control $u_2(x, 0)$ for $\gamma = 0.01$.



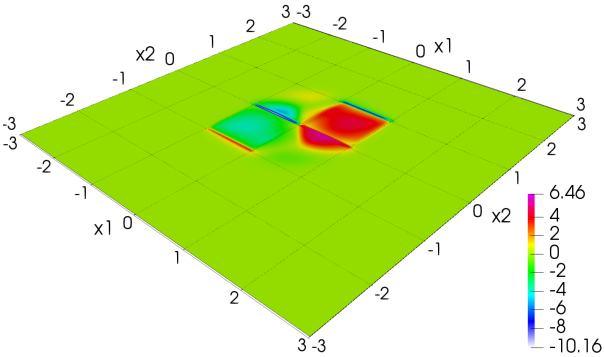
(c) Control $u_1(x, 4.5)$ for $\gamma = 0.01$.



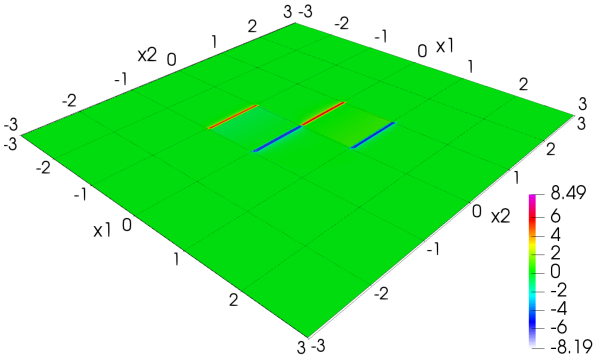
(d) Control $u_2(x, 4.5)$ for $\gamma = 0.01$.



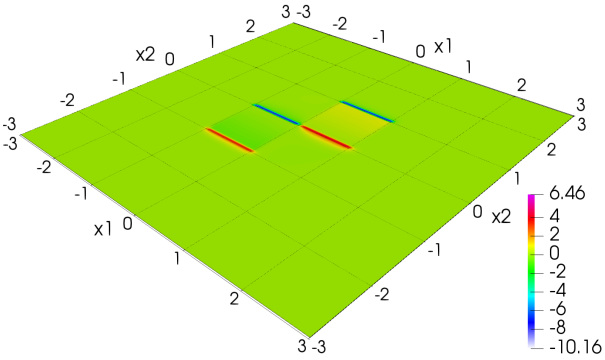
(e) Control $u_1(x, 0)$ for $\gamma = 0.001$.



(f) Control $u_2(x, 0)$ for $\gamma = 0.001$.

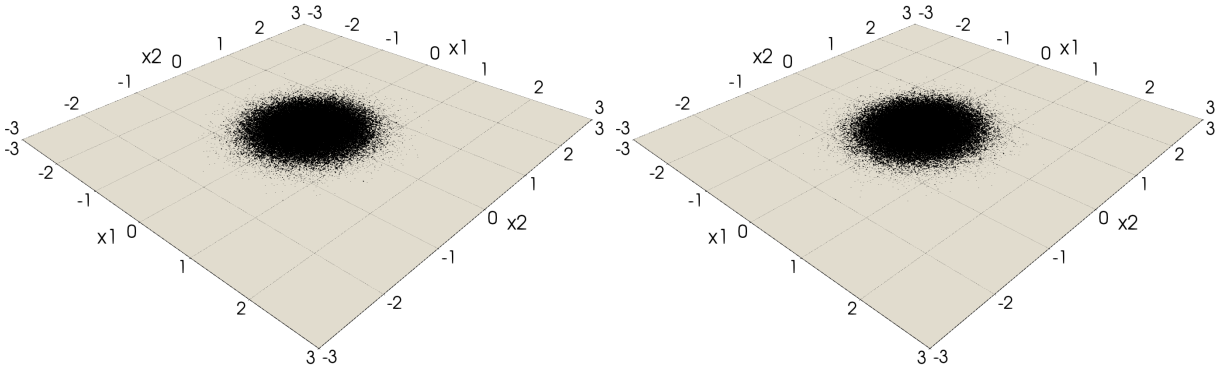


(g) Control $u_1(x, 4.5)$ for $\gamma = 0.001$.



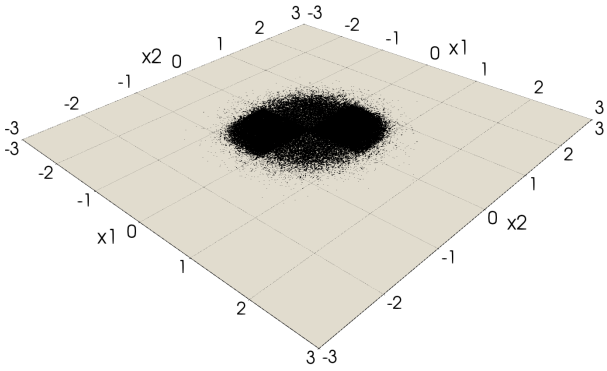
(h) Control $u_2(x, 4.5)$ for $\gamma = 0.001$.

Figure 7.10: Controls $u_1(x, t)$ and $u_2(x, t)$ for Example 7.4 (Bimodal Uniform Target) and various regularization parameters γ. Note the different scales for the different values of γ.

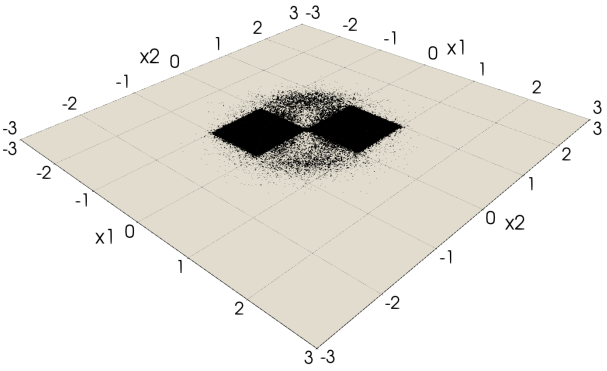


(a) Controlled SDE at $t = 0$ ($\gamma = 0.01$).

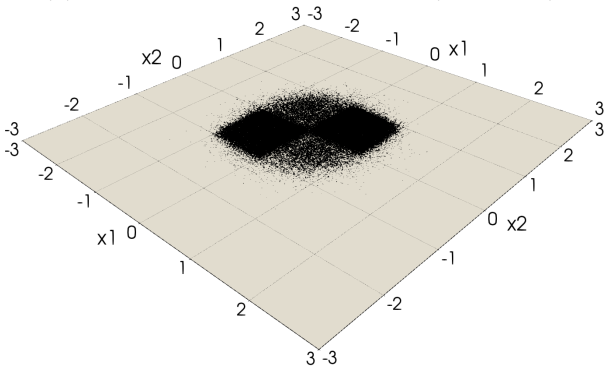
(b) Controlled SDE at $t = 0$ ($\gamma = 0.001$).



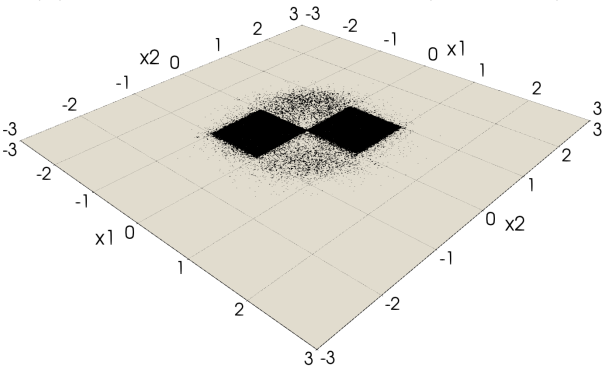
(c) Controlled SDE at $t = 0.1$ ($\gamma = 0.01$).



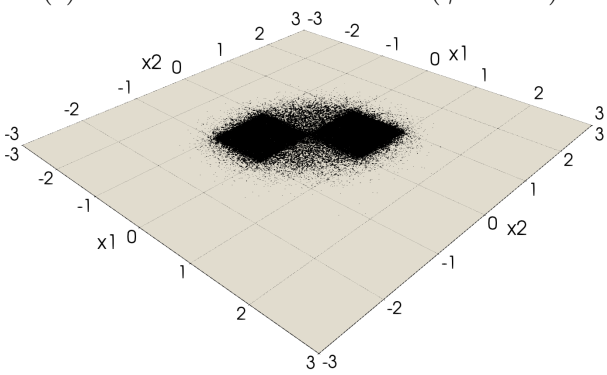
(d) Controlled SDE at $t = 0.1$ ($\gamma = 0.001$).



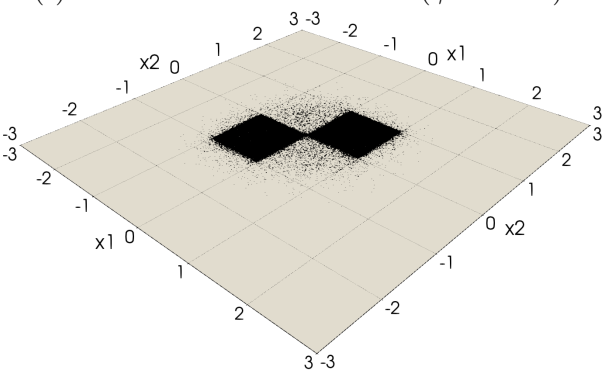
(e) Controlled SDE at $t = 0.2$ ($\gamma = 0.01$).



(f) Controlled SDE at $t = 0.2$ ($\gamma = 0.001$).



(g) Controlled SDE at $t = 1$ ($\gamma = 0.01$).



(h) Controlled SDE at $t = 1$ ($\gamma = 0.001$).

Figure 7.11: SDE simulation for Example 7.4 (Moving Uniform Target) and various regularization parameters γ; 100000 paths.

Future Research

8

In this concluding chapter we outline various extensions and possibilities for future research.

8.1 Generalization of existing results

One way to deepen the understanding to what extent Model Predictive Control works well in the Fokker–Planck optimal control framework (cf. Section 1.1) is to generalize the existing results in both the stabilizing MPC case (cf. Section 3.2) and the economic MPC case (cf. Section 3.3).

8.1.1 Minimal Stabilizing Horizon

In the case of stabilizing MPC we have shown that the MPC approach we employ "works" for linear stochastic processes and Gaussian PDFs, cf. Theorem 5.11, provided the horizon is sufficiently large. It would be interesting to generalize this result on the Fokker–Planck PDE level and to incorporate nonlinear stochastic processes or linear processes with a nonlinear control, cf. Section 7.4.

Another challenge is to determine or at least approximate the minimal stabilizing horizon. In this context, we have studied the Ornstein–Uhlenbeck process in more detail, cf. Section 4.2, where the control is space-independent and Section 5.3.2, where the space-dependent control is linear in space. A common trait we encountered is that the minimal stabilizing horizon is $N = 2$, i.e., the shortest possible horizon, although the optimal value function grows, cf. Example 5.14 or Section 4.3 and Figure 4.5. For nonlinear as well as for other linear processes, strategies have to be developed to cope with this. One such strategy—which was persued in Section 4.2—is to suitably modify the stage cost without affecting the resulting optimal control sequence, such that Theorem 3.4 can still be employed.

8.1.2 Strict Dissipativity

In the case of economic MPC we have investigated the strict dissipativity of optimal control problems subject to a bilinear discrete-time dynamics that approximates the Ornstein–Uhlenbeck process. While various stage costs were considered, it would be very desirable to extend this analysis to more complicated dynamics, ideally to a class of stochastic processes through a combination of a stage cost ℓ and a suitable storage function λ such that the modified cost $\tilde{\ell}$ is (strongly) convex.

Another open question, although minor compared to the previous one, is whether the conditions in Assumption 3.10 hold, which is necessary in order to apply the stability result from Theorem 3.13. While the continuity of the storage function is usually directly provable, the remaining properties likely need to be shown indirectly, via local controllability, cf. Definition 3.11.

8.2 New Fields of Application

A different research direction, which is possibly more promising in terms of more readily available results, is to study Model Predictive Control in connection with the Fokker–Planck equation in the context of new fields of application, e.g., mean-field games and mean-field type control problems [12].

8.2.1 Mean-Field Games

One related problem to steering the PDF via control of the Fokker–Planck equation is the planning problem in the context of mean-field games, considered in [76]. In mean-field games one considers strategic decision making of an agent in the case of a (large) population of (indistinguishable) players. In the case of an infinite number of players, the Fokker–Planck equation is coupled with a Hamilton–Jacobi–Bellman equation and models the evolution of the density of players, who are characterized by their strategy [76, 77]. The strategy of an individual agent depends on that density, i.e., on the behavior of other players, but it is treated as an external parameter that cannot be influenced by the agent [12]. The goal of the planning problem is to steer an initial density (of players) to a desired one "through the optimal decisions of the agents" [76]. It would be interesting to see to what extent Model Predictive Control can be applied in this context.

8.2.2 Mean-Field Type Control Problems

Similar to mean-field games, mean-field type control problems can be formulated in the Fokker–Planck optimal control framework, where the Fokker–Planck equation is coupled with a Hamilton–Jacobi–Bellman equation. For more details, see, e.g., the monograph [12] or the survey [6]. The difference to mean-field games is that the mean-field term, i.e., the density, can be influenced by the agent, resulting in a coupled system of McKean-Vlasov type [12].

Although nowadays powerful numerical solution techniques exist, solving this control problem on long or variable horizons is computationally hard. In this context, Model Predictive Control seems to be a viable approach, as first results in [15] indicate.

List of Figures

List of Tables

Bibliography

[1] A. Addou and A. Benbrik. Existence and uniqueness of optimal control for a distributed-parameter bilinear system. *J. Dynam. Control Systems*, 8(2):141–152, 2002.

[2] N. Altmüller and L. Grüne. Distributed and boundary Model Predictive Control for the heat equation. *GAMM-Mitt.*, 35(2):131–145, 2012.

[3] D. Angeli, R. Amrit, and J. B. Rawlings. On average performance and stability of economic model predictive control. *IEEE Trans. Autom. Control*, 57(7):1615–1626, 2012.

[4] M. Annunziato and A. Borzì. Optimal control of probability density functions of stochastic processes. *Math. Model. Anal.*, 15(4):393–407, 2010.

[5] M. Annunziato and A. Borzì. A Fokker-Planck control framework for multidimensional stochastic processes. *J. Comput. Appl. Math.*, 237(1):487–507, 2013.

[6] M. Annunziato and A. Borzì. A fokker–planck control framework for stochastic systems. *EMS Surveys in Mathematical Sciences*, 5(1):65–98, 2018.

[7] M. Annunziato, A. Borzì, F. Nobile, and R. Tempone. On the connection between the Hamilton-Jacobi-Bellman and the Fokker-Planck control frameworks. *Applied Mathematics*, 5(16):2476–2484, 2014.

[8] D. G. Aronson. Non-negative solutions of linear parabolic equations. *Ann. Scuola Norm. Sup. Pisa (3)*, 22:607–694, 1968.

[9] D. G. Aronson and J. Serrin. Local behavior of solutions of quasilinear parabolic equations. *Arch. Rational Mech. Anal.*, 25:81–122, 1967.

[10] J.-P. Aubin. Un théorème de compacité. *C. R. Acad. Sci. Paris*, 256:5042–5044, 1963.

[11] U. Ayachit. *The ParaView Guide: A Parallel Visualization Application*. Kitware, Inc., Clifton Park, NY, USA, 2015.

[12] A. Bensoussan, J. Frehse, and P. Yam. *Mean Field Games and Mean Field Type Control Theory*, volume 101. Springer New York, 2013.

[13] A. Blaquière. Controllability of a Fokker-Planck equation, the Schrödinger system, and a related stochastic optimal control (revised version). *Dynam. Control*, 2(3):235–253, 1992.

[14] A. Boccia, L. Grüne, and K. Worthmann. Stability and feasibility of state constrained MPC without stabilizing terminal constraints. *Systems & Control Letters*, 72:14–21, 2014.

[15] A. Borzì and L. Grüne. Towards a solution of mean-field control problems using model predictive control. *IFAC-PapersOnLine*, 53(2):4973–4978, 2020. 21st IFAC World Congress.

[16] T. Breiten, K. Kunisch, and L. Pfeiffer. Control strategies for the Fokker-Planck equation. *ESAIM: COCV*, 24(2):741–763, 2018.

[17] R. Brockett. New issues in the mathematics of control. In *Mathematics unlimited—2001 and beyond*, pages 189–219. Springer, Berlin, 2001.

[18] R. Brockett. Notes on the control of the Liouville equation. In *Control of Partial Differential Equations*, pages 101–129. Springer, 2012.

[19] E. A. Buehler, J. A. Paulson, and A. Mesbah. Lyapunov-based stochastic nonlinear model predictive control: Shaping the state probability distribution functions. In *2016 American Control Conference (ACC)*, pages 5389–5394, 2016.

[20] E. Casas. Pontryagin's principle for state-constrained boundary control problems of semilinear parabolic equations. *SIAM J. Control Optim.*, 35(4):1297–1327, 1997.

[21] J. Chang and G. Cooper. A practical difference scheme for fokker-planck equations. *Journal of Computational Physics*, 6(1):1–16, 1970.

[22] Y. Chen, T. T. Georgiou, and M. Pavon. Optimal steering of a linear stochastic system to a final probability distribution, part II. *IEEE Transactions on Automatic Control*, 61(5):1170–1180, 2016.

[23] J. Conway. *A Course in Functional Analysis*, volume 96 of *Graduate Texts in Mathematics*. Springer New York, 2nd edition, 1990.

[24] T. Damm, L. Grüne, M. Stieler, and K. Worthmann. An exponential turnpike theorem for dissipative discrete time optimal control problems. *SIAM J. Control Optim.*, 52(3):1935–1957, 2014.

[25] M. Diehl, R. Amrit, and J. B. Rawlings. A Lyapunov function for economic optimizing model predictive control. *IEEE Trans. Autom. Control*, 56:703–707, 2011.

[26] R. Dorfman, P. A. Samuelson, and R. M. Solow. *Linear Programming and Economic Analysis*. Dover Publications, New York, 1987. Reprint of the 1958 original.

[27] S. Dubljevic and P. Christofides. Boundary predictive control of parabolic PDEs. In *American Control Conference, 2006*, pages 49–56, 2006.

[28] S. Dubljevic, N. H. El-Farra, P. Mhaskar, and P. D. Christofides. Predictive control of parabolic PDEs with state and control constraints. *International Journal of Robust and Nonlinear Control*, 16:749–772, 2006.

[29] T. Faulwasser, L. Grüne, and M. A. Müller. Economic nonlinear model predictive control. *Foundations and Trends® in Systems and Control*, 5(1):1–98, 2018.

[30] W. Feller. Diffusion processes in one dimension. *Trans. Amer. Math. Soc.*, 77:1–31, 1954.

[31] A. Figalli. Existence and uniqueness of martingale solutions for SDEs with rough or degenerate coefficients. *J. Funct. Anal.*, 254(1):109–153, 2008.

[32] A. Fleig and L. Grüne. On dissipativity of the Fokker–Planck equation for the Ornstein–Uhlenbeck process. *IFAC-PapersOnLine*, 52(2):13–18, 2019. 3rd IFAC Workshop on Control of Systems Governed by Partial Differential Equations CPDE 2019.

[33] A. Fleig and L. Grüne. Estimates on the minimal stabilizing horizon length in Model Predictive Control for the Fokker-Planck equation. *IFAC-PapersOnLine*, 49(8):260–265, 2016. 2nd IFAC Workshop on Control of Systems Governed by Partial Differential Equations CPDE 2016.

[34] A. Fleig and L. Grüne. L^2-tracking of Gaussian distributions via Model Predictive Control for the Fokker-Planck equation. *Vietnam J. Math.*, 46(4):915–948, Dec 2018.

[35] A. Fleig and L. Grüne. Strict dissipativity analysis for classes of optimal control problems involving probability density functions. *Math. Control Relat. F.*, 2020. doi: 10.3934/mcrf.2020053.

[36] A. Fleig, L. Grüne, and R. Guglielmi. Some results on Model Predictive Control for the Fokker-Planck equation. In *MTNS 2014: 21st International Symposium on Mathematical Theory of Networks and Systems, July 7-11, 2014*, University of Groningen, The Netherlands, pages 1203–1206, 2014.

[37] A. Fleig and R. Guglielmi. Bilinear optimal control of the Fokker-Planck equation. *IFAC-PapersOnLine*, 49(8):254–259, 2016. 2nd IFAC Workshop on Control of Systems Governed by Partial Differential Equations CPDE 2016.

[38] A. Fleig and R. Guglielmi. Optimal control of the Fokker-Planck equation with space-dependent controls. *J. Optim. Theory Appl.*, 174(2):408–427, 2017.

[39] W. H. Fleming and R. W. Rishel. *Deterministic and stochastic optimal control*. Springer-Verlag, Berlin-New York, 1975. Applications of Mathematics, No. 1.

[40] M. G. Forbes, J. F. Forbes, and M. Guay. Regulating discrete-time stochastic systems: focusing on the probability density function. *Dyn. Contin. Discrete Impuls. Syst. Ser. B Appl. Algorithms*, 11(1-2):81–100, 2004.

[41] C. Gardiner. *Stochastic methods*. Springer Series in Synergetics. Springer-Verlag, Berlin, fourth edition, 2009. A handbook for the natural and social sciences.

[42] T. T. Georgiou. The structure of state covariances and its relation to the power spectrum of the input. *IEEE Transactions on Automatic Control*, 47(7):1056–1066, 2002.

[43] Ĭ. Ī. Gīhman and A. V. Skorohod. *Stochastic differential equations*. Springer-Verlag, New York-Heidelberg, 1972. Translated from the Russian by Kenneth Wickwire, Ergebnisse der Mathematik und ihrer Grenzgebiete, Band 72.

[44] C. R. Givens and R. M. Shortt. A class of wasserstein metrics for probability distributions. *Michigan Math. J.*, 31(2):231–240, 1984.

[45] L. Grüne. Analysis and design of unconstrained nonlinear MPC schemes for finite and infinite dimensional systems. *SIAM J. Control Optim.*, 48:1206–1228, 2009.

[46] L. Grüne. Economic receding horizon control without terminal constraints. *Automatica*, 49:725–734, 2013.

[47] L. Grüne. Approximation properties of receding horizon optimal control. *Jahresbericht der Deutschen Mathematiker-Vereinigung*, 118(1):3–37, 2016.

[48] L. Grüne and M. A. Müller. On the relation between strict dissipativity and the turnpike property. *Syst. Contr. Lett.*, 90:45–53, 2016.

[49] L. Grüne and J. Pannek. *Nonlinear Model Predictive Control. Theory and Algorithms*. Springer, London, 2nd edition, 2017.

[50] L. Grüne, J. Pannek, M. Seehafer, and K. Worthmann. Analysis of unconstrained nonlinear MPC schemes with time varying control horizon. *SIAM J. Control Optim.*, 48:4938–4962, 2010.

[51] L. Grüne, M. Schaller, and A. Schiela. Sensitivity analysis of optimal control for a class of parabolic PDEs motivated by model predictive control. *SIAM J. Control Optim.*, 57(4):2753–2774, 2019.

[52] L. Grüne, M. Schaller, and A. Schiela. Exponential sensitivity and turnpike analysis for linear quadratic optimal control of general evolution equations. *J. Differ. Equ.*, 268(12):7311–7341, 2020.

[53] L. Grüne and M. Stieler. Asymptotic stability and transient optimality of economic MPC without terminal conditions. *J. Proc. Control*, 24(8):1187–1196, 2014.

[54] G. Guennebaud, B. Jacob, et al. Eigen v3. http://eigen.tuxfamily.org, 2010.

[55] W. Hahn. *Stability of motion*, volume 138. Springer, 1967.

[56] A. W. Heemink. Stochastic modelling of dispersion in shallow water. *Stochastic Hydrology and Hydraulics*, 4(2):161–174, Jun 1990.

[57] 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.

[58] W. Horsthemke and R. Lefever. *Noise-Induced Transitions: Theory and Applications in Physics, Chemistry, and Biology*, volume 15 of *Springer Series in Synergetics*. Springer-Verlag, Berlin, 1984.

[59] K. Ito and K. Kunisch. Receding horizon optimal control for infinite dimensional systems. *ESAIM: COCV*, 8:741–760, 2002.

[60] G. Jumarie. Tracking control of nonlinear stochastic systems by using path cross-entropy and Fokker-Planck equation. *Internat. J. Systems Sci.*, 23(7):1101–1114, 1992.

[61] M. Kárný. Towards fully probabilistic control design. *Automatica J. IFAC*, 32(12):1719–1722, 1996.

[62] P. E. Kloeden and E. Platen. *Numerical solution of stochastic differential equations*, volume 23. Springer Science & Business Media, 2013.

[63] A. Kolmogoroff. Über die analytischen Methoden in der Wahrscheinlichkeitsrechnung. *Math. Ann.*, 104(1):415–458, 1931.

[64] O. Ladyzhenskaya, V. Solonnikov, and N. Ural'tseva. *Linear and quasilinear equations of parabolic type.* Moskva: Izdat. 'Nauka'. 736 pp., 1967.

[65] C. Le Bris and P.-L. Lions. Existence and uniqueness of solutions to Fokker-Planck type equations with irregular coefficients. *Comm. Partial Differential Equations*, 33(7-9):1272–1317, 2008.

[66] B. Lincoln and A. Rantzer. Relaxing dynamic programming. *IEEE Transactions on Automatic Control*, 51(8):1249–1260, 08 2006.

[67] J.-L. Lions. *Quelques méthodes de résolution des problèmes aux limites non linéaires.* Dunod; Gauthier-Villars, Paris, 1969.

[68] J.-L. Lions. *Optimal control of systems governed by partial differential equations.* Translated from the French by S. K. Mitter. Die Grundlehren der mathematischen Wissenschaften, Band 170. Springer-Verlag, New York-Berlin, 1971.

[69] M. Manuel. A two-mean reverting-factor model of the term structure of interest rates. *Journal of Futures Markets*, 23(11):1075–1105, 2003.

[70] Maplesoft, a division of Waterloo Maple Inc. Maple (version 2019), 2019. https://maplesoft.com.

[71] M. A. Müller, D. Angeli, and F. Allgöwer. On necessity and robustness of dissipativity in economic model predictive control. *IEEE Transactions on Automatic Control*, 60(6):1671–1676, June 2015.

[72] M. Mohammadi and A. Borzì. Analysis of the Chang-Cooper discretization scheme for a class of Fokker-Planck equations. *Journal of Numerical Mathematics*, 23(3):271–288, 2015.

[73] J. Nocedal and S. J. Wright. *Numerical Optimization.* Springer, New York, NY, USA, second edition, 2006.

[74] G. Peskir. On the fundamental solution of the kolmogorov–shiryaev equation. In *From Stochastic Calculus to Mathematical Finance: The Shiryaev Festschrift*, pages 535–546. Springer Berlin Heidelberg, Berlin, Heidelberg, 2006.

[75] K. B. Petersen and M. S. Pedersen. The matrix cookbook, nov 2012. Version 20121115.

[76] A. Porretta. On the planning problem for the mean field games system. *Dyn. Games Appl.*, 4(2):231–256, 2014.

[77] A. Porretta. Weak solutions to Fokker-Planck equations and mean field games. *Arch. Ration. Mech. Anal.*, 216(1):1–62, 2015.

[78] A. Porretta and E. Zuazua. Long time versus steady state optimal control. *SIAM J. Control Optim.*, 51(6):4242–4273, 2013.

[79] S. Primak, V. Kontorovich, and V. Lyandres. *Stochastic methods and their applications to communications*. John Wiley & Sons, Inc., Hoboken, NJ, 2004.

[80] P. E. Protter. *Stochastic Integration and Differential Equations*, volume 21 of *Stochastic Modelling and Applied Probability*. Springer-Verlag, Berlin, 2005.

[81] J. Rawlings, D. Mayne, and M. Diehl. *Model Predictive Control: Theory and Design*. Nob Hill Publishing, 2nd edition, 2017.

[82] J. B. Rawlings, D. Bonné, J. B. Jorgensen, A. N. Venkat, and S. B. Jorgensen. Unreachable setpoints in model predictive control. *IEEE Transactions on Automatic Control*, 53(9):2209–2215, 2008.

[83] J. P. Raymond and H. Zidani. Hamiltonian Pontryagin's principles for control problems governed by semilinear parabolic equations. *Appl. Math. Optim.*, 39(2):143–177, 1999.

[84] H. Risken. *The Fokker-Planck Equation*, volume 18 of *Springer Series in Synergetics*. Springer-Verlag, Berlin, 2nd edition, 1989.

[85] S. Roy, M. Annunziato, and A. Borzì. A Fokker-Planck feedback control-constrained approach for modelling crowd motion. *J. Comput. Theor. Transp.*, 45(6):442–458, 2016.

[86] S. Roy, M. Annunziato, A. Borzì, and C. Klingenberg. A fokker–planck approach to control collective motion. *Computational Optimization and Applications*, 69(2):423–459, Mar 2018.

[87] W. E. Schiesser. *The numerical method of lines: integration of partial differential equations*. Elsevier, 2012.

[88] N. Sévérien and A. S. Ejaz. Shrinkage drift parameter estimation for multi-factor Ornstein-Uhlenbeck processes. *Applied Stochastic Models in Business and Industry*, 26(2):103–124, 2009.

[89] J. Simon. Compact sets in the space $L^p(0,T;B)$. *Ann. Mat. Pura Appl. (4)*, 146:65–96, 1987.

[90] E. D. Sontag. Smooth stabilization implies coprime factorization. *IEEE Transactions on Automatic Control*, 34(4):435–443, 1989.

[91] B. Stroustrup. *The C++ programming language*. Pearson Education India, 2000.

[92] H. Tanaka. Stochastic differential equations with reflecting boundary condition in convex regions. *Hiroshima Math. J.*, 9(1):163–177, 1979.

[93] E. Trélat, C. Zhang, and E. Zuazua. Steady-state and periodic exponential turnpike property for optimal control problems in Hilbert spaces. *SIAM J. Control Optim.*, 56(2):1222–1252, 2018.

[94] E. Trélat and E. Zuazua. The turnpike property in finite-dimensional nonlinear optimal control. *Journal of Differential Equations*, 258(1):81–114, 2015.

[95] F. Tröltzsch. *Optimal Control of Partial Differential Equations*, volume 112 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2010.

[96] G. E. Uhlenbeck and L. S. Ornstein. On the theory of the brownian motion. *Phys. Rev.*, 36:823–841, Sep 1930.

[97] G. Van Rossum and F. L. Drake. *Python 3 Reference Manual*. CreateSpace, Scotts Valley, CA, 2009.

[98] O. Vasicek. An equilibrium characterization of the term structure. *Journal of Financial Economics*, 5(2):177–188, 1977.

[99] H. Wang. Robust control of the output probability density functions for multivariable stochastic systems with guaranteed stability. *IEEE Trans. Automat. Control*, 44(11):2103–2107, 1999.

[100] P. Wieschollek. Cppoptimizationlibrary. `https://github.com/PatWie/CppNumericalSolvers`, 2016.

[101] J. C. Willems. Dissipative dynamical systems. I. General theory. *Arch. Rational Mech. Anal.*, 45:321–351, 1972.

[102] J. Wloka. *Partial differential equations*. Cambridge University Press, Cambridge, 1987.

[103] D. W. Yeung and S. E. Stewart. Stationary probability distributions of some lotka-volterra types of stochastic predation systems. *Stochastic Analysis and Applications*, 13(4):503–516, 1995.

Publications

[1] A. Fleig and L. Grüne. Strict dissipativity analysis for classes of optimal control problems involving probability density functions. *Math. Control Relat. F.*, 2020. doi: 10.3934/mcrf.2020053.

[2] A. Fleig and L. Grüne. On dissipativity of the Fokker–Planck equation for the Ornstein–Uhlenbeck process. *IFAC-PapersOnLine*, 52(2):13–18, 2019. 3rd IFAC Workshop on Control of Systems Governed by Partial Differential Equations CPDE 2019.

[3] A. Fleig and L. Grüne. L^2-tracking of Gaussian distributions via Model Predictive Control for the Fokker-Planck equation. *Vietnam J. Math.*, 46(4):915–948, Dec 2018.

[4] A. Fleig and R. Guglielmi. Optimal control of the Fokker-Planck equation with space-dependent controls. *J. Optim. Theory Appl.*, 174(2):408–427, 2017.

[5] A. Fleig and L. Grüne. Model Predictive Control for the Fokker-Planck equation: analysis and structural insight. In *Proceedings of the 22nd International Symposium on Mathematical Theory of Networks and Systems*, University of Minnesota, Minneapolis, pages 689–690, 2016.

[6] A. Fleig and L. Grüne. Estimates on the minimal stabilizing horizon length in Model Predictive Control for the Fokker-Planck equation. *IFAC-PapersOnLine*, 49(8):260–265, 2016. 2nd IFAC Workshop on Control of Systems Governed by Partial Differential Equations CPDE 2016.

[7] A. Fleig and R. Guglielmi. Bilinear optimal control of the Fokker-Planck equation. *IFAC-PapersOnLine*, 49(8):254–259, 2016. 2nd IFAC Workshop on Control of Systems Governed by Partial Differential Equations CPDE 2016.

[8] A. Fleig, L. Grüne, and R. Guglielmi. Some results on Model Predictive Control for the Fokker-Planck equation. In *MTNS 2014: 21st International Symposium on Mathematical Theory of Networks and Systems, July 7-11, 2014*, University of Groningen, The Netherlands, pages 1203–1206, 2014.

Eidesstattliche Versicherung

Hiermit versichere ich an Eides statt, dass ich die vorliegende Arbeit selbstständig verfasst und keine anderen als die von mir angegebenen Quellen und Hilfsmittel verwendet habe.

Weiterhin erkläre ich, dass ich die Hilfe von gewerblichen Promotionsberatern bzw. Promotionsvermittlern oder ähnlichen Dienstleistern weder bisher in Anspruch genommen habe, noch künftig in Anspruch nehmen werde.

Zusätzlich erkläre ich hiermit, dass ich keinerlei frühere Promotionsversuche unternommen habe.

Bayreuth, den

_____

(Arthur Fleig)