Sensitivity Analysis and Goal Oriented Error Estimation for Model Predictive Control

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

von
Manuel Schaller
geboren in Ansbach

1. Gutachter:	Prof. Dr. Lars Grüne
2. Gutachterin:	Prof. Dr. Birgit Jacob
3. Gutachter:	Prof. Dr. Stefan Volkwein

Tag der Einreichung: 10.11.2020
Tag des Kolloquiums: 30.04.2021

Mai, 2021

Zusammenfassung

Gegenstand dieser Arbeit ist die Sensitivitätsanalyse und die spezialisierte adaptive Diskretisierung für die modellprädiktive Regelung von Optimalsteuerungsproblemen mit partiellen Differentialgleichungen. In jedem Schritt eines modellprädiktiven Reglers wird ein Optimalsteuerungsproblem auf einem möglicherweise langen Zeithorizont gelöst. Nur ein Anfangsteil der optimalen Lösung wird als Regelung für das zu steuernde System verwendet. Dies motiviert die Verwendung von effizienten Diskretisierungsschemata, die genau auf dieses Vorgehen zugeschnitten sind, die also Orts- und Zeitgitter verwenden, welche am Anfang des Zeithorizonts fein sind und gegen Ende immer gröber werden.

In dieser Arbeit wird eine umfangreiche Sensitivitätsanalyse durchgeführt, um den Einfluss von Störungen, die in ferner Zukunft auftreten, auf die Rückkopplung des modellprädiktiven Reglers, also die optimale Steuerung auf einem Anfangsteil des Lösungshorizonts, abzuschätzen. Es wird unter Stabilisierbarkeitsannahmen an die zugrundeliegenden Operatoren gezeigt, dass der Einfluss von Störungen lokaler Natur ist, d.h., dass Diskretisierungsfehler, die in ferner Zukunft auftreten, einen vernachlässigbaren Einfluss auf die Rückkopplung der modellprädiktiven Regelung haben. Diese Eigenschaft wird für eine Vielzahl von Problemklassen bewiesen, darunter Probleme, deren Dynamik durch eine stark stetige Halbgruppe, durch eine nichtautonome parabolische Gleichung oder durch eine semilineare parabolische Gleichung beschrieben wird. Weiterhin wird gezeigt, dass dieses Abklingen von Störungen im Falle eines autonomen Problems sehr nah mit der Turnpike Eigenschaft verwandt ist – einer strukturellen Eigenschaft von optimalen Lösungen, die sich dadurch auszeichnet, dass die Lösungen von Optimalsteuerungsproblemen auf langen Zeithorizonten die meiste Zeit nahe eines Gleichgewichts verweilen. In diesem Kontext werden neue Turnpike Resultate gezeigt.

Diese theoretische Analyse bietet die Grundlage für effiziente Diskretisierungsverfahren für die modellprädiktive Regelung. Wir schlagen dazu verschiedene Methoden zur a-priori-Diskretisierung in Ort und Zeit vor. Weiter analysieren wir die zielorientierte a-posteriori-Fehlerschätzung mit einer bestimmten Interessensfunktion, die nur einen Anfangsteil des Horizonts mit einbezieht, als wirksames Werkzeug für die adaptive modellprädiktive Regelung. Dazu werden wir unter Stabilisierbarkeitsannahmen beweisen, dass die Fehlerindikatoren außerhalb des Trägers dieser spezialisierten Interessensfunktion exponentiell abfallen. Wir werden das Verhalten und die Performanz dieser adaptiven Diskretisierungsmethoden im Kontext der modellprädiktiven Regelung an einer Vielzahl von numerischen Beispielen testen, darunter Probleme mit linearen, semilinearen und quasilinearen Dynamiken unter verteilter Steuerung sowie Randsteuerung.

Abstract

Subject of this thesis is the sensitivity analysis and the specialized adaptive discretization for the Model Predictive Control (MPC) of optimal control problems with partial differential equations. In every iteration of an MPC controller, an optimal control problem on a possibly long time horizon is solved. Only an initial part of the optimal solution is used as a feedback for the system to be controlled. This motivates the use of efficient discretization schemes tailored to this approach, i.e., space and time grids, which are fine at the beginning of the time interval and become coarser towards the end.

In this work, a comprehensive sensitivity analysis is performed to estimate the influence of perturbations that occur in the far future on the MPC feedback, i.e., the optimal control on an initial part. Under stabilizability conditions on the involved operators it will be shown that the influence of perturbations is of local nature, meaning that discretization errors that occur in the far future only have a negligible effect on the MPC feedback. This property will be proven for various problem classes, covering problems governed by strongly continuous semigroups, by non-autonomous parabolic equations or by semilinear parabolic equations. It is further shown that, in case of an autonomous problem, the exponential decay of perturbations is strongly connected to the turnpike property—a structural feature of optimal solutions stating that solutions of autonomous optimal control problems on a long time horizon reside close to a steady state for the majority of the time. In that context, novel turnpike results for optimal control problems are given.

The theoretical analysis serves as a foundation for efficient discretization methods for MPC. Thus, we propose several a priori space and time discretization schemes. Further, we analyze goal oriented a posteriori error estimation with a specialized objective for refinement, which only incorporates an initial part of the horizon, as a powerful tool for adaptive MPC. We will prove under stabilizability assumptions that the error indicators decay exponentially outside the support of this specialized quantity of interest. Finally, we illustrate the behavior and performance of these specialized discretization algorithms in an MPC context by various numerical examples, including problems governed by linear, semilinear, and quasilinear dynamics with distributed and with boundary control.

Acknowledgments

I am grateful to my supervisors Lars Grüne and Anton Schiela for suggesting this topic to me, their valuable input in countless discussions, for giving me the opportunity to attend several conferences and for providing me freedom and a fruitful research environment.

Moreover, I would like to thank Birgit Jacob and Stefan Volkwein for committing their time to referee this thesis.

Thanks also go to Felix Schwenninger for sparking my interest in semigroup theory and for many stimulating discussions. I would also like to thank Enrique Zuazua for valuable comments on my work concerning turnpike properties during several research stays. Further I am thankful to Georg Müller for carefully reading this thesis and giving feedback on various aspects.

I would like to thank my father for his support over all the years, my brother for sharing many enjoyable memories with me and my girlfriend Anja for her encouragement and patience. Finally, I would like to express my gratitude towards my mother for her endless care and example of perseverance.

Contents

Chapter 1

Introduction

Model Predictive Control (MPC) is a control technique which is widely used in many applications, such as chemical process engineering, electrical engineering, aerospace engineering or automotive engineering, cf. [26, 115]. It represents an optimization-based feedback controller, in which the solution of an optimal control problem (OCP) on an indefinite or infinite horizon is split into the successive solution of problems on a finite but possibly long horizon $T > 0$. Only an initial part up to a time $\tau > 0$, where often $\tau \ll T$, is implemented in the system under control. The resulting state is then measured or estimated and set as an initial condition, and the process is repeated. This procedure is depicted in Algorithm 1.

Algorithm 1 Standard MPC Algorithm

Given: Prediction horizon $T > 0$, implementation horizon $0 < \tau \leq T$, initial state x_0
$k = 0$
while controller active **do**
    Solve OCP on $[k\tau, T + k\tau]$ with initial state x_k, save optimal control in u
    Implement $u\big|_{[k\tau,(k+1)\tau]}$ as feedback, measure/estimate resulting state and save in x_{k+1}
    $k = k + 1$
end while

The resulting trajectories arising from an MPC algorithm can, in many applications, be proven to be quasi-optimal for the original problem on the infinite horizon. For this and many other aspects, we refer the interested reader to the paper [61] and the books [66, 118] which provide a mathematical foundation by covering topics including approximation properties, stability analysis, feasibility, robustness and efficient numerical implementation.

A rigorous stability analysis and performance estimates for MPC without terminal constraints or terminal cost can be concluded if a turnpike property is present, cf. [61]. The turnpike property is a feature of solutions to optimal control problems and, qualitatively speaking, states that the solution trajectory of an autonomous OCP on a long time horizon subject to an evolution equation resides close to an optimal steady state for the majority of the time. This behavior is depicted by the green trajectory in Figure 1.1. Loosely speaking, the turnpike property allows to replace the infinite horizon in the optimal control problem by a finite but large horizon without significantly changing the behavior of optimal solutions at small time instances. Also outside an MPC context, the turnpike property is a useful tool to understand and capture the structure and main features of solutions to problems on large time intervals.

After having been observed in the midst of the last century in the context of economics analysis, cf. [43], the turnpike property has since received interest in various fields of mathematics and economics, cf.,e.g., [7, 43, 53, 65, 75, 76, 79, 112, 134, 156, 157]. A particular kind of turnpike behavior is the so called exponential turnpike property, where the convergence of the dynamic problem's solution to the optimal steady state is exponential, cf. the recent works [24, 36, 62, 63, 113, 114, 123, 135, 136]. Recently, turnpike properties for non-observable systems [51, 111], for problems arising in deep learning [47] and for fractional parabolic problems [142] were presented.

As can be observed in Algorithm 1, an important feature of MPC is that only a first part on $[0, \tau]$ of the optimal control is used as a feedback and thus, only the solution on this part has to be computed accurately. This motivates the use of discretizations in space and time that are fine on $[0, \tau]$ and coarse on the remainder. In optimal control of dynamical systems, however, the optimal solution is subject to an adjoint equation which is formulated backwards in time. Hence, the optimal solution is subject to a fully coupled system of forward and backward equations and it is not clear a priori, that discretization errors stay local in time. However, if a turnpike property is present, it seems intuitively clear that perturbations of the system (e.g. by discretization errors) that occur in the far future will only marginally affect the optimal control at present time, cf. Figure 1.1.

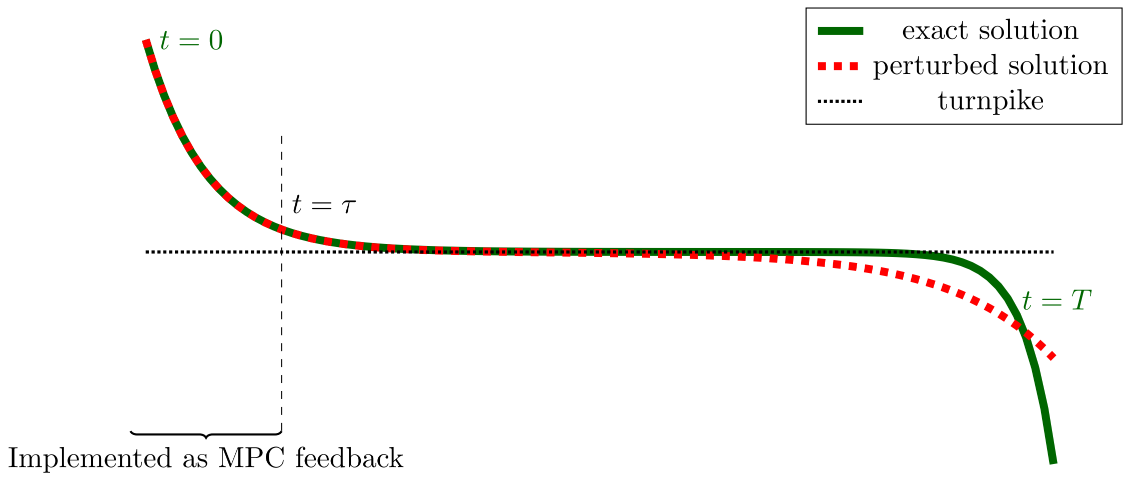


Figure 1.1: Depiction of steady state turnpike behavior and sketch of a solution being accurate only on an initial part.

One main goal of this thesis is to prove this property of exponential decay of perturbations under suitable stabilizability assumptions. We will analyze this topic in a very broad framework by means of sensitivity analysis and we will see that this property is very closely connected with the turnpike property, which can be interpreted as a property of exponential decay of perturbations of initial (and terminal) values. In that context, the abstract analysis presented in this thesis enables us to provide novel results in turnpike theory. Outside the context of MPC and turnpike theory, this property of decay of perturbations can also be used for efficient domain decomposition methods, cf. [105].

The stability and sensitivity analysis in this work will be carried out for problems governed by general evolution equations with bounded and unbounded control or observation and purely initial or initial and terminal condition. Moreover, we consider the case of OCPs governed by non-autonomous parabolic equations under a particular stabilizability condition. Further, we utilize the linear analysis to derive a local nonlinear result for semilinear parabolic equations. Eventually, we will show how the theoretical results lead to very efficient MPC schemes using goal oriented error estimation. We will present a particular objective for refinement that is tailored to an MPC context. Due to the exponential decay of perturbations shown before, we show for various examples that the controller performance is significantly increased when using this specialized refinement objective.

1.1 Contributions and outline

This work is organized as follows.

Chapter 1 - Introduction. The remainder of this chapter will consist of introducing the notation used in this work.

Chapter 2 - Sensitivity and turnpike analysis for linear quadratic optimal control of general evolution equations. We show for optimal control problems governed by strongly continuous semigroups that the influence of perturbations of the extremal equations decays exponentially in time if the operators satisfy a stabilizability and detectability assumption. Under the same assumptions, we provide a turnpike result. We prove these results for bounded control and observation operators in Theorem 2.27 and Theorem 2.30, unbounded but admissible control or observation in Theorem 2.48 and Theorem 2.49, and, under a controllability assumption, for problems including terminal conditions on the state in Theorem 2.55 and Theorem 2.56. We further provide sharper estimates for the particular case of a parabolic equation by a bootstrapping argument in Section 2.6.1 and by maximal parabolic regularity in Section 2.6.2. We accompany the theoretical results by various examples including heat and wave equations.

Chapter 3 - Sensitivity analysis for linear quadratic optimal control of non-autonomous parabolic equations. In this chapter we show in Theorem 3.14 for non-autonomous problems satisfying a particular stabilizability notion that perturbations of the optimality conditions decay exponentially in time. Moreover, assuming that the problem is autonomous, we derive a turnpike result in Sobolev norms in Theorem 3.16. Finally, in Section 3.3, we numerically illustrate the turnpike property for optimal control of a heat equation. Additionally, we put forward a priori time and space grid generation techniques specialized for MPC and evaluate their performance by means of examples with distributed and boundary control of a heat equation.

Chapter 4 - Sensitivity and turnpike analysis for nonlinear optimal control problems. We analyze nonlinear problems by formulating the extremal equations as a nonlinear operator equation. We first present an abstract implicit function theorem with scaled norms, which enables us to extend the sensitivity and turnpike results from the linear quadratic setting to a nonlinear setting. A central assumption in this result is a T-independent bound on the solution operator corresponding to the linearized system as well as T-uniform differentiability of the corresponding nonlinearities. We will present two applications of this abstract analysis and provide a turnpike and sensitivity result for finite dimensional problems in Corollary 4.18 and Corollary 4.19 and for semilinear parabolic problems in Corollary 4.30 and Corollary 4.31, respectively. Further, in Section 4.5, we illustrate the turnpike property by means of numerical examples of distributed control of a semilinear and boundary control of a quasilinear equation and evaluate the performance of different a priori grid generation techniques.

Chapter 5 - Goal oriented error estimation for Model Predictive Control. We utilize goal oriented a posteriori error estimation to efficiently and adaptively solve optimal control problems arising in a Model Predictive Controller. To this end, we formulate a particular functional for refinement tailored to MPC. We evaluate the space and time grids resulting from refinement via this specialized objective and compare it to classical a posteriori error estimation with respect to the cost functional. We prove in Theorem 5.2 and Theorem 5.6 under stabilizability conditions that if one uses a localized objective for refinement, the error indicators decay exponentially outside the support of this functional. Finally, we inspect the behavior and the performance gain from using this specialized goal oriented error estimator in an MPC loop. Thus, in Section 5.3, we present examples including autonomous and non-autonomous optimal control of linear, semilinear and quasilinear parabolic equations with distributed or boundary control and a domain with a reentrant corner. We conclude the chapter by providing implementation details for efficient adaptive nonlinear MPC algorithms covering topics such as parallelization, grid warm starts and solution warm starts.

1.2 Notation

Throughout this thesis $\Omega \subset \mathbb{R}^n$, $n \in \{2,3\}$ is considered to be a bounded domain with Lipschitz boundary $\partial\Omega$ in the sense of [50, Definition 4.4] and [60]. If $(X, \|\cdot\|_X)$ is a Banach space, we denote the topological dual space by X^* and the duality product by $\langle \cdot,\cdot \rangle_{X^*\times X}$, where $\langle \varphi, v\rangle_{X^*\times X} := \varphi(v)$ for $\varphi \in X^*$, $v \in X$. By $L_p(\Omega)$, $1 \le p \le \infty$ (and analogously for the boundary $\partial\Omega$), we denote the standard Lebesgue spaces of measurable functions $v\colon \Omega \to \mathbb{R}$ for which

$$\|v\|_{L_p(\Omega)} := \left(\int_\Omega |v(\omega)|^p \, d\omega\right)^{\frac{1}{p}} < \infty \qquad \text{for } 1 \le p < \infty,$$
$$\|v\|_{L_\infty(\Omega)} := \operatorname*{ess\,sup}_{\omega\in\Omega} |v(\omega)| < \infty.$$

By $\partial_i v$ we mean the (weak) derivative of a space dependent function v with respect to the i-th spatial variable. ∇v is the (weak) gradient of v and $\Delta v := \sum_{i=1}^n \frac{\partial^2 v}{\partial x_i}$ the Laplacian. For integers m, p we denote by $W^{m,p}(\Omega)$ the usual Sobolev space endowed with the norm

$$\|v\|_{W^{m,p}(\Omega)} := \left(\sum_{0\le|\alpha|\le m} \int_\Omega \|D^\alpha v(\omega)\|^p \, d\omega\right)^{\frac{1}{p}},$$

where $\alpha = (\alpha_1, \dots, \alpha_n)$, $|\alpha| = \sum_{i=1}^n \alpha_i$ and D^α denotes the mixed (weak) partial derivative. We adopt the usual notation and write $H^m(\Omega) = W^{m,2}(\Omega)$. We will denote by $\mathrm{tr} : H^1(\Omega) \to L_2(\partial\Omega)$ the Dirichlet trace operator, cf. [138, Theorem 2.1] or [107, Section 2]. By $H_0^1(\Omega)$, we mean all functions in $H^1(\Omega)$ that are zero a.e. on the boundary. By $H^{-1}(\Omega)$ we denote the topological dual of $H_0^1(\Omega)$. For a precise definition of these Sobolev spaces the reader is referred to [1, Chapter 3]. For a Lebesgue exponent $1 \le p \le \infty$, we will write p' for the dual exponent, i.e., $\frac{1}{p} + \frac{1}{p'} = 1$, where we use the convention $\frac{1}{\infty} = 0$.

Let $[0,T]$ a bounded proper interval. If $(X, \|\cdot\|_X)$ is a Banach space, we denote by $L_p(0,T;X)$ for $1 \le p \le \infty$ the space of (Bochner)-measurable functions $v : (0,T) \to X$ for which we have

$$\|v\|_{L_p(0,T;X)} := \left(\int_0^T \|v(t)\|_X^p \, dt\right)^{\frac{1}{p}} < \infty \qquad \text{if } p < \infty,$$
$$\|v\|_{L_\infty(0,T;X)} := \operatorname*{ess\,sup}_{t\in[0,T]} \|v(t)\|_X < \infty.$$

$C(0,T;X)$ denotes the space of all continuous functions $v : [0,T] \to X$ with norm

$$\|v\|_{C(0,T;X)} := \max_{t\in[0,T]} \|v(t)\|.$$

For a precise definition of Bochner spaces of vector-valued functions, the reader is referred to [158, Section 23.2] and [138, Section 3.4]. We will gather a few important properties of these spaces.

$C(0,T;X)$ and $L_p(0,T;X)$ together with the respective norms form Banach spaces, in case of $L_p(0,T;X)$ after forming equivalence classes of functions who are equal a.e.. If X is a Hilbert space with scalar product $\langle \cdot,\cdot \rangle_X$, then $L_2(0,T;X)$ is, with the scalar product

$$\langle u,v\rangle_{L_2(0,T;X)} = \int_0^T \langle u(t),v(t)\rangle_X \, dt.$$

If $(Y,\|\cdot\|_Y)$ is a Banach space and if the embedding $X \hookrightarrow Y$ is continuous, then $L_p(0,T;X) \hookrightarrow L_p(0,T;Y)$ continuously for $1 \le p \le \infty$. Moreover we have that $L_p(0,T;X)^* \cong L_{p'}(0,T;X^*)$. Analogously, for any measurable subset $S \subset \mathbb{R}^n$, $n \in \mathbb{N}$ we denote by $L_p(S;X)$ the space of functions $v : S \to X$ such that

$$\begin{aligned} \|v\|_{L_p(S;X)} &:= \left(\int_S \|v(s)\|_X^p \, ds \right)^{\frac{1}{p}} < \infty \qquad \text{if } p < \infty, \\ \|v\|_{L_\infty(S;X)} &:= \operatorname*{ess\,sup}_{s \in S} \|v(s)\|_X < \infty. \end{aligned}$$

Eventually, we will denote the space of infinitely differentiable test functions $\varphi : [0,T] \to X$ by $C^\infty([0,T];X)$ and if $X = \mathbb{R}$, we may write $C^\infty(0,T) = C^\infty([0,T];X)$. For a function $v : [0,T] \to X$ we mean by v' or $\frac{d}{dt}v$ the (distributional) time derivative of v.

If $(X,\|\cdot\|_X)$ and $(Y,\|\cdot\|_Y)$ are Banach spaces we denote by $L(X,Y)$ the space of linear bounded operators from X to Y endowed with the usual norm

$$\|M\|_{L(X,Y)} := \sup_{\|x\|_X \neq 0} \frac{\|Mx\|_Y}{\|x\|_X}$$

and we may abbreviate $L(X) = L(X,X)$.

Chapter 2

Sensitivity and turnpike analysis for linear quadratic optimal control of general evolution equations

In this chapter, we will analyze the sensitivity of general optimal control problems that are subject to dynamics governed by a strongly continuous semigroup. Strongly continuous semigroups are a very powerful tool for studying linear dynamical systems and their properties, cf. [44, 109, 133, 153]. The case of (optimal) control of dynamical systems using a semigroup approach is extensively treated in, e.g., [19, 35, 90, 91, 95, 132, 139, 154]. We will utilize this theory to obtain sensitivity results in a very general setting. Additionally, we will make use of the concept of admissible control and observation operators and well-posed linear systems as introduced in the seminal papers [124, 125, 145], the monographs [132, 139] and the survey articles [140, 146] in order to cover the case of unbounded observation or control operators. Such unboundedness can occur when the control or observation acts on the boundary or at isolated points. Semigroup theory and admissibility can be seen as the most general framework to obtain trajectories that are continuous in time, which itself is crucial to make sense of initial conditions.

The analysis presented here is based on investigating the first-order necessary optimality conditions (sometimes also referred to as extremal equations) of the optimal control problem and characterizing their stability via bounds of the corresponding solution operator that are independent of the time horizon T. The key to establishing these uniform operator bounds are stabilizability and detectability assumptions and the main step is to consider special test functions, similar to [113] and [135], that decay exponentially. Consequently, we will be able to show that perturbations of the extremal equations' dynamics decay exponentially in time. Concerning temporal regularity, we show uniform estimates as well as L_2-type estimates for perturbations of L_2 and L_1 temporal integrability. As described in Chapter 1, an important motivation for our sensitivity analysis is Model Predictive Control (MPC). The analysis in this part shows that even in a very general setting, under appropriate stability assumptions, it can be shown that perturbations occurring towards the end of the optimization horizon only have negligible influence on the MPC feedback, if the optimization horizon is large. As stated in Chapter 1, this particular feature will allow for a very efficient adaptive discretization of optimal control problems governed by PDEs in a Model Predictive Controller, i.e., only refining the spatial and temporal grid on the initial part.

As a second result, we show an exponential turnpike property as depicted in Figure 1.1. The proofs establishing the turnpike property in recent works [24, 135] are based on a stabilizability and detectability assumption on the system. We will also depend on these assumptions, however under significantly weaker structural assumptions on the semigroup, which allows us to extend the existing results to a very general setting. In particular, turnpike theorems in Hilbert spaces were given in [135] for general strongly continuous semigroups with bounded control and observation operators as well as for boundary controlled parabolic equations. The proofs in [135], however, make use of the Algebraic Riccati Equation—a theory that is well established for admissible boundary control of parabolic equations but not for general evolution equations. Here, we will show a turnpike result for unbounded but admissible control of non-parabolic equations that has not been available until now. This is possible, as we avoid using Riccati theory in our approach. Moreover, we present results in the case of initial and terminal conditions on the state under a controllability condition. To the authors' best knowledge, such a result was also not available in a general Hilbert space setting.

Additionally, the analysis sheds light on the close connection of exponential sensitivity analysis and the turnpike property, both emerging from the T-uniform boundedness of the operator corresponding to the extremal equations. This becomes clear by comparing the abstract scaling results in Theorem 2.27 and Theorem 2.30.

Finally, we will see two approaches for refining the sensitivity estimates if one assumes additional structure on the system, i.e., that the underlying semigroup is analytic. For the turnpike case, this relates to the analysis performed in [24], where the authors deduce a turnpike property for analytic semigroups in Sobolev norms.

We will accompany all theoretical considerations by various examples of parabolic and hyperbolic systems with boundary or distributed control and observation.

Structure. First, in Section 2.1 we present our theoretical framework, the optimal control problem and optimality conditions. In the first part of Section 2.2, namely Section 2.2.1, we derive a general result on the propagation of perturbations over time in Theorem 2.27, under the assumption that various norms of the extremal equations' solution operator, which itself may indeed depend on the horizon T, can be bounded independently of T. Further, in Section 2.2.2, under the same assumptions of T-independent bounds on several solution operator's norms, we show a turnpike result in Theorem 2.30 for general evolution equations. In Section 2.2.3 we show that the T-independent bounds on these operator norms hold, if the dynamics are exponentially stabilizable and detectable. In Section 2.3, we will extend the results to unbounded but admissible control operators and discuss the necessary modifications to the proofs. Then, in Section 2.4, under an exact controllability assumption, we extend our results to the case of a terminal condition on the state. We then present two examples that fulfill the assumptions of our analysis, namely the interior control of a heat equation and the Dirichlet boundary control of a wave equation, in Section 2.5. Eventually, in Section 2.6, we discuss the case of an analytic semigroup and present two approaches to obtain stronger estimates. Finally, we will illustrate these refined results by means of an example with a heat equation.

This chapter comprises the results of [71]. In Section 2.6, we present previously unpublished results considering the particular case of an analytic semigroup.

2.1 Setting and preliminaries

In this section, we will introduce the solution concept for the dynamics we will consider, namely the mild solution defined by a strongly continuous semigroup. In the first part we will cover homogeneous equations, whereas in the second, we will include nonzero right hand sides. Last, we will move to optimal control problems involving strongly continuous semigroups and recall known results concerning existence of minimizers and optimality conditions. The majority of this introduction is based on the books [35, 44, 109].

2.1.1 Strongly continuous semigroups and their generators

We are interested in solutions of an abstract dynamical system described by

$$x' = Ax, \qquad x(0) = x_0, \tag{ACP}$$

where $A\colon D(A) \subset X \to X$ is a possibly unbounded but closed operator, $x_0 \in X$ is an initial datum and X is a Hilbert space with scalar product and induced norm denoted by $\langle \cdot, \cdot \rangle$ and $\| \cdot \|$, respectively. This initial value problem is often referred to as an *abstract Cauchy problem* (ACP). To facilitate notation, we will indicate the operator norm by the same norm symbol, i.e., for $T \in L(X, X) =: L(X)$ we set $\|T\| = \|T\|_{L(X)}$ if no ambiguity is possible.

In order to discuss the existence of solutions, a very powerful concept is to characterize the solutions of the system above via a family of linear operators, parameterized by time, mapping initial values to the state at the current time. Three important features are demanded from this family: First, a semigroup property, second, that the operator corresponding to time zero is the identity on X, and third, a strong continuity property, i.e., continuity in time at time zero for all initial values. This family of operators is called a *strongly continuous semigroup* or C_0*-semigroup*, denoted by $(\mathcal{T}(t))_{t\geq 0}$. We will only consider semigroups that are strongly continuous. Thus, for the sake of brevity, we will sometimes not explicitly annotate the strong continuity and only write semigroup.

Definition 2.1 (Strongly continuous semigroup)**.** *An operator valued map $\mathcal{T}\colon \mathbb{R}_{\geq 0} \to L(X)$ is called a strongly continuous semigroup if the following conditions are satisfied:*

i) $\mathcal{T}(t)\mathcal{T}(s) = \mathcal{T}(t+s) \quad \forall t, s \geq 0,$

ii) $\mathcal{T}(0) = I,$

iii) $\|\mathcal{T}(t)x_0 - x_0\| \overset{t\to 0}{\to} 0$ *for all* $x_0 \in X$.

A direct consequence of this definition is continuity of trajectories $x(t) = \mathcal{T}(t)x_0$.

Theorem 2.2. *Consider a strongly continuous semigroup* $(\mathcal{T}(t))_{t\geq 0}$*. Then the map* $t \mapsto \mathcal{T}(t)x_0$ *is continuous for all* $t \geq 0$ *and* $x_0 \in X$.

Proof. See [109, Corollary 2.3]. □

In order to establish a connection between the abstract Cauchy problem (ACP) and a strongly continuous semigroup, we define the infinitesimal generator.

Definition 2.3 (Infinitesimal generator)**.** *A linear operator* $A\colon D(A) \subset X \to X$ *is called the infinitesimal generator of a strongly continuous semigroup* $(\mathcal{T}(t))_{t\geq 0}$ *if*

$$Ax_0 = \lim_{t\to 0} \frac{\mathcal{T}(t)x_0 - x_0}{t} \qquad \forall x_0 \in D(A),$$

where $D(A) := \left\{x_0 \in X \mid \lim_{t\to 0} \frac{\mathcal{T}(t)x_0 - x_0}{t} \text{ exists}\right\}$ *is called the domain of* A.

Remark 2.4. *Another class of operator semigroups is formed by uniformly continuous semigroups, which can be defined via continuity at zero in the uniform operator topology, i.e.,* $\|\mathcal{T}(t) - I\|_{L(X)} \overset{t\to 0}{\to} 0$ *as opposed to the strong operator topology in Definition 2.1 iii). It can be shown that every uniformly continuous semigroup is of the form*

$$\mathcal{T}(t) = e^{tA} := \sum_{k=0}^{\infty} \frac{t^k A^k}{k!}, \tag{2.1}$$

where $A \in L(X)$ *[44, Chapter I, Theorem 3.7]. Moreover, boundedness of the generator on* X*, closedness of* $D(A)$ *in* X *and uniform continuity of the semigroup are equivalent, cf. [109, Theorem 1.2] or [44, Chapter II, Corollary 1.5]. In this case,* $t \mapsto \mathcal{T}(t)x_0$ *is continuously differentiable in* t*, cf. [44, p.48f] or [109, Corollary 1.4]. However, demanding* A *to be bounded on* X *is too restrictive in terms of applications, e.g., if* A *is the Laplace operator and* $X = L_2(\Omega)$.

The definition of the generator can also be interpreted as the derivative of the orbit map $t \mapsto \mathcal{T}(t)x_0$ at time $t = 0$. The following theorem shows that if $x_0 \in D(A)$, then the orbit maps are differentiable for $t \geq 0$. In this case semigroup and generator commute. It additionally shows that not only the generator and its domain are defined uniquely by the strongly continuous semigroup, but that the converse is also true.

Theorem 2.5. *Let* A *with corresponding domain* $D(A)$ *be the generator of a strongly continuous semigroup* $(\mathcal{T}(t))_{t\geq 0}$*. Then, the following properties hold:*

i) $A\colon D(A) \subset X \to X$ *is a linear operator.*

ii) A *is a closed and densely defined operator that defines the strongly continuous semigroup* $(\mathcal{T}(t))_{t\geq 0}$ *uniquely.*

iii) If $x_0 \in D(A)$, then $\mathcal{T}(t)x_0 \in D(A)$ for all $t \geq 0$ and

$$\frac{d}{dt}\mathcal{T}(t)x_0 = \mathcal{T}(t)Ax_0 = A\mathcal{T}(t)x_0 \qquad \forall t \geq 0.$$

Proof. See [44, Chapter II, Lemma 1.3 and Theorem 1.4] or [109, Theorem 2.4 and Corollary 2.5]. □

We can now define mild and classical solutions to the abstract problem (ACP).

Definition 2.6 (Solution concepts, homogeneous case)**.** *Consider (ACP) and let A generate a strongly continuous semigroup $(\mathcal{T}(t))_{t\geq 0}$.*

i) For $x_0 \in X$, we call $x(t) := \mathcal{T}(t)x_0 \in C(0,T;X)$ the mild solution of the initial value problem (ACP).

ii) A function $x\colon \mathbb{R}_{\geq 0} \to X$ is called a classical solution *of (ACP) if it satisfies (ACP) in the classical sense, i.e.,*

(a) $x'(t) \in X$, $x(t) \in D(A)$ and $x'(t) = Ax(t)$ in X $\forall t \geq 0$,

(b) $x(0) = x_0$.

While the classical solution is defined via the initial value problem (ACP), the mild solution is defined via the strongly continuous semigroup $(\mathcal{T}(t))_{t\geq 0}$. However, the two solution concepts are strongly connected, with mild solutions being a generalization of classical solutions, as the following theorem shows. For a further discussion on the topic of well-posedness and existence of particular types of solutions, the interested reader is referred to [44, Section II.6].

Theorem 2.7 (Relation of classical and mild solution)**.** *A classical solution to (ACP) exists if and only if $x_0 \in D(A)$. In this case, the mild solution and the classical solution coincide.*

Proof. The proof follows directly from the definition of mild and classical solutions and Theorem 2.5. □

Up to now, we introduced the semigroup and its infinitesimal generator. In many theoretical considerations, a third component comes into play: the *resolvent operator*. This operator will play a role when we investigate a particular class of semigroups, namely analytic semigroups, at the end of this chapter, cf. Section 2.6.

Definition 2.8 (Resolvent operator)**.** *The set*

$$\rho(A) := \{\lambda \in \mathbb{C} \,|\, \lambda I - A\colon D(A) \to X \text{ is bijective}\}$$

is called the resolvent set of A. By the closed graph theorem [84, Theorem 5.20], for any $\lambda \in \rho(A)$, the operator

$$R(\lambda, A) := (\lambda I - A)^{-1}$$

is a bounded linear operator in X, called the resolvent operator. The complement of $\rho(A)$ in $\mathbb{C}$ is called the spectrum of A which we will denote by $\sigma(A)$.

Remark 2.9. *It is common to define the semigroup first and then the generator as the right derivative of the orbit maps in zero. In certain cases, there are also ways to define the semigroup by the generator or the resolvent. A strongly continuous semigroup can sometimes be defined via a Cauchy integral formula* $\mathcal{T}(t) := \frac{1}{2\pi i}\int_{\partial U} e^{\lambda t} R(\lambda, A)\, d\lambda$ *where* $U \subset \mathbb{C}$ *is an open neighborhood of* $\sigma(A)$ *with smooth positively oriented boundary* ∂U *(cf. [44, Section II.4.a]), an analogon to Eulers formula* $\mathcal{T}(t) := \lim_{n\to\infty}(\frac{n}{t}R(\frac{n}{t}, A))^n$ *([44, Chapter III, Corollary 5.5]), or by approximating the unbounded generator* A *by a sequence of bounded operators* $(A_n)_{n\in\mathbb{N}}$ *and defining the semigroup via* $\mathcal{T}(t) := \lim_{n\to\infty} e^{tA_n}$ *with the exponential defined in* (2.1) *(*A_n *are called Yosida approximations, cf. [44, Chapter II, Theorem 3.5]).*

In concrete applications, the semigroup $(\mathcal{T}(t))_{t\geq 0}$ is unknown, whereas the operator A resp. the initial value problem (ACP) is known. In this case, it is necessary to show that A indeed is the generator of a strongly continuous semigroup. We shortly present the most important theorems that establish such a result. The most general theorem is the *Hille-Yosida theorem*, see [44, Chapter II, Theorem 3.8] and [35, Theorem 2.1.12]. The second one is the Lumer-Phillips theorem for dissipative operators A, i.e., operators such that $\|(\lambda - A)x\| \geq \lambda\|x\|$ for all $x \in D(A)$ and $\lambda > 0$, which, under additional assumptions, generate a contraction semigroup, i.e., a semigroup such that $\|\mathcal{T}(t)\| \leq 1$ cf. [109, Theorem 4.3]. In a Hilbert space setting, Stone's theorem states that any densely defined skew adjoint operator $A^* = -A$ generates a unitary group, i.e., $\|\mathcal{T}(t)\| = 1$ for all $t \in \mathbb{R}$, see, e.g., [44, Chapter II, Theorem 3.24].

In the following, we will discuss the asymptotic behavior of $\|\mathcal{T}(t)\|$ for $t \to \infty$.

Definition 2.10 (Type)**.** *The number* $\omega_0(\mathcal{T}) := \inf_{t>0} \frac{1}{t}\log\|\mathcal{T}(t)\|$ *is called the* type *of a strongly continuous semigroup* $(\mathcal{T}(t))_{t\geq 0}$*.*

Theorem 2.11. *Let* $(\mathcal{T}(t))_{t\geq 0}$ *be a strongly continuous semigroup. Then*

i) $\omega_0(\mathcal{T})$ *is finite or* $-\infty$*.*

ii) For every $\omega > \omega_0(\mathcal{T})$*, there exists* $M_\omega \geq 1$ *such that*

$$\|\mathcal{T}(t)\| \leq M_\omega e^{\omega t} \qquad \forall t \geq 0.$$

Proof. For i) see [19, Part II-1, Proposition 2.2] and for ii) see [19, Part II-1, Corollary 2.1]. □

Definition 2.12 (Exponential stability)**.** *A strongly continuous semigroup* $(\mathcal{T}(t))_{t\geq 0}$ *is called* exponentially stable *if there exists* $\alpha > 0$ *and* $M \geq 1$ *such that*

$$\|\mathcal{T}(t)\| \leq Me^{-\alpha t} \qquad \forall t \geq 0.$$

The following theorem is a slightly modified version of [19, Part II-1, Theorem 2.2] and sheds light on the connection of stability of a strongly continuous semigroup and its type. Additionally, it shows that whenever the operator norm of a strongly continuous semigroup decays to zero for $t \to \infty$, then it decays exponentially.

Theorem 2.13 (Characterizations of exponential stability)**.** *Let $(\mathcal{T}(t))_{t\geq 0}$ be a strongly continuous semigroup with generator A and $1 \leq p < \infty$. Then the following are equivalent:*

i) $\omega_0(\mathcal{T}) < 0$.

ii) There is a constant $c > 0$ such that

$$\int_0^\infty \|\mathcal{T}(t)x_0\|^p \leq c^p \|x_0\|^p \qquad \forall x_0 \in X.$$

iii) $(\mathcal{T}(t))_{t\geq 0}$ *is exponentially stable in the sense of Definition 2.12.*

iv) $(\mathcal{T}(t))_{t\geq 0}$ *is asymptotically stable in $L(X)$, i.e.,*

$$\|\mathcal{T}(t)\| \to 0 \qquad \text{for } t \to \infty.$$

In particular, for all $\omega > \omega_0(\mathcal{T})$ the operator $A - \omega I$ generates an exponentially stable semigroup $(\mathcal{T}_\omega(t))_{t\geq 0}$ with

$$\mathcal{T}_\omega(t) = e^{-\omega t}\mathcal{T}(t) \qquad \forall t \geq 0.$$

Proof. See [19, Part II-1, Theorem 2.2] and [19, Part II-1, Corollary 2.2] □

Remark 2.14. *The equivalence of Theorem 2.13 ii) and exponential stability is also known as the Datko-Pazy theorem, cf. [109, Theorem 4.1]. The characterization Theorem 2.13 ii) can also be interpreted as T-uniform boundedness of the solution operator $S\colon x_0 \to \mathcal{T}(t)x_0$ to the abstract Cauchy problem as a map from X to $L_p(0,T;X)$, i.e., $S \in L(X, L_p(0,T;X))$ for $1 \leq p < \infty$ with operator norm independent of the time horizon T. This interpretation will be useful in the next section, where we will establish T-independent bounds on solution operator norms under stabilizability conditions.*

2.1.2 Inhomogeneous equations

While we only considered homogeneous initial value problems in the previous subsection, we will discuss the solutions to inhomogeneous equations in this part. To this end we replace the problem of interest (ACP) by the inhomogeneous abstract Cauchy problem

$$x' = Ax + f, \qquad x(0) = x_0, \tag{iACP}$$

where again $x_0 \in X$ and $A\colon D(A) \subset X \to X$ is a possibly unbounded but closed operator.

Definition 2.15 (Mild solution, inhomogeneous case)**.** *Let A generate a strongly continuous semigroup $(\mathcal{T}(t))_{t\geq 0}$ and consider $x_0 \in X$ and $f \in L_1(0,T;X)$. The function*

$$x(t) := \mathcal{T}(t)x_0 + \int_0^t \mathcal{T}(t-s)f(s)\,ds \tag{2.2}$$

is called the mild solution *of the inhomogeneous initial value problem* (iACP).

The formula (2.2) is sometimes referred to as the variation of constants formula. By the definition above, the mild solution exists, it is continuous in time and depends continuously on the problem data. The following lemma shows that the mild solution is the unique solution of (iACP). We again refer to classical solutions, being defined completely analogous to classical solutions of homogeneous equations, cf. Definition 2.6.

Lemma 2.16 (Uniqueness of solutions, [109, Chapter 4, Corollary 2.2]). *Let $f \in L_1(0, T; X)$ and $x_0 \in X$. Then* (iACP) *has at most one classical solution which is a mild solution in the sense of Definition 2.15.*

Definition 2.17 (Weak solution). *Let $x_0 \in X$. A function $x \in C(0, T; X)$ is called a* weak solution *of the inhomogeneous initial value problem* (iACP) *if*

i) $x(0) = x_0$,

ii) $t \to \langle x(t), v\rangle$ *is absolutely continuous for* $v \in D(A^*)$,

iii) $\frac{d}{dt}\langle x(t), v\rangle = \langle x(t), A^* v\rangle + \langle f(t), v\rangle$ *for* $v \in D(A^*)$ *and a.e.* $t \in [0, T]$.

Theorem 2.18 (Equivalence of weak and mild solution [9]). *Let $x_0 \in X$ and $f \in L_1(0, T; X)$. There exists a unique weak solution of* (iACP) *if and only if A generates a strongly continuous semigroup on X, and in this case the weak solution is the mild solution, i.e., satisfies* (2.15).

Remark 2.19. *In the inhomogeneous case, one can define further meaningful solution concepts than the mild, weak and classical solutions presented here. In particular, we refer to the notion of strict and strong solutions, cf. [19, Part II-1, Definition 3.1].*

In view of optimal control, we will need the definition of an adjoint semigroup. The following theorem shows that, in a Hilbert space, the semigroup consisting of the adjoint operators is generated by the adjoint of the generator.

Theorem 2.20 (Dual semigroup, [35, Theorem 2.2.6]). *If A with domain $D(A)$ generates a strongly continuous semigroup $(\mathcal{T}(t))_{t\geq 0}$ on a Hilbert space X, then A^* with domain $D(A^*)$ generates the dual semigroup $(\mathcal{T}(t)^*)_{t\geq 0}$ on X.*

Remark 2.21 (Backwards-in-time equations). *Let A generate a strongly continuous semigroup $(\mathcal{T}(t))_{t\geq 0}$ and consider a backwards differential equation*

$$-x' = Ax + f, \qquad x(T) = x_T$$

with terminal condition $x_T \in X$. By a simple time transformation $t \to T - t$, it can be seen that the unique mild solution is given by

$$x(t) = \mathcal{T}(T - t)x_T + \int_t^T \mathcal{T}(s - t)f(s)\, ds.$$

We furthermore recall a well-known scaling argument, of which we will make use in the sensitivity analysis. Whenever we refer to the solution of an abstract Cauchy problem, it is meant in the sense of the mild solution, cf. Definition 2.6 i) resp. Definition 2.15.

Lemma 2.22. *Let A generate a strongly continuous semigroup $(\mathcal{T}(t))_{t\geq 0}$ on X, $f_1, f_2 \in L_1(0,T;X)$ and $x_0, x_T \in X$. Assume $x_1, x_2 \in C(0,T;X)$ solve the abstract Cauchy problems*

$$\begin{aligned} x_1' &= Ax_1 + f_1, & x_1(0) &= x_0, \\ -x_2' &= A^*x_2 + f_2, & x_2(T) &= x_T. \end{aligned}$$

Then for any $\mu \in \mathbb{R}$:

i) $\widetilde{x}_1(t) := e^{-\mu t}x_1(t)$ and $\widetilde{x}_2(t) := e^{-\mu t}x_2(t)$ solve

$$\begin{aligned} \widetilde{x}_1' &= (A-\mu I)\widetilde{x}_1 + e^{-\mu t}f_1, & \widetilde{x}_1(0) &= x_0, \\ -\widetilde{x}_2' &= (A+\mu I)^*\widetilde{x}_2 + e^{-\mu t}f_2, & \widetilde{x}_2(T) &= e^{-\mu T}x_T. \end{aligned}$$

ii) For all $0 \leq s \leq t \leq T$,

$$\langle x_1(t), x_2(t)\rangle - \langle x_1(s), x_2(s)\rangle = \int_s^t (\langle x_2(\tau), f_1(\tau)\rangle - \langle f_2(\tau), x_1(\tau)\rangle)\, d\tau.$$

Proof. For i), we multiply the variation of constants formula for $x(t)$ by $e^{-\mu t}$ and get

$$\tilde{x}_1(t) = e^{-\mu t}\left(\mathcal{T}(t)x_0 + \int_0^t \mathcal{T}(t-s)f_1(s)\,ds\right) = e^{-\mu t}\mathcal{T}(t)x_0 + \int_0^t e^{-\mu(t-s)}\mathcal{T}(t-s)e^{-\mu s}f_1(s)\,ds.$$

Moreover, if a semigroup $(\mathcal{T}(t))_{t\geq 0}$ has generator A, the scaled semigroup $(e^{-\mu t}\mathcal{T}(t))_{t\geq 0}$ has generator $A-\mu I$ [44, p.60] with the same domain as A, as the domain does not change under bounded perturbations, cf. [44, Chapter III]. The result for $\tilde{x}_2$ follows analogously. For ii), see [95, Proposition 5.7]. □

2.1.3 Optimal control with bounded control and observation

In this part, we will consider the case of optimal control of dynamics governed by the generator of a strongly continuous semigroup. Two further ingredients will come into play: On the one hand an input operator B, which allows us to influence the dynamics via, e.g., actuators, and on the other hand an output operator C, that could model, e.g., sensors.

Problem 2.23.

$$\begin{aligned} \min_{(x,u)} \frac{1}{2}\int_0^T \|C(x(t)-x_d(t))\|_Y^2 + \|R(u(t)-u_d(t))\|_U^2\,dt \\ s.t.\ x' = Ax + Bu + f, \\ x(0) = x_0 \end{aligned} \tag{2.3}$$

with the following standing assumptions:

i) $T > 0$ is a fixed time horizon,

ii) X is a real Hilbert space and $A\colon D(A) \subset X \to X$ is a (possibly unbounded) generator of a strongly continuous semigroup $(\mathcal{T}(t))_{t\geq 0}$ on X, $f \in L_1(0,T;X)$ and $x_0 \in X$,

iii) U is a real Hilbert space with scalar product $\langle\cdot,\cdot\rangle_U$ and induced norm $\|\cdot\|_U$, $B \in L(U,X)$, $u_d \in L_2(0,T;U)$,

iv) $R \in L(U,U)$ with $\|Ru\|_U^2 \geq \alpha\|u\|_U^2$ for $\alpha > 0$ and all $u \in U$,

v) Y is a real Hilbert space with scalar product $\langle\cdot,\cdot\rangle_Y$ and induced norm $\|\cdot\|_Y$, $C \in L(X,Y)$, $x_d \in L_2(0,T;X)$.

Theorem 2.24 (Existence of optimal solution and optimality conditions)**.** *There exists a unique minimizer $(x,u) \in C(0,T;X) \times L_2(0,T;U)$ to Problem 2.23. Further, there is an adjoint state $\lambda \in C(0,T;X)$ such that*

$$\begin{aligned} C^*Cx - \lambda' - A^*\lambda &= C^*Cx_d, \\ R^*Ru - B^*\lambda &= R^*Ru_d, \\ x' - Ax - Bu &= f, \end{aligned} \tag{2.4}$$

$\lambda(T) = 0$ and $x(0) = x_0$. The second equation is to be understood in $U^* \cong U$ for a.e. $t \in [0,T]$ and the first and third in a mild sense along $[0,T]$.

Proof. First, observe that the control-to-state map $S\colon L_2(0,T;U) \to C(0,T;X)$ is given by

$$x(t) = (Su)(t) := \mathcal{T}(t)x_0 + \int_0^t \mathcal{T}(t-s)(Bu(s) + f(s))\,ds. \tag{2.5}$$

Inserting this into the cost functional yields the reduced cost functional

$$J(u) := \frac{1}{2}\int_0^T \|C((Su)(t) - x_\mathrm{d}(t))\|_Y^2 + \|R(u(t) - u_\mathrm{d}(t))\|_U^2\,dt,$$

where $J\colon L_2(0,T;U) \to \mathbb{R}$ is radially unbounded, i.e., $J(u) \to \infty$ if $\|u\|_{L_2(0,T;U)} \to \infty$ due to $\|Ru\|_U \geq \alpha\|u\|_U$ $\forall u \in U$, cf. Assumption iv) in Problem 2.23. By standard arguments, this yields the existence of an optimal control, cf. [138, Theorem 2.14]. In order to derive optimality conditions, we take the derivative of J at the optimal control $u \in L_2(0,T;U)$ in direction $\delta u \in L_2(0,T;U)$:

$$J'(u)\delta u = \int_0^T \langle C((Su)(t) - x_\mathrm{d}(t)), C(\int_0^t \mathcal{T}(t-s)B\delta u(s)\,ds)\rangle_Y + \langle R(u(t) - u_\mathrm{d}(t)), R\delta u(t)\rangle_U\,dt.$$

The adjoint of $(L\delta u)(t) := \int_0^t \mathcal{T}(t-s)B\delta u(s)\,ds$ as a mapping from $L_2(0,T;U)$ to $L_2(0,T;X)$ is given by $(L^*d)(t) = \int_t^T B^*\mathcal{T}^*(s-t)d(s)\,ds$, cf. [90, Section 0.4]. Further, requiring $J'(u)\delta u = 0$ for all $\delta u \in L_2(0,T;U)$ yields

$$\int_0^T \langle \int_t^T B^*\mathcal{T}^*(s-t)C^*C((Su)(s) - x_\mathrm{d}(s))\,ds, \delta u(t)\rangle_U + \langle R^*R(u(t) - u_\mathrm{d}(t)), \delta u(t)\rangle_U\,dt = 0.$$

As this equation needs to be fulfilled for all $\delta u(t) \in L_2(0,T;U)$, we get

$$B^* \int_t^T \mathcal{T}^*(s-t)C^*C((\mathcal{S}u)(s) - x_{\mathrm{d}}(s))\,ds + R^*R(u(t) - u_{\mathrm{d}}(t)) = 0$$

for a.e. $t \in [0,T]$. Defining $\lambda(t) = -\int_t^T \mathcal{T}^*(s-t)C^*C((\mathcal{S}u)(s) - x_{\mathrm{d}}(s))\,ds$ together with (2.5) yields the system (2.4). □

Remark 2.25. *A different and more involved proof of existence of a solution and optimality conditions for bounded control and observation in a nonlinear setting is given in [95, Chapter 3 and 4]. However, in this linear-quadratic setting, we presented a simpler proof, as this yields the possibility to be extended to the unbounded control case, cf. the discussion in Remark 2.44.*

In order to simplify notation and for a clear presentation, we will rewrite the optimality system as a linear operator equation.

Definition 2.26 (Time evaluation operator)**.** *For $t \in [0,T]$, we define a linear bounded operator $E_t\colon C(0,T;X) \to X$ by $E_t x := x(t)$ for $x \in C(0,T;X)$.*

Defining $Q := R^*R$ and eliminating the control via the second equation with $u = Q^{-1}B^*\lambda + u_{\mathrm{d}}$ leads to the linear system of equations

$$\underbrace{\begin{pmatrix} C^*C & -\frac{d}{dt} - A^* \\ 0 & E_T \\ \frac{d}{dt} - A & -BQ^{-1}B^* \\ E_0 & 0 \end{pmatrix}}_{=:M} \begin{pmatrix} x \\ \lambda \end{pmatrix} = \begin{pmatrix} C^*Cx_{\mathrm{d}} \\ 0 \\ Bu_{\mathrm{d}} + f \\ x_0 \end{pmatrix}. \tag{2.6}$$

The operator M corresponds to the two abstract inhomogeneous evolution equations (2.4) with initial and terminal condition after elimination of the control, i.e., the adjoint equation in the first two rows and the state equation in the last two rows, and allows for a brief notation. The solution operator of this system, which we denote by M^{-1}, maps initial values and source terms for the state and the adjoint equation to the solution. A central question in the following will be the dependence of the norm of M^{-1} on the time T. Here we recall Remark 2.14, where we observed that for strongly continuous semigroups the solution operator has an T-independent bound as operator from X to $L_p(0,T;X)$ with $1 \le p < \infty$ if and only if the semigroup is exponentially stable. In the optimal control setting, instead of assuming exponential stability of the strongly continuous semigroup generated by A, a weaker notion, namely stabilizability and detectability of (A,B) resp. (A,C) will suffice to derive T-independent bounds for the solution operator M^{-1}.

2.2 The case of bounded control and observation

This section is split up into three major parts. The first two subsections give two preliminary results, the first stating that perturbations of the right hand side stay local in time, whereas the second yields an exponential turnpike result. They are preliminary in the sense that they include assumptions on T-independent bounds on M^{-1} as defined in (2.6). Under a stabilizability and detectability assumption, these bounds will be derived in the third part of this section.

2.2.1 An abstract exponential sensitivity result

We will refer to the solution $(x, \lambda) \in C(0,T;X)^2$ of (2.6) as the *exact solution* and assume that there is a second pair of variables $(\tilde{x}, \tilde{\lambda}) \in C(0,T;X)^2$ that satisfies the perturbed system

$$\begin{pmatrix} C^*C & -\frac{d}{dt} - A^* \\ 0 & E_T \\ \frac{d}{dt} - A & -BQ^{-1}B^* \\ E_0 & 0 \end{pmatrix} \begin{pmatrix} \tilde{x} \\ \tilde{\lambda} \end{pmatrix} = \begin{pmatrix} C^*Cx_{\mathrm{d}} \\ 0 \\ Bu_{\mathrm{d}} + f \\ x_0 \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_T \\ \varepsilon_2 \\ \varepsilon_0 \end{pmatrix} \tag{2.7}$$

for perturbations $(\varepsilon_1, \varepsilon_2) \in L_1(0,T;X)^2$ and $(\varepsilon_0, \varepsilon_T) \in X^2$. The solution $(\tilde{x}, \tilde{\lambda})$ will be referred to as the *perturbed solution*. The terms ε_1 and ε_2 are perturbations of the dynamics which could be caused by discretization errors in time or space over the time interval $[0,T]$, whereas ε_0 and ε_T resemble perturbations from space discretization errors in the initial and terminal datum, respectively. The question we want to answer is the following: How does $(\varepsilon_0, \varepsilon_T)$ and the behavior of the perturbation of the dynamics ε_1 and ε_2 over time influence the temporal behavior of $\delta x := \tilde{x} - x$ and $\delta\lambda := \tilde{\lambda} - \lambda$? To answer this question, we subtract (2.6) from (2.7) and conclude by linearity

$$\begin{pmatrix} C^*C & -\frac{d}{dt} - A^* \\ 0 & E_T \\ \frac{d}{dt} - A & -BQ^{-1}B^* \\ E_0 & 0 \end{pmatrix} \begin{pmatrix} \delta x \\ \delta\lambda \end{pmatrix} = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_T \\ \varepsilon_2 \\ \varepsilon_0 \end{pmatrix} =: \varepsilon. \tag{2.8}$$

Directly from the solvability of the extremal equations, one would obtain the estimates

$$\|(\delta x, \delta\lambda)\|_{C(0,T;X)^2} \le c\|\varepsilon\|_{(L_1(0,T;X)\times X)^2},$$
$$\|(\delta x, \delta\lambda)\|_{L_p(0,T;X)^2} \le c\|\varepsilon\|_{(L_1(0,T;X)\times X)^2},$$

for any $1 \le p \le \infty$, meaning the absolute error will be small if the perturbation is small. However, there are two downsides of this estimate. First, we do not know how the constant $c \ge 0$ depends on T and second, motivated by the particular application to MPC, we would like $(\varepsilon_1, \varepsilon_2)$ to be increasing towards T, modeling grids that coarsen up exponentially. In that case, this estimate would yield no useful information. It turns out that the key towards deriving local-in-time estimates for the absolute error $(\delta x, \delta\lambda)$ is a scaling argument combined with T-independent bounds on the solution operator M^{-1}, as stated in the following theorem.

Theorem 2.27. *Let $(\delta x, \delta\lambda) \in C(0,T;X)^2$ solve (2.8), where $\varepsilon_1, \varepsilon_2 \in L_1(0,T;X)$ and $\varepsilon_0, \varepsilon_T \in X$. Moreover, let $\delta u = Q^{-1}B^*\delta\lambda$. Assume the solution operator's norms*

$$\begin{aligned} &\|M^{-1}\|_{L((L_1(0,T;X)\times X)^2, C(0,T;X)^2)}, \qquad \|M^{-1}\|_{L((L_2(0,T;X)\times X)^2, C(0,T;X)^2)}, \\ &\|M^{-1}\|_{L((L_1(0,T;X)\times X)^2, L_2(0,T;X)^2)}, \qquad \|M^{-1}\|_{L((L_2(0,T;X)\times X)^2, L_2(0,T;X)^2)} \end{aligned} \tag{2.9}$$

can be bounded independently of T. Then there is a scaling factor $\mu > 0$ satisfying

$$\mu < \frac{1}{\|M^{-1}\|_{L((L_2(0,T;X)\times X)^2, L_2(0,T;X)^2)}}$$

and a constant $c \geq 0$, both independent of T, such that, defining

$$\rho := \left\|e^{-\mu t}\varepsilon_1(t)\right\|_S + \left\|e^{-T}\varepsilon_T\right\| + \left\|e^{-\mu t}\varepsilon_2(t)\right\|_S + \|\varepsilon_0\|$$

for $S := L_1(0,T;X)$ or $S := L_2(0,T;X)$, we have

$$\begin{aligned} \left\|e^{-\mu t}\delta x(t)\right\|_{L_2(0,T;X)} + \left\|e^{-\mu t}\delta\lambda(t)\right\|_{L_2(0,T;X)} &\leq c\rho, \\ \left\|e^{-\mu t}\delta u(t)\right\|_{L_2(0,T;U)} &\leq c\rho \end{aligned} \tag{2.10}$$

and

$$\begin{aligned} \left\|e^{-\mu t}\delta x(t)\right\|_{C(0,T;X)} + \left\|e^{-\mu t}\delta\lambda(t)\right\|_{C(0,T;X)} &\leq c\rho, \\ \left\|e^{-\mu t}\delta u(t)\right\|_{L_\infty(0,T;U)} &\leq c\rho. \end{aligned} \tag{2.11}$$

Proof. For $\mu > 0$ arbitrary we introduce scaled variables $\widetilde{\delta x}(t) := e^{-\mu t}\delta x(t)$, $\widetilde{\delta\lambda}(t) := e^{-\mu t}\delta\lambda(t)$, $\widetilde{\varepsilon}_1(t) := e^{-\mu t}\varepsilon_1(t)$ and $\widetilde{\varepsilon}_2(t) := e^{-\mu t}\varepsilon_2(t)$ and apply Lemma 2.22 i). This yields

$$(2.8) \quad \iff \quad \left(\begin{pmatrix} C^*C & -\frac{d}{dt} - A^* \\ 0 & E_T \\ \frac{d}{dt} - A & -BQ^{-1}B^* \\ E_0 & 0 \end{pmatrix} + \mu \underbrace{\begin{pmatrix} 0 & -I \\ 0 & 0 \\ I & 0 \\ 0 & 0 \end{pmatrix}}_{=:P} \right) \begin{pmatrix} \widetilde{\delta x} \\ \widetilde{\delta\lambda} \end{pmatrix} = \begin{pmatrix} e^{-\mu t}\varepsilon_1 \\ e^{-T}\varepsilon_T \\ e^{-\mu t}\varepsilon_2 \\ \varepsilon_0 \end{pmatrix}.$$

Introducing $\widetilde{z} := \left(\widetilde{\delta x}, \widetilde{\delta\lambda}\right)$ and $\widetilde{\varepsilon} := (e^{-\mu t}\varepsilon_1, e^{-T}\varepsilon_T, e^{-\mu t}\varepsilon_2, \varepsilon_0)$, we compute

$$(M + \mu P)\widetilde{z} = \widetilde{\varepsilon} \quad \Rightarrow \quad (I + \mu P M^{-1})M\widetilde{z} = \widetilde{\varepsilon} \quad \Rightarrow \quad \widetilde{z} = M^{-1}(I + \mu P M^{-1})^{-1}\widetilde{\varepsilon}. \tag{2.12}$$

Next, we expand $(I+\mu PM^{-1})^{-1}$ into a Neumann series, cf. [85, Theorem 2.14]. In the following, denote $E = (L_2(0,T;X) \times X)^2$. While the previous computation is valid for all $\mu \in \mathbb{R}$, we now choose $\mu > 0$ small enough such that $\beta := \mu\|M^{-1}\|_{L(E,L_2(0,T;X)^2)} < 1$. By assumption, $\|M^{-1}\|_{L(E,L_2(0,T;X)^2)}$ is bounded independently of T, so we can choose $\mu > 0$ independently of T. Since $\|P\|_{L(L_2(0,T;X)^2,E)} \leq 1$, it follows that $\|\mu PM^{-1}\|_{L(E,E)} \leq \beta < 1$. A Neumann series expansion of $(I + \mu PM^{-1})^{-1}$ yields

$$\|(I + \mu PM^{-1})^{-1}\|_{L(E)} \leq \sum_{i=0}^{\infty} \|((\mu PM^{-1})^i\|_{L(E)} \leq \sum_{i=0}^{\infty} \beta^i = \frac{1}{1-\beta}. \tag{2.13}$$

Hence, we obtain the desired L_2-estimate on state and adjoint for the case $S = L_2(0,T;X)$:

$$\|\tilde{z}\|^2_{L_2(0,T;X)} \leq \|M^{-1}\|_{L(E,L_2(0,T;X)^2)}\|(I + \mu PM^{-1})^{-1}\|_{L(E,E)}\|\tilde{\varepsilon}\|_E.$$

To prove the remaining pointwise estimates or the case $S = L_1(0,T;X)$, we use the following alternative representation, which can be verified by premultiplication with $(I+\mu PM^{-1})$:

$$(I+\mu PM^{-1})^{-1} = I - (I+\mu PM^{-1})^{-1}\mu PM^{-1}.$$

Now, we estimate the operator norm for $Z = C(0,T;X)^2$ or $Z = L_2(0,T;X)^2$ via

$$\begin{aligned}
&\|(M^{-1}(I+\mu PM^{-1})^{-1}\|_{L((S\times X)^2,Z)} = \|(M^{-1} - M^{-1}(I+\mu PM^{-1})^{-1}\mu PM^{-1}\|_{L((S\times X)^2,Z)} \\
&\leq \|M^{-1}\|_{L((S\times X)^2,Z)} + \|M^{-1}\|_{L(E,Z)}\|((I+\mu PM^{-1})^{-1}\|_{L(E,E)}\|\mu PM^{-1}\|_{L((S\times X)^2,E)} \\
&\leq \|M^{-1}\|_{L((S\times X)^2,Z)} + \frac{\mu\|M^{-1}\|_{L(E,Z)}\|M^{-1}\|_{L((S\times X)^2,L_2(0,T;X)^2)}}{1-\beta}
\end{aligned}$$

using $\|P\|_{L(L_2(0,T;X)^2,E)} \leq 1$ and (2.13). Thus, by (2.12) it follows with taking norms that

$$\|\widetilde{z}\|_Z \leq \left(\|M^{-1}\|_{L((S\times X)^2,Z)} + \frac{\mu\|M^{-1}\|_{L(E,Z)}\|M^{-1}\|_{L((S\times X)^2,L_2(0,T;X)^2)}}{1-\beta}\right)\|\widetilde{\varepsilon}\|_{S^2}, \tag{2.14}$$

for $Z = C(0,T;X)^2$ or $Z = L_2(0,T;X)^2$. Using the assumption on the T-independent bound on the operator norms and the definition $\rho := \|\widetilde{\varepsilon}\|_{(S\times X)^2}$, we obtain the result for the state and the adjoint by going back to the original variables via $\widetilde{z} = (e^{-\mu t}\delta x, e^{-\mu t}\delta\lambda)$. For the control, we conclude

$$\begin{aligned}
\|e^{-\mu t}\delta u(t)\|_{L_2(0,T;U)} = \|e^{-\mu t}Q^{-1}B^*\delta\lambda\|_{L_2(0,T;U)} \leq \|Q^{-1}B^*\|_{L(X,U)}\left\|\widetilde{\delta\lambda}\right\|_{L_2(0,T;X)} \leq c\rho, \\
\|e^{-\mu t}\delta u(t)\|_U = \|e^{-\mu t}Q^{-1}B^*\delta\lambda(t)\|_U \leq \|Q^{-1}B^*\|_{L(X,U)}\left\|\widetilde{\delta\lambda}(t)\right\| \leq c\rho
\end{aligned} \tag{2.15}$$

for a.e. $t\in[0,T]$, where we used the bound on $\|e^{-\mu t}\delta\lambda\|_{L_2(0,T;X)}$ and $\|e^{-\mu t}\delta\lambda\|_{C(0,T;X)}$ and the fact that B and Q are local in time. This yields (2.10) and (2.11). □

Remark 2.28. *We will briefly comment on the Neumann series occurring in* (2.13)*. The operator $(I+\mu PM^{-1})^{-1}$ can be represented by its Neumann series, i.e., $(I+\mu PM^{-1})^{-1} = \sum_{i=0}^{\infty}(-\mu PM^{-1})^k$, see [85, Theorem 2.14]. We provide an illustration for the summand for $k=2$, i.e., $(\mu PM^{-1})^2 = \mu PM^{-1}\mu PM^{-1}$. The application of this operator can be interpreted as the following. M^{-1} solves the corresponding Cauchy problems with right hand side including initial and terminal condition. Afterwards, the operator μP maps the solutions to source terms scaled by μ, i.e., $\mu P(\delta x,\delta\lambda) = (-\mu\delta\lambda, 0, \mu\delta x, 0)$. This right hand side then enters M^{-1} again, the Cauchy problems are solved with zero initial data and source terms $-\mu\delta\lambda$ and $\mu\delta x$ and the process is repeated.*

The crucial assumption in Theorem 2.27 is that the operator norms in (2.9) can be bounded independently of T. This ensures that the scaling factor and the constants in the upper bound do not deteriorate for $T\to\infty$. In Section 2.2.3, we will show that these T-independent bounds can be derived if the dynamics are exponentially stabilizable and detectable. It will turn out that all bounds in (2.9) can be shown simultaneously. Yet before, we will present another scaling theorem which, again under T-independent boundedness of M^{-1}, provides an exponential turnpike property. The approach will be very closely connected to the proof of Theorem 2.27.

2.2.2 An exponential turnpike result

In this section, we modify the scaling approach employed in Theorem 2.27 to deduce a turnpike result for the optimal solution of Problem 2.23 and the corresponding adjoint state. In the case of optimal control problems governed by general evolution equations in Hilbert spaces, turnpike theorems were given for dynamics governed by a strongly continuous semigroup with bounded control and observation operators as well as for boundary controlled parabolic equations in [135]. We give an alternative proof for the case of bounded control and observation that can be generalized to unbounded control or observation, see Section 2.3, and a terminal condition on the state, see Section 2.4. First, we introduce the steady state optimization problem corresponding to Problem 2.23. To this end, we assume that x_d, u_d and f are independent of time, i.e., $x_d \equiv \bar{x}_d \in X$, $u_d \equiv \bar{u}_d \in U$ and $f \equiv \bar{f} \in X$. The steady state control problem then reads

$$\begin{aligned} \min_{\bar{x},\bar{u}} \ & \frac{1}{2}\|C(\bar{x} - \bar{x}_\mathrm{d})\|_Y^2 + \frac{1}{2}\|R(\bar{u} - \bar{u}_\mathrm{d})\|_U^2 \\ \text{s.t.} \ & -A\bar{x} - B\bar{u} = \bar{f}. \end{aligned} \tag{2.16}$$

We tacitly assume that A is continuously invertible to ensure the existence of a control-to-state map. In the case of parabolic PDEs where A is a differential operator of second order, this could, e.g., be achieved by assuming coercivity of the bilinear form induced by A via the Lax-Milgram lemma. By coercivity of R, cf. Problem 2.23 iv), the problem is convex and $(\bar{x}, \bar{\lambda})$ solves the corresponding necessary and sufficient first-order conditions

$$\begin{pmatrix} C^*C & -A^* \\ -A & -BQ^{-1}B^* \end{pmatrix} \begin{pmatrix} \bar{x} \\ \bar{\lambda} \end{pmatrix} = \begin{pmatrix} C^*C\bar{x}_\mathrm{d} \\ B\bar{u}_\mathrm{d} + \bar{f} \end{pmatrix}, \tag{2.17}$$

where $Q = R^*R$ and where we eliminated the control via $\bar{u} = Q^{-1}B^*\bar{\lambda} + \bar{u}_\mathrm{d}$.

Lemma 2.29. *Let (x, u, λ) solve Problem 2.23 with $f = 0$. Moreover, let $(\bar{x}, \bar{u}, \bar{\lambda})$ solve the corresponding steady state problem (2.16). Then $(\delta x, \delta\lambda) := (x - \bar{x}, \lambda - \bar{\lambda})$ solves*

$$\begin{pmatrix} C^*C & -\frac{d}{dt} - A^* \\ 0 & E_T \\ \frac{d}{dt} - A & -BQ^{-1}B^* \\ E_0 & 0 \end{pmatrix} \begin{pmatrix} \delta x \\ \delta\lambda \end{pmatrix} = \begin{pmatrix} 0 \\ -\bar{\lambda} \\ 0 \\ x_0 - \bar{x} \end{pmatrix}, \tag{2.18}$$

where $\delta u := u - \bar{u} = Q^{-1}B^*\delta\lambda$ and E_t for $t \in [0, T]$ is defined in Definition 2.26.

Proof. Using (2.17) and $\frac{d}{dt}\bar{x} = \frac{d}{dt}\bar{\lambda} = 0$ yields that $(\bar{x}, \bar{\lambda})$ satisfies

$$\begin{pmatrix} C^*C & -\frac{d}{dt} - A^* \\ 0 & E_T \\ \frac{d}{dt} - A & -BQ^{-1}B^* \\ E_0 & 0 \end{pmatrix} \begin{pmatrix} \bar{x} \\ \bar{\lambda} \end{pmatrix} = \begin{pmatrix} C^*C\bar{x}_\mathrm{d} \\ \bar{\lambda} \\ B\bar{u}_\mathrm{d} + \bar{f} \\ \bar{x} \end{pmatrix}. \tag{2.19}$$

We conclude the result by subtracting (2.19) from (2.6). □

Theorem 2.30. *Let (x, u, λ) solve Problem 2.23 and let $(\bar{x}, \bar{u}, \bar{\lambda})$ solve the corresponding steady state problem (2.16). Assume the solution operator's norms*

$$\|M^{-1}\|_{L((L_2(0,T;X)\times X)^2, C(0,T;X)^2)} \quad \textit{and} \quad \|M^{-1}\|_{L((L_2(0,T;X)\times X)^2, L_2(0,T;X)^2)} \tag{2.20}$$

can be bounded independently of T. Then, defining $(\delta x, \delta u, \delta\lambda) := (x - \bar{x}, u - \bar{u}, \lambda - \bar{\lambda})$, there exist a scaling factor $\mu > 0$ satisfying

$$\mu < \frac{1}{\|M^{-1}\|_{L((L_2(0,T;X)\times X)^2, L_2(0,T;X)^2)}}$$

and a constant $c \geq 0$, both independent of T, such that

$$\begin{aligned} \left\| \frac{1}{e^{-\mu t} + e^{-\mu(T-t)}} \delta x(t) \right\|_{L_2(0,T;X)} &+ \left\| \frac{1}{e^{-\mu t} + e^{-\mu(T-t)}} \delta u(t) \right\|_{L_2(0,T;U)} \\ &+ \left\| \frac{1}{e^{-\mu t} + e^{-\mu(T-t)}} \delta \lambda(t) \right\|_{L_2(0,T;X)} \leq c \left(\|x_0 - \bar{x}\| + \|\bar{\lambda}\| \right), \end{aligned} \tag{2.21}$$

$$\begin{aligned} \left\| \frac{1}{e^{-\mu t} + e^{-\mu(T-t)}} \delta x(t) \right\|_{C(0,T;X)} &+ \left\| \frac{1}{e^{-\mu t} + e^{-\mu(T-t)}} \delta u(t) \right\|_{L_\infty(0,T;U)} \\ &+ \left\| \frac{1}{e^{-\mu t} e^{-\mu(T-t)}} \delta \lambda(t) \right\|_{C(0,T;X)} \leq c \left(\|x_0 - \bar{x}\| + \|\bar{\lambda}\| \right). \end{aligned} \tag{2.22}$$

Proof. We proceed similarly to the proof of Theorem 2.27 and introduce a scaling factor $0 < \mu < \frac{1}{\|M^{-1}\|_{L((L_2(0,T;X)\times X)^2, L_2(0,T;X)^2)}}$ and scaled variables $\widetilde{\delta x} := \frac{1}{e^{-\mu t} + e^{-\mu(T-t)}} \delta x$ and $\widetilde{\delta \lambda} := \frac{1}{e^{-\mu t} + e^{-\mu(T-t)}} \delta \lambda$, and compute

$$(2.18) \quad \iff \quad \left(\underbrace{\begin{pmatrix} C^*C & -\frac{d}{dt} - A^* \\ 0 & E_T \\ \frac{d}{dt} - A & -BQ^{-1}B^* \\ E_0 & 0 \end{pmatrix}}_{=M} + \mu \underbrace{\begin{pmatrix} 0 & F \\ 0 & 0 \\ -F & 0 \\ 0 & 0 \end{pmatrix}}_{=:P} \right) \begin{pmatrix} \widetilde{\delta x} \\ \widetilde{\delta \lambda} \end{pmatrix} = \frac{1}{1 + e^{-\mu T}} \begin{pmatrix} 0 \\ -\bar{\lambda} \\ 0 \\ x_0 - \bar{x} \end{pmatrix}$$

where $F := \frac{e^{-\mu(T-t)} - e^{-\mu t}}{e^{-\mu t} + e^{-\mu(T-t)}}$ and the factor $\frac{1}{1+e^{-\mu T}}$ arises due to the scaling of the initial values. The proof for the estimate of the state and the adjoint in (2.21) and (2.22) is analogous to the one of Theorem 2.27. Defining $\tilde{z} := (\widetilde{\delta x}, \widetilde{\delta \lambda})$ and $\tilde{r} := \frac{1}{1+e^{-\mu T}}(0, -\bar{\lambda}, 0, x_0 - \bar{x})$, we get

$$(M + \mu P)\tilde{z} = \tilde{r} \quad \Rightarrow \quad (I + \mu P M^{-1}) M \tilde{z} = \tilde{r} \quad \Rightarrow \quad \tilde{z} = M^{-1}(I + \mu P M^{-1})^{-1} \tilde{r}. \tag{2.23}$$

We observe that $\|F\|_{L(L_2(0,T;X), L_2(0,T;X))} \leq 1$ and thus $\|P\|_{L(L_2(0,T;X)^2, (L_2(0,T;X)\times X)^2)} \leq 1$. Thus, as in the proof of Theorem 2.27, by a standard Neumann series argument, cf. [85, Theorem 2.14] and by choosing μ such that $\beta := \mu\|M^{-1}\|_{L((L_2(0,T;X)\times X)^2, L_2(0,T;X)^2)} < 1$, the mapping

$(I + \mu P M^{-1})$ is continuously invertible. Therefore, using the Neumann series representation of $(I + \mu P M^{-1})^{-1}$ yields that

$$\begin{aligned} &\|(I + \mu P M^{-1})^{-1}\|_{L((L_2(0,T;X)\times X)^2,(L_2(0,T;X)\times X)^2)} \\ &\qquad \le \sum_{i=0}^{\infty} \|(\mu P M^{-1})^i\|_{L((L_2(0,T;X)\times X)^2,(L_2(0,T;X)\times X)^2)} \le \sum_{i=0}^{\infty} \beta^i = \frac{1}{1-\beta}. \end{aligned} \tag{2.24}$$

Hence, we conclude with (2.23) and (2.24)

$$\begin{aligned} \|\tilde{z}\|_{L_2(0,T;X)^2} &\le \frac{\|M^{-1}\|_{L((L_2(0,T;X)\times X)^2,L_2(0,T;X)^2)}}{1-\beta} \|\tilde{r}\|_{(L_2(0,T;X)\times X)^2}, \\ \|\tilde{z}\|_{C(0,T;X)^2} &\le \frac{\|M^{-1}\|_{L((L_2(0,T;X)\times X)^2,C(0,T;X)^2)}}{1-\beta} \|\tilde{r}\|_{(L_2(0,T;X)\times X)^2}. \end{aligned} \tag{2.25}$$

Finally, $\|\tilde{r}\|_{(L_2(0,T;X)\times X)^2} \le \|x_0 - \bar{x}\| + \|\bar{\lambda}\|$ and going back to the original variables yields the result for the state and adjoint. We set $\widetilde{\delta u} := \frac{1}{e^{-\mu t}+e^{-\mu(T-t)}}\delta u$ and from boundedness of B, we compute

$$\begin{aligned} \|\widetilde{\delta u}\|_{L_2(0,T;U)} &= \left\| \frac{1}{e^{-\mu t} + e^{-\mu(T-t)}} Q^{-1} B^* \delta\lambda \right\|_{L_2(0,T;U)} \\ &\le \|Q^{-1}B^*\|_{L(X,U)} \|\widetilde{\delta\lambda}\|_{L_2(0,T;X)} \le c\|\tilde{r}\|_{(L_2(0,T;X)\times X)^2}, \\ \|\widetilde{\delta u}(t)\|_U &= \left\| \frac{1}{e^{-\mu t} + e^{-\mu(T-t)}} Q^{-1} B^* \widetilde{\delta\lambda}(t) \right\| \\ &\le \|Q^{-1}B^*\|_{L(X,U)} \|\widetilde{\delta\lambda}(t)\| \le c\|\tilde{r}\|_{(L_2(0,T;X)\times X)^2} \qquad \text{for a.e } t \in [0,T], \end{aligned} \tag{2.26}$$

which completes the estimate (2.22). □

We give a short interpretation of the two estimates given in Theorem 2.30. For the first inequality (2.21) consider a fixed $\varepsilon \in (0, \frac{1}{2})$. For $t \in [\varepsilon T, (1-\varepsilon)T]$ and if $T \to \infty$, the two scaling terms $e^{-\mu t}$ and $e^{-\mu(T-t)}$ approach zero exponentially fast and we estimate

$$\frac{1}{e^{-\mu t} + e^{-\mu(T-t)}} \ge \frac{1}{e^{-\mu\varepsilon T} + e^{-\mu(T-(1-\varepsilon)T)}} = \frac{1}{2e^{-\mu\varepsilon T}}.$$

Hence, e.g., for the difference of state x and its turnpike $\bar{x}$,

$$\int_0^T \left\| \frac{x(t) - \bar{x}}{e^{-\mu t} + e^{-\mu(T-t)}} \right\|^2 dt \ge \int_{\varepsilon T}^{(1-\varepsilon)T} \left\| \frac{x(t) - \bar{x}}{e^{-\mu t} + e^{-\mu(T-t)}} \right\|^2 dt \ge \frac{1}{4e^{-2\mu\varepsilon T}} \int_{\varepsilon T}^{(1-\varepsilon)T} \|x(t) - \bar{x}\|^2 \, dt$$

which, using (2.21) of Theorem 2.30, implies that

$$\int_{\varepsilon T}^{(1-\varepsilon)T} \|x(t) - \bar{x}\|^2 \, dt \le c e^{-2\mu\varepsilon T} \left(\|x_0 - \bar{x}\| + \|\bar{\lambda}\| \right)^2.$$

Proceeding analogously for the adjoint and the control, we get

$$\|x-\bar{x}\|_{L_2(\varepsilon T,(1-\varepsilon)T;X)}, \|u-\bar{u}\|_{L_2(\varepsilon T,(1-\varepsilon)T;U)}, \left\|\lambda-\bar{\lambda}\right\|_{L_2(\varepsilon T,(1-\varepsilon)T;X)} \to 0 \qquad \text{if } T\to\infty$$

i.e., $L_2(0,T;X)$-convergence on a part $I_\varepsilon = [\varepsilon T,(1-\varepsilon)T]$ of the whole time interval [0,T]. The convergence rate is exponential and the size of I_ε grows linearly in T, as its length is $(1-2\varepsilon)T$. Hence, the share of I_ε of the whole interval is constant because $(1-2\varepsilon)T/T = 1-2\varepsilon$. Thus, on a fixed percentage of the interval $[0,T]$, the state, the control and the adjoint converge to the turnpike in the L_2 norm, as the horizon T goes to infinity.

For the second inequality, i.e., (2.22), rewriting the pointwise estimate, we have

$$\|x(t)-\bar{x}\| + \left\|\lambda(t)-\bar{\lambda}\right\| \le c(e^{-\mu t}+e^{-\mu(T-t)})$$

for every $t\in[0,T]$. Therefore, if we fix $\varepsilon\in(0,\frac{1}{2})$ and take the maximum over all $t\in[\varepsilon T,(1-\varepsilon)T]$, we get

$$\|x-\bar{x}\|_{C(\varepsilon T,(1-\varepsilon)T;X)} + \left\|\lambda-\bar{\lambda}\right\|_{C(\varepsilon T,(1-\varepsilon)T;X)} \le 2ce^{-\mu\varepsilon T}.$$

The right hand side approaches zero exponentially fast as $T\to\infty$. Hence, for each $\varepsilon\in(0,\frac{1}{2})$, we obtain uniform exponential convergence on the interval $[\varepsilon T,(1-\varepsilon)T]$ of x and λ to $\bar{x}$ and $\bar{\lambda}$, respectively, as $T\to\infty$. Again, as $(1-2\varepsilon)T/T=(1-2\varepsilon)$, we conclude that on a fixed fraction of the whole interval $[0,T]$, the state and adjoint converge to the turnpike in the maximum norm.

Under the assumption of T-independent bounds on various solution operator's norms, we have deduced two results: First, we proved that perturbations of the right hand side stay local in time, cf. Theorem 2.27 and second, we obtained a turnpike result, cf. Theorem 2.30. In the following, we will show that these T-independent bounds indeed hold, provided that the dynamics are stabilizable and detectable.

2.2.3 T-independent bounds for the solution operator

In this section, we will derive T-independent bounds on the norm of the solution operator M^{-1}, which is a central assumption in the abstract scaling result of Theorem 2.27 and Theorem 2.30. Since $[0,T]$ is bounded, we have the continuous embeddings

$$C(0,T;X)\hookrightarrow L_2(0,T;X)\hookrightarrow L_1(0,T;X).$$

Hence, we may denote

$$\|v\|_{1\vee 2} := \min\{\|v\|_{L_1(0,T;X)}, \|v\|_{L_2(0,T,X)}\}$$

for $v\in L_1(0,T;X)$ (setting $\|v\|_{L_2(0,T;X)}=\infty$ if $v\notin L_2(0,T;X)$), where

$$\min\left\{1,\frac{1}{\sqrt{T}}\right\}\|v\|_{L_1(0,T;X)} \le \|v\|_{1\vee 2} \le \|v\|_{L_1(0,T;X)}.$$

Likewise, we write

$$\|v\|_{2\wedge\infty} := \max\{\|v\|_{L_2(0,T;X)}, \|v\|_{C(0,T;X)}\},$$

for $v \in C(0,T;X)$ satisfying

$$\|v\|_{C(0,T;X)} \le \|v\|_{2\wedge\infty} \le \max\left\{1, \sqrt{T}\right\} \|v\|_{C(0,T;X)}.$$

We note that $\|\cdot\|_{2\wedge\infty}$ induces an equivalent norm on $C(0,T;X)$ where the constants in the equivalence of norms deteriorate for $T \to \infty$. This is not the case for $\|\cdot\|_{1\vee 2}$, which does not satisfy a triangle inequality. For brevity of notation, for $z := (v_1, v_2) \in L_1(0,T;X)^2$, we will write $\|z\|_{1\vee 2}^2 := \|v_1\|_{1\vee 2}^2 + \|v_2\|_{1\vee 2}^2$. Similarly for $z = (v_1, v_2) \in C(0,T;X)^2$ we abbreviate $\|z\|_{2\wedge\infty}^2 := \|v_1\|_{2\wedge\infty}^2 + \|v_1\|_{2\wedge\infty}^2$. In the following, $c > 0$ denotes a generic constant and will be renamed accordingly over the course of a proof. It is very important, however, that the constants in the proofs will never depend on the horizon T. Also, we tacitly use equivalence of norms in $\mathbb{R}^2$: $\max\{|a|, |b|\} \approx \sqrt{|a|^2 + |b|^2} \approx |a| + |b|$ for all $a, b \in \mathbb{R}^2$.

We first present a Hölder-like inequality for the notions introduced above.

Lemma 2.31. *Let $v \in C(0,T;X)$ and $w \in L_1(0,T;X)$. Then,*

$$\int_0^T \langle v(s), w(s)\rangle \, ds \le \|v\|_{2\wedge\infty}\|w\|_{1\vee 2}.$$

Proof.

$$\begin{aligned}\int_0^T \langle v(s), w(s)\rangle \, ds &\le \min\{\|v\|_{C(0,T;X)}\|w\|_{L_1(0,T;X)}, \|v\|_{L_2(0,T;X)}\|w\|_{L_2(0,T;X)}\} \\ &\le \min\{\|v\|_{2\wedge\infty}\|w\|_{L_1(0,T;X)}, \|v\|_{2\wedge\infty}\|w\|_{L_2(0,T;X)}\} \le \|v\|_{2\wedge\infty}\|w\|_{1\vee 2}.\end{aligned}$$

□

The main result of this section will be a T-independent bound for

$$\|M^{-1}\|_{L(((L_1(0,T;X),\|\cdot\|_{1\vee 2})\times X)^2, (C(0,T;X),\|\cdot\|_{2\wedge\infty})^2)}.$$

This implies all desired T-independent bounds required by Theorem 2.27 and Theorem 2.30. To this end, consider the mild solution (x, λ) of the system

$$\begin{pmatrix} C^*C & -\frac{d}{dt} - A^* \\ 0 & E_T \\ \frac{d}{dt} - A & -BQ^{-1}B^* \\ E_0 & 0 \end{pmatrix} \begin{pmatrix} x \\ \lambda \end{pmatrix} = \begin{pmatrix} l_1 \\ \lambda_T \\ l_2 \\ x_0 \end{pmatrix} \tag{2.27}$$

in $[0,T]$, where $l_1, l_2 \in L_1(0,T;X)$ and $x_0, \lambda_T \in X$ are given. Again, we abbreviate $z = M^{-1}r$ with $z = (x, \lambda)$ and $r = (l_1, \lambda_T, l_2, x_0)$.

Next, we introduce the main assumption that ensures a T-independent bound. Recall from Definition 2.12 that a strongly continuous semigroup $(\mathcal{T}(t))_{t\ge 0}$ is called exponentially stable if there exist $M, \mu > 0$ such that $\|\mathcal{T}(t)\|_{L(X)} \le Me^{-\mu t}$ for all $t > 0$.

Assumption 2.32.

i) (A, C) *is exponentially detectable, i.e., there exists a feedback operator* $K_C \in L(Y, X)$ *such that the semigroup generated by* $A^* + C^* K_C^*$ *is exponentially stable.*

ii) (A, B) *is exponentially stabilizable, i.e., there exists a feedback operator* $K_B \in L(X, U)$ *such that the strongly continuous semigroup generated by* $A + BK_B$ *is exponentially stable.*

By the norm identity of an operator and its adjoint, i.e., for a semigroup $(\mathcal{T}(t))_{t\geq 0}$, $\|\mathcal{T}^*(t)\| = \|\mathcal{T}(t)\|$, detectability can equivalently be formulated by exponential stability of the semigroup generated by $A + K_C C$.

The assumption above relates to Remark 2.14, where a T-independent bound for the solution operator of a Cauchy problem can be shown if (and only if) the underlying semigroup is exponentially stable. In the case of optimal control, we are able to replace exponential stability by mere stabilizability and detectability. This weaker assumption allows to include unstable or conservative systems, e.g., the undamped wave equation, where the corresponding (uncontrolled) semigroup is a group of isometries, i.e., $\|\mathcal{T}(t)\| = 1$ for all $t \in \mathbb{R}$.

Using the stabilizability and detectability assumption, we first define suitable exponentially stable test functions. The approach in Lemma 2.33 and Lemma 2.34 is inspired by the stability estimates in [135, Lemma 2] and [113, Lemma 3.5].

Lemma 2.33. *Consider* $x \in C(0, T; X)$, $t \in [0, T]$ *and let* $\varphi \in C(0, t; X)$ *solve*

$$\begin{aligned} -\varphi' &= (A^* + C^* K_C^*)\varphi \qquad \text{in } [0, t], \\ \varphi(t) &= x(t), \end{aligned} \tag{2.28}$$

where K_C^* *is a stabilizing feedback for* (A^*, C^*)*. Then, there are constants* $M_\varphi, k_\varphi > 0$ *such that for test functions* $v \in L_2(0, t; X)$*:*

$$\int_0^t |\langle v(s), \varphi(s)\rangle| \, ds \leq \|x(t)\| \frac{M_\varphi}{\sqrt{k_\varphi}} \sqrt{\int_0^t \|v(s)\|^2 e^{-k_\varphi (t-s)} \, ds}. \tag{2.29}$$

Additionally, consider $\lambda \in C(0, T; X)$, $t \in [0, T]$ *and let* $\psi \in C(t, T; X)$ *solve*

$$\begin{aligned} \psi' &= (A + BK_B)\psi \qquad \text{in } [t, T], \\ \psi(t) &= \lambda(t), \end{aligned} \tag{2.30}$$

where K_B *is a stabilizing feedback for* (A, B)*. Then, there are constants* $M_\psi, k_\psi > 0$ *such that for test functions* $v \in L_2(t, T; X)$*:*

$$\int_t^T |\langle v(s), \psi(s)\rangle| \, ds \leq \|\lambda(t)\| \frac{M_\psi}{\sqrt{k_\psi}} \sqrt{\int_t^T \|v(s)\|^2 e^{-k_\psi (s-t)} \, ds}. \tag{2.31}$$

Proof. We will first prove (2.29). By exponential stability of the strongly continuous semigroup, there exist $M_\varphi, k_\varphi > 0$, such that

$$\|\varphi(s)\| \leq M_\varphi e^{-k_\varphi (t-s)} \|x(t)\| \qquad \forall\, 0 \leq s \leq t.$$

Using this exponential stability, we get

$$\int_0^t |\langle v(s), \varphi(s)\rangle| \, ds \leq \int_0^t \|v(s)\| \|\varphi(s)\| \, ds \leq \|x(t)\| \int_0^t \|v(s)\| M_\varphi e^{-k_\varphi(t-s)} \, ds.$$

For $v \in L_2(0,t;X)$, the integral term can be estimated via:

$$\begin{aligned} \int_0^t \|v(s)\| M_\varphi e^{-k_\varphi(t-s)} \, ds &= \int_0^t \|v(s)\| M_\varphi e^{-\frac{k_\varphi}{2}(t-s)} \cdot e^{-\frac{k_\varphi}{2}(t-s)} \, ds \\ &\leq \sqrt{\int_0^t \|v(s)\|^2 M_\varphi^2 e^{-k_\varphi(t-s)} \, ds} \cdot \underbrace{\sqrt{\int_0^t e^{-k_\varphi(t-s)} \, ds}}_{< \frac{1}{\sqrt{k_\varphi}}}. \end{aligned}$$

The estimate (2.31) follows analogously. □

By using φ and ψ from (2.28) and (2.30) as test functions for (2.27), respectively, we obtain the following pointwise-in-time identities:

Lemma 2.34. *Let (x, λ) solve (2.27). If φ solves (2.28), then*

$$\|x(t)\|^2 = \int_0^t -\langle \varphi(s), K_C C x(s)\rangle + \langle R^{-*}B^*\varphi(s), R^{-*}B^*\lambda(s)\rangle_U + \langle \varphi(s), l_2(s)\rangle \, ds + \langle x_0, \varphi(0)\rangle \quad (2.32)$$

for all $0 \leq t \leq T$. If ψ solves (2.30), then

$$\|\lambda(t)\|^2 = \int_t^T -\langle K_B^* B^* \lambda(s), \psi(s)\rangle - \langle Cx(s), C\psi(s)\rangle_Y + \langle l_1(s), \psi(s)\rangle \, ds + \langle \psi(T), \lambda_T\rangle \quad (2.33)$$

for all $0 \leq t \leq T$.

Proof. We begin with the proof of (2.32). Testing the state equation with φ solving (2.28), integration over $[0,t]$ and integration by parts in the sense of Lemma 2.22 ii) on $[0,t]$ with $x_1 = x$, $f_1 = BQ^{-1}B^*\lambda + l_2$, $x_2 = \varphi$, and $f_2 = C^* K_C^* \varphi$ yields

$$\langle x(t), \varphi(t)\rangle - \langle x(0), \varphi(0)\rangle = \int_0^t \langle \varphi(s), BQ^{-1}B^*\lambda(s) + l_2(s)\rangle - \langle C^* K_C^* \varphi(s), x(s)\rangle \, ds.$$

Rearranging the terms, using the terminal condition $\varphi(t) = x(t)$ and $Q^{-1} = (R^*R)^{-1} = R^{-1}R^{-*}$, we get

$$\|x(t)\|^2 = \int_0^t -\langle \varphi(s), K_C C x(s)\rangle + \langle R^{-*}B^*\varphi(s), R^{-*}B^*\lambda(s)\rangle_U + \langle \varphi(s), l_2(s)\rangle \, ds + \langle x_0, \varphi(0)\rangle.$$

Formula (2.33) follows analogously by testing the adjoint equation with ψ solving (2.30). □

Based on (2.32) and (2.33), we will derive norm estimates for M^{-1} as a mapping into $L_2(0,T;X)^2$ and $C(0,T;X)^2$. While the latter turns out to be rather straightforward, the L_2-estimate requires integrating (2.32) and (2.33) over $[0,T]$. The crucial observation is that the integrals on the right hand side of (2.32) and (2.33) can be converted into convolutions with exponentially decaying functions. This will allow us to derive an L_2-estimate without any constants depending on the time T with the help of the following general lemma:

Lemma 2.35. *For $w \in L_1(0,T;(0,\infty))$, consider*

$$h_1(t) := \int_0^t w(s)e^{-k_\varphi(t-s)}\,ds, \text{ where } k_\varphi > 0,$$
$$h_2(t) := \int_t^T w(s)e^{-k_\psi(s-t)}\,ds, \text{ where } k_\psi > 0.$$

Then, there is a constant $c \geq 0$ independent of T, such that

$$\|h_i\|_{L_p(0,T)} \leq c\|w\|_{L_1(0,T)} \quad \text{for } i = 1,2 \text{ and } 1 \leq p \leq \infty.$$

Proof. Extending w by 0 from $[0,T]$ to $\mathbb{R}$ and defining $g_1(\tau) := e^{-k_\varphi \tau}$ for $\tau \geq 0$ and $g_1(\tau) = 0$ otherwise, we can write h_1 as the convolution

$$h_1(t) = (g_1 * w)(t) = \int_{\mathbb{R}} g_1(t-s)w(s)\,ds$$

and apply Young's inequality, cf. [147, Theorem II.4.4] to obtain

$$\|h_1\|_{L_p(0,T)} = \|g_1 * w\|_{L_p(\mathbb{R})} \leq \|g_1\|_{L_p(\mathbb{R})}\|w\|_{L_1(\mathbb{R})} \leq c\|w\|_{L_1(0,T)}$$

from $\|g_1\|_{L_p(\mathbb{R})} = \|e^{-k_\varphi t}\|_{L_p(\mathbb{R}_+)} \leq c(k_\varphi)$. For h_2, the estimate follows in the same way, setting $g_2(\tau) = e^{k_\psi \tau}$ for $\tau \leq 0$ and 0 otherwise. □

Using these convolution estimates, we can conclude:

Lemma 2.36. *Let Assumption 2.32 hold and let (x,λ) solve (2.27). Then there exists a constant $c \geq 0$ independent of T, such that*

$$\|x\|^2_{2\wedge\infty} + \|\lambda\|^2_{2\wedge\infty} \leq c\left(\|Cx\|^2_{L_2(0,T;Y)} + \|R^{-*}B^*\lambda\|^2_{L_2(0,T;U)} + \|r\|^2_{1\vee 2}\right), \tag{2.34}$$

where $r = (l_1, \lambda_T, l_2, y_0)$ and $\|r\|^2_{1\vee 2} := \|l_1\|^2_{1\vee 2} + \|\lambda_T\|^2 + \|l_2\|^2_{1\vee 2} + \|x_0\|^2$.

Proof. Our first step will be to derive an estimate for $\|x(t)\|$ from (2.32). By Lemma 2.33, we estimate the terms occurring in (2.32) as follows:

$$\int_0^t |\langle K_C Cx(s), \varphi(s)\rangle| ds \leq \|x(t)\| \frac{M_\varphi \|K_C\|_{L(Y,X)}}{\sqrt{k_\varphi}} \sqrt{\int_0^t \|Cx(s)\|_Y^2 e^{-k_\varphi(t-s)} ds}, \quad (2.35)$$

$$\int_0^t |\langle R^{-*}B^*\lambda(s), R^{-*}B^*\varphi(s)\rangle_U|\, ds \leq \|x(t)\| \frac{M_\varphi \|BR^{-1}\|_{L(U,X)}}{\sqrt{k_\varphi}} \sqrt{\int_0^t \|R^{-*}B^*\lambda(s)\|_U^2 e^{-k_\varphi(t-s)}\, ds}, \quad (2.36)$$

$$\int_0^t \langle l_2(s), \varphi(s)\rangle\, ds \leq \|x(t)\| \frac{M_\varphi}{\sqrt{k_\varphi}} \sqrt{\int_0^t \|l_2(s)\|^2 e^{-k_\varphi(t-s)}\, ds}, \quad (2.37)$$

or alternatively that

$$\int_0^t \langle l_2(s), \varphi(s)\rangle\, ds \leq \int_0^t \|l_2(s)\| \|\varphi(s)\|\, ds \leq \|x(t)\| M_\varphi \int_0^t \|l_2(s)\| e^{-k_\varphi(t-s)}\, ds, \quad (2.38)$$

and finally that

$$\langle x_0, \varphi(0)\rangle \leq \|x_0\| \|x(t)\| M_\varphi e^{-k_\varphi t} \leq \|x_0\| \|x(t)\| M_\varphi \sqrt{e^{-k_\varphi t}}. \quad (2.39)$$

Now we substitute all estimates (2.35)-(2.39) into (2.32) while taking the minimum over (2.37) and (2.38) and cancel $\|x(t)\|$ on both sides. Taking squares on both sides and using the simple inequality

$$(a+b+c+d)^2 \leq 4(a^2+b^2+c^2+d^2),$$

for $a, b, c, d \geq 0$, we obtain the following pointwise estimate for x:

$$\begin{aligned} \|x(t)\|^2 \leq c \left(\int_0^t (\|Cx(s)\|_Y^2 + \|R^{-*}B\lambda(s)\|_U^2) e^{-k_\varphi(t-s)}\, dt + \|x_0\|^2 e^{-k_\varphi t} \right) \\ + c \min\left\{ \int_0^t \|l_2(s)\|^2 e^{-k_\varphi(t-s)}\, ds, \left(\int_0^t \|l_2(s)\| e^{-k_\varphi(t-s)}\, ds \right)^2 \right\}. \end{aligned} \quad (2.40)$$

To derive an estimate for $\|x\|_{C(0,T;X)}^2$, we estimate all exponential functions by 1, extend the domains of integration from $[0,t]$ to $[0,T]$, and take the maximum over all $t \in [0,T]$:

$$\begin{aligned} &\|x\|_{C(0,T;X)}^2 \leq \\ &c\Big(\|Cx\|_{L_2(0,T;Y)}^2 + \|R^{-*}B^*\lambda\|_{L_2(0,T;U)}^2 + \min\{\|l_2\|_{L_2(0,T;X)}^2, \|l_2\|_{L_1(0,T;X)}^2\} + \|x_0\|^2 \Big). \end{aligned} \quad (2.41)$$

Similarly via (2.33) we get:

$$\begin{aligned} &\|\lambda\|_{C(0,T;X)}^2 \leq \\ &c\Big(\|Cx\|_{L_2(0,T;Y)}^2 + \|R^{-*}B^*\lambda\|_{L_2(0,T;U)}^2 + \min\{\|l_1\|_{L_2(0,T;X)}^2, \|l_1\|_{L_1(0,T;X)}^2\} + \|\lambda_T\|^2 \Big). \end{aligned} \quad (2.42)$$

To derive an estimate for $\|x\|^2_{L_2(0,T;X)}$, we have to integrate (2.40) over $[0,T]$ and apply Lemma 2.35 to the integral terms in (2.40). Setting $w(s) := \|l_2(s)\|^2$ in Lemma 2.35, we obtain

$$h_1(t) = \int_0^t \|l_2(s)\|^2 e^{-k_\varphi(t-s)}\, ds$$

and we conclude that

$$\int_0^T \int_0^t \|l_2(s)\|^2 e^{-k_\varphi(t-s)}\, ds\, dt = \|h_1\|_{L_1(0,T)} \le c\|w\|_{L_1(0,T)} = c\|l_2\|^2_{L_2(0,T;X)}$$

with Lemma 2.35 and similarly that

$$\int_0^T \int_0^t (\|Cx(s)\|_Y^2 + \|R^{-*}B^*\lambda(s)\|_U^2) e^{-k_\varphi(t-s)} ds\, dt \\ \le c\left(\|Cx(s)\|^2_{L_2(0,T;Y)} + \|R^{-*}B^*\lambda(s)\|^2_{L_2(0,T;U)}\right).$$

If we set $w(s) := \|l_2(s)\|$ Lemma 2.35, then

$$h_1(t) = \int_0^t \|l_2(s)\| e^{-k_\varphi(t-s)}\, ds,$$

and we obtain that

$$\int_0^T \left(\int_0^t \|l_2(s)\| e^{-k_\varphi(t-s)}\right)^2 ds\, dt = \|h_1\|^2_{L_2(0,T)} \le c\|w\|^2_{L_1(0,T)} = c\|l_2\|^2_{L_1(0,T;X)}.$$

This yields the desired L_2-estimate:

$$\|x\|^2_{L_2(0,T;X)} \le \\ c\left(\|Cx\|^2_{L_2(0,T;Y)} + R^{-*}B^*\lambda(s)\|^2_{L_2(0,T;U)} + \min\{\|l_2\|^2_{L_2(0,T;X)}, \|l_2\|^2_{L_1(0,T;X)}\} + \|x_0\|^2\right). \tag{2.43}$$

In the same way we compute

$$\|\lambda\|^2_{L_2(0,T;X)} \le \\ c\left(\|Cx\|^2_{L_2(0,T;Y)} + \|R^{-*}B^*\lambda(s)\|^2_{L_2(0,T;U)} + \min\{\|l_1\|^2_{L_2(0,T;X)}, \|l_1\|^2_{L_1(0,T;X)}\} + \|\lambda_T\|^2\right). \tag{2.44}$$

Now, we take the maximum of (2.41) and (2.43) and add it to the maximum of (2.42) and (2.44). Using the definition

$$\|v\|^2_{2\wedge\infty} = \max\{\|v\|^2_{L_2(0,T;X)}, \|v\|^2_{C(0,T;X)}\}, \quad \|w\|^2_{1\vee 2} = \min\{\|w\|^2_{L_1(0,T;X)}, \|w\|^2_{L_2(0,T;X)}\},$$

our result follows. □

The first two terms on the right hand side of (2.34) still depend on the state and the adjoint. We therefore present the following representation formula, motivated by [135, Proof of Theorem 1].

Lemma 2.37. *Let (x, λ) solve (2.27). Then*

$$\begin{aligned}
\|Cx\|^2_{L_2(0,T;Y)} &+ \|R^{-*}B^*\lambda\|^2_{L_2(0,T;U)} \\
&= -\langle \lambda_T, x(T)\rangle + \langle x_0, \lambda(0)\rangle + \int_0^T \langle l_2(s), \lambda(s)\rangle - \langle l_1(s), x(s)\rangle\, ds \\
&\le c\left(\|\lambda_T\|\|x(T)\| + \|x_0\|\|\lambda(0)\| + \|l_2\|_{1\vee 2}\|\lambda\|_{2\wedge\infty} + \|l_1\|_{1\vee 2}\|x\|_{2\wedge\infty}\right).
\end{aligned} \tag{2.45}$$

Proof. We apply Lemma 2.22 ii) to the state and adjoint equation, which yields

$$\langle \lambda_T, x(T)\rangle - \langle x_0, \lambda(0)\rangle = \int_0^T \langle l_2(s), \lambda(s)\rangle - \langle l_1(s), x(s)\rangle - \|Cx(s)\|^2_Y - \|R^{-*}B^*\lambda(s)\|^2_U\, ds.$$

Rearranging the terms and the Cauchy-Schwarz inequality yields the result. □

Theorem 2.38. *Let Assumption 2.32 hold. Then there is $c \ge 0$ independent of T such that*

$$\|M^{-1}\|_{L(((L_1(0,T;X),\|\cdot\|_{1\vee 2})\times X)^2,(C(0,T;X),\|\cdot\|_{2\wedge\infty})^2)} \le c.$$

Proof. Consider $z := (x, \lambda) \in C(0, T; X)^2$ and $r := (l_1, \lambda_T, l_2, x_0) \in (L_1(0, T; X) \times X)^2$ that satisfy (2.27). Thus, as shown in Lemma 2.36, the estimate (2.34) applies. We substitute (2.45) into (2.34) and apply the Hölder-like inequality of Lemma 2.31 to the integral terms to obtain that

$$\begin{aligned}
\|z\|^2_{2\wedge\infty} &:= \|x\|^2_{2\wedge\infty} + \|\lambda\|^2_{2\wedge\infty} \\
&\le c\left(\|\lambda_T\|\|x(T)\| + \|x_0\|\|\lambda(0)\| + \|l_2\|_{1\vee 2}\|\lambda\|_{2\wedge\infty} + \|l_1\|_{1\vee 2}\|x\|_{2\wedge\infty} + \|r\|^2_{1\vee 2}\right) \\
&\le c\left((\|\lambda_T\| + \|l_1\|_{1\vee 2})\|x\|_{2\wedge\infty} + (\|x_0\| + \|l_2\|_{1\vee 2})\|\lambda\|_{2\wedge\infty} + \|r\|^2_{1\vee 2}\right) \\
&\le c\left(\|r\|_{1\vee 2}\|z\|_{2\wedge\infty} + \|r\|^2_{1\vee 2}\right).
\end{aligned}$$

The application of the simple estimate $c\|r\|_{1\vee 2}\|z\|_{2\wedge\infty} \le \frac{1}{2}(c^2\|r\|^2_{1\vee 2} + \|z\|^2_{2\wedge\infty})$ implies the estimate

$$\|M^{-1}r\|_{2\wedge\infty} = \|z\|_{2\wedge\infty} \le c\|r\|_{1\vee 2}$$

and hence the desired result. □

2.3 The case of unbounded control or observation

In this section, we extend the results of Section 2.2 to the case of a control operator B that is unbounded as a mapping into X but admissible for the strongly continuous semigroup generated by A in the sense of [139, Chapter 4]. The reader is referred to [139] for an in-depth introduction to this topic. An unbounded control operator often arises in the case of boundary control. As a consequence, the operator norm $\|B\|_{L(U,X)} = \|B^*\|_{L(X,U)}$ is no longer finite. Additionally, we will allow for a state feedback operator K_B in Assumption 2.32 that is unbounded but admissible. Inspection of the proofs in Sections 2.2.1 and 2.2.3 yields that norms of the control resp. observation operator are only used in inequality (2.36) and the dual version of (2.35). In addition, the estimate of the control in Theorem 2.27 is performed via $R^*Ru = B^*\lambda$ using $\|B^*\|_{L(X,U)}$ in (2.15). All remaining estimates and constants do not involve norms of B or K_B.

The goal of this section is to replace boundedness of B and the feedback operator K_B in Assumption 2.32 by a weaker property, which is known as *admissibility* while maintaining the stability result of Lemma 2.36. Our strategy of proof will be to show surrogates for (2.36) and the dual version of (2.35) involving K_B that allow us to generalize our main results to the case of admissible control and feedback operators. All further steps of the proofs remain unchanged. The sensitivity results in Theorem 2.27 for the state and adjoint directly carry over as stated in Theorem 2.48, whereas the estimate for the control in the proof of Theorem 2.27 involves the norm of the control operator. We therefore modify the proof to obtain the results in Theorem 2.48 below. However, we only obtain an integral estimate but no uniform estimate of the control in Theorem 2.27. This is the only price to pay for going from bounded to unbounded but admissible control operators.

2.3.1 Well-posed linear systems and admissibility

We recall the definition of admissible control and observation operators. Let $A\colon D(A) \subset X \to X$ be the generator of a strongly continuous semigroup $(\mathcal{T}(t))_{t\geq 0}$ on X. Moreover, let A^* be the adjoint operator of A with domain $D(A^*)$. Let X_1 be $D(A)$ equipped with the norm $\|\cdot\|_1 := \|(\beta I - A)\cdot\|$ for $\beta \in \rho(A)$, where $\rho(A) := \{\beta \in \mathbb{C} \,|\, \beta I - A \text{ is continuously invertible and } (\beta I - A)^{-1} \in L(X)\}$ is the resolvent set of A as defined in Definition 2.8. Second, again for $\beta \in \rho(A)$, we define X_{-1} to be the completion of X with respect to the norm $\|\cdot\|_{-1} := \big\|(\beta I - A)^{-1}\cdot\big\|$. We note that the norms $\|\cdot\|_1$ for different β are equivalent, see [139, Proposition 2.10.1], and the same also holds true for $\|\cdot\|_{-1}$, see [139, Proposition 2.10.2]. Furthermore, by, e.g., [139, Proposition 2.10.4], the strongly continuous semigroup $(\mathcal{T}(t))_{t\geq 0}$ can be extended to a strongly continuous semigroup on X_{-1}, which we will denote by the same symbol $(\mathcal{T}(t))_{t\geq 0}$.

Definition 2.39. *([139, Definition 4.2.1, Definition 4.3.1])*

i) *$B \in L(U, X_{-1})$ is called an admissible control operator for the strongly continuous semigroup $(\mathcal{T}(t))_{t\geq 0}$ if for some $\tau > 0$, $\operatorname{ran}\Phi_\tau \subset X$, where*

$$\Phi_\tau u := \int_0^\tau \mathcal{T}(\tau - s)Bu(s)\,ds$$

for $u \in L_2(0, \infty; U)$.

ii) *$C \in L(X_1, Y)$ is called an admissible observation operator for the strongly continuous semigroup $(\mathcal{T}(t))_{t\geq 0}$ if for some $\tau > 0$, $\Psi_\tau \in L(X_1, L_2(0, \infty, Y))$ has a continuous extension to X, where*

$$(\Psi_\tau z_0)(t) := \begin{cases} C\mathcal{T}(t)z_0 & \text{for } t \in [0, \tau] \\ 0 & \text{for } t > \tau \end{cases}$$

for $z_0 \in X_1$.

Note that if i) and ii) in Definition 2.39 are satisfied for one $\tau \geq 0$, they hold for all $\tau \geq 0$, see [139, Proposition 4.2.2, Proposition 4.3.2].

We briefly recall some properties of admissible control operators that will be important in the remainder of this section.

Proposition 2.40. *Let B be an admissible control operator for the strongly continuous semigroup $(\mathcal{T}(t))_{t\geq 0}$. Then,*

i) B^* is an admissible observation operator for the adjoint semigroup $(\mathcal{T}^*(t))_{t\geq 0}$.

ii) For all $t \geq 0$ and $x(t) \in D(A^*)$, there exists a constant $K_t \geq 0$ such that

$$\int_0^t \|B^*\mathcal{T}^*(t-s)x(t)\|_U^2 \, ds \leq K_t^2 \|x(t)\|^2. \tag{2.46}$$

iii) If $(\mathcal{T}^*(t))_{t\geq 0}$ is exponentially stable, the constant K_t can be chosen independently of t.

Proof. Part i) follows from the duality result [139, Theorem 4.4.3]. For ii), see [139, Definition 4.3.1]. The fact that the bound can be chosen independently of t, as stated in iii), follows from [139, Remark 4.3.5]. □

Remark 2.41. *We briefly comment on the inequality (2.46).*

- *As the norm in the upper bound is the norm in X, this estimate can be extended to all $x(t) \in X$ by density of $D(A^*)$ in X.*
- *A very prominent example, where an estimate like (2.46) holds for an unbounded control operator is in the case of the wave equation with $B^* = \frac{\partial}{\partial \nu} = \langle \nabla \cdot, \nu \rangle$, where ν is the outer unit normal, cf. [139, Section 7.1]. This feature is often referred to as hidden regularity. It can be shown that the estimate (2.46) holds for solutions of the wave equation even for initial displacements in $H_0^1(\Omega)$ and $U = L_2(\partial\Omega)$. This is not obvious, as due to the lack of a smoothing property of the wave equation, the displacements are in $H_0^1(\Omega)$ for every time point and B^* is not bounded from $H_0^1(\Omega)$ to $L_2(\partial\Omega)$. For a proof of this hidden regularity property, the reader is referred to [89] or [139, Theorem 7.1.3].*

In addition to the concept of admissible control and observation operators, we will make use of the notion of *well-posed linear systems*, for which certain desirable properties hold. Besides continuity of the state trajectory, these systems enjoy boundedness of input-to-state, state-to-output and input-to-output maps, as partly defined in Definition 2.39, although the generating operators can be unbounded. For an in-depth treatment of this topic, the interested reader is referred to the seminal papers [124, 125, 145], to the monograph [132], and the survey articles [140, 146].

A possible approach of defining well-posed linear systems is to require particular properties of the maps mentioned above, cf. [140, Definition 3.1], which is very similarly to the way strongly continuous semigroups were introduced in Definition 2.1. In the context of well-posed linear systems these properties include, e.g., that the initial-value-to-state map is given by a strongly continuous semigroup, that the state does not depend on the future input and that the past output does not depend on the future input, the latter two often being called causality. In order to keep the presentation concise, the reader is referred to [140, Definition 3.1] for a precise definition of the properties mentioned above. An important fact in our context is that the class of operators (A, B, C) generating such well-posed linear systems are precisely characterized by A generating a strongly continuous semigroup, admissibility of B and C with respect to the strongly continuous semigroup, and a condition on the transfer function, assuring boundedness of the input-output map, cf. [140, Proposition 4.9] and [132, Theorem 4.2.1, Theorem 4.4.2].

Proposition 2.42. *([140, Proposition 4.9])*
A triple of operators (A, B, C) is well-posed on (U, X, Y) if and only if the following conditions hold:

- *A generates a strongly continuous semigroup $(\mathcal{T}(t))_{t\geq 0}$ on X,*
- *$B \in L(U, X_{-1})$ is an admissible control operator for $(\mathcal{T}(t))_{t\geq 0}$,*
- *$C \in L(X_1, Y)$ is an admissible observation operator for $(\mathcal{T}(t))_{t\geq 0}$,*
- *some (hence every) transfer function associated with (A, B, C) is proper,*

where the transfer function is characterized by $G(s_1) - G(s_2) = C\left((s_1 I - A)^{-1} - (s_2 I - A)^{-1}\right) B$ for every $s_1, s_2 \in \{s \in \mathbb{C} \,|\, \Re(s) > \omega_0(\mathcal{T})\}$ and $\omega_0(\mathcal{T})$ is the type of the strongly continuous semigroup, cf. Definition 2.10. An analytic function with a domain in some right half plane is called proper if it is bounded on some right half plane [140, p.8].

The optimality conditions (2.6) are derived assuming boundedness of the control and observation operator. For specific cases, i.e., e.g., boundary controlled wave equations, optimality conditions are given in, e.g., [96] using functional analytic methods. However, to the author's best knowledge, optimality conditions in the abstract setting with unbounded control and observation as presented here are not yet available. Thus, in the remainder, we will assume that (x, u, λ) solves the extremal equations, rather than that (x, u) solve the optimal control problem. Then, of course, the question of solvability of the extremal equation arises, i.e., existence and uniqueness of solutions. The solution operators norm that we will derive in Theorem 2.47 implies that if a solution to the extremal equations exists, it is unique. The following theorem gives a partial answer to the question of existence.

Theorem 2.43. *Let (A, B, C) form a well-posed system and I resp. Q^{-1} be admissible feedback operators in the sense of [145, Definition 3.5]. Then the operator $\begin{pmatrix} C^*C & A^* \\ A & BQ^{-1}B^* \end{pmatrix}$ in (2.6) with domain $D(\mathcal{A} + \mathcal{B}K\mathcal{C})$, where*

$$\mathcal{A} = \begin{pmatrix} 0 & A^* \\ A & 0 \end{pmatrix}, \mathcal{B} = \begin{pmatrix} 0 & C^* \\ B & 0 \end{pmatrix}, \mathcal{C} = \begin{pmatrix} 0 & B^* \\ C & 0 \end{pmatrix}, K = \begin{pmatrix} I & 0 \\ 0 & Q^{-1} \end{pmatrix}$$

generates a strongly continuous semigroup on $X \times X$.

Proof. By the the well-posedness of (A, B, C), it follows that $(\mathcal{A}, \mathcal{B}, \mathcal{C})$ form a well-posed system on $X \times X$. Choosing the admissible feedback operator $K = \begin{pmatrix} I & 0 \\ 0 & Q^{-1} \end{pmatrix}$, the closed-loop system $(\mathcal{A}+\mathcal{B}K\mathcal{C}, \mathcal{B}, \mathcal{C})$ with $\mathcal{A}+\mathcal{B}K\mathcal{C} = \begin{pmatrix} C^*C & A^* \\ A & BQ^{-*}B^* \end{pmatrix}$ forms another well-posed system, cf. [132, Theorem 7.1.2] or [145]. In particular, the operator $\mathcal{A}+\mathcal{B}K\mathcal{C}$ generates a strongly continuous semigroup and $\mathcal{C} = \begin{pmatrix} 0 & B^* \\ C & 0 \end{pmatrix}$ is an admissible observation operator for the semigroup generated by $\mathcal{A}+\mathcal{B}K\mathcal{C}$. □

Remark 2.44. *Theorem 2.43 shows that, even in the unbounded setting, assuming a well-posed system, the operator occurring in the extremal equations (2.27) generates a semigroup on $X \times X$. However, it is not clear how this corresponds to solvability of a forward-backward systems of the type*

$$\begin{pmatrix} x'(t) \\ -\lambda'(t) \end{pmatrix} = \mathcal{A}_{cl} \begin{pmatrix} x(t) \\ \lambda(t) \end{pmatrix}, \qquad \begin{pmatrix} x(0) \\ \lambda(T) \end{pmatrix} = \begin{pmatrix} x_0 \\ \lambda_T \end{pmatrix},$$

where $\mathcal{A}_{cl} = \begin{pmatrix} C^*C & A^* \\ A & BQ^{-*}B^* \end{pmatrix}$. The main problem is the terminal condition $\lambda(T) = \lambda_T$. One strategy for showing solvability could be to substitute $p(t) = \lambda(T-t)$, which yields

$$\begin{pmatrix} x'(t) \\ p'(T-t) \end{pmatrix} = \mathcal{A}_{cl} \begin{pmatrix} x(t) \\ p(T-t) \end{pmatrix}, \qquad \begin{pmatrix} x(0) \\ p(0) \end{pmatrix} = \begin{pmatrix} x_0 \\ \lambda_T \end{pmatrix},$$

where now the temporal argument has changed in the second equation. Moreover, even though $\mathcal{A}_{cl}$ generates a semigroup, the solutions are only defined for $t \in [0, T]$, not, as by the definition of a semigroup, for $t \geq 0$.

Assumption 2.45. *Let A generate a semigroup $(\mathcal{T}(t))_{t\geq 0}$ on X, $C \in L(X, Y)$ and $B \in L(U, X_{-1})$ be an admissible control operator for $(\mathcal{T}(t))_{t\geq 0}$. Further, assume that*

i) (A, C) is exponentially detectable, i.e., there exists $K_C \in L(Y, X)$ such that the semigroup generated by $A^* + C^*K_C^*$ is exponentially stable.

ii) (A, B) is exponentially stabilizable, i.e., there exists $K_B \in L(X_1, U)$ such that

- *(A, B, K_B) is well-posed on (U, X, U) and*
- *$A + BK_B$ with domain $D(A + BK_B)$ generates an exponentially stable semigroup.*

Remark 2.46. *In a similar fashion, one could allow for unbounded but admissible C and K_C when assuming B and K_B to be bounded. The case where all of the operators C, B, K_C and K_B are unbounded but admissible cannot be included in all generality. This is due to the non-existence of perturbation results for this case, i.e., not every admissible observation operator C for A is admissible for $A+BK$ if B is an unbounded but admissible control operator and K is an* *unbounded but admissible feedback operator. Mixed perturbation results of this kind are a very delicate matter and the reader is referred to [139, Proposition 5.5.2, Example 5.5.3, Proposition 10.1.10] and the discussion after [139, Corollary 5.5.1]. In particular, such a perturbation result is obtained in the example in [139, Section 10.8]. Finally, for the particular case of* $K = LC$ *with* $L \in L(Y, U)$*, such perturbation results are established in [132, Theorem 7.1.2] or [145].*

The strategy we will pursue in the following will be to apply the admissibility estimate of Proposition 2.40 to φ and ψ in Lemma 2.33. As the underlying semigroups in Lemma 2.33 are exponentially stable, we may apply the estimate (2.46) to the respective admissible control operators with K_t independent of t. This will turn out to be an appropriate replacement for assuming finiteness of $\|B^*\|_{L(X,U)}$ and $\|K_B\|_{L(X,U)}$.

2.3.2 Scaling results and T-independent bounds

With the help of the concept of admissible operators as defined in the previous section, we now present all necessary modifications of the proofs of Section 2.2 to the case of an unbounded but admissible control and feedback operator.

Theorem 2.47. *Consider* M *defined in* (2.6) *and let Assumption 2.45 hold. Then there is* $c \geq 0$ *independent of* T *such that*

$$\|M^{-1}\|_{L(((L_1(0,T;X),\|\cdot\|_{1\vee 2})\times X)^2,(C(0,T;X),\|\cdot\|_{2\wedge\infty})^2)} \leq c.$$

Proof. As already noted, the only step in the proof of Theorem 2.38 where the operator norm of B and K_B is needed, is the proof of (2.36) and (2.35) with their respective dual counterparts. It is thus sufficient to show a modification of this inequality that circumvents this operator norm estimate by using exponential stability of the test functions and the fact that B is an admissible control operator. We will sketch the proof for the estimate (2.36) and the estimate including K_B in the dual counterpart of (2.35) follows completely analogously.

Let φ solve (2.28), i.e., $\varphi(s) = \mathcal{T}_{\text{cl}}^*(t-s)x(t)$, where $(\mathcal{T}_{\text{cl}}^*(t))_{t\geq 0}$ is the strongly continuous semigroup generated by $(A^* + C^*K_C^*)$. By assumption, B is an admissible control operator for the semigroup generated by A and $K_C C \in L(X,X)$. We obtain that B is an admissible control operator for $(A + K_C C)$ by a perturbation result, cf. [139, Theorem 5.4.2]. Hence B^* is an admissible observation operator for the adjoint semigroup generated by $(A + K_C C)^*$. We show that the critical estimate (2.36) still holds with different constants, which do not involve the operator norm of B. First, a simple calculation using the fact that R^{-*} is bounded from U to U and applying the Cauchy-Schwarz inequality twice yields

$$\begin{aligned}\int_0^t |\langle R^{-*}B^*\lambda(s), R^{-*}B^*\varphi(s)\rangle_U|\,ds &\leq c\int_0^t e^{-\frac{k_\varphi}{2}(t-s)}\|R^{-*}B^*\lambda(s)\|_U \left\|B^* e^{\frac{k_\varphi}{2}(t-s)}\varphi(s)\right\|_U ds \\ &\leq c\sqrt{\int_0^t e^{-k_\varphi(t-s)}\|R^{-*}B^*\lambda(s)\|_U^2\,ds}\sqrt{\int_0^t \left\|B^* e^{\frac{k_\varphi}{2}(t-s)}\varphi(s)\right\|_U^2 ds.}\end{aligned} \tag{2.47}$$

By the exponential stability $\|\mathcal{T}_{\mathrm{cl}}^*(t-s)\| \le M_\varphi e^{-k_\varphi(t-s)}$, the scaled semigroup $\left(e^{\frac{k_\varphi}{2}(t-s)}\mathcal{T}_{\mathrm{cl}}^*(t)\right)_{t\ge 0}$ is still exponentially stable, cf. Definition 2.12. Hence we employ Proposition 2.40 ii) and iii) and obtain

$$\sqrt{\int_0^t \left\|B^* e^{\frac{k_\varphi}{2}(t-s)}\varphi(s)\right\|^2 ds} = \sqrt{\int_0^t \left\|B^* e^{\frac{k_\varphi}{2}(t-s)}\mathcal{T}_{\mathrm{cl}}^*(t-s)x(t)\right\|^2 ds} \le K\|x(t)\|$$

with K being independent of t. Inserting this into (2.47), we conclude that

$$\int_0^t |\langle R^{-*}B^*\lambda(s), R^{-*}B^*\varphi(s)\rangle|\, ds \le c\|x(t)\|\sqrt{\int_0^t e^{-k_\varphi(t-s)}\|R^{-*}B^*\lambda(s)\|^2\, ds},$$

which yields the desired replacement for (2.36) with a different constant independent of the norm of B. As the remaining results in Section 2.2.3, namely Lemma 2.36, Lemma 2.37 and Theorem 2.38 do not hinge on boundedness, we can conclude the result analogously to the bounded case. □

As a consequence of Theorem 2.47, the estimates in Theorem 2.27 also hold true in the case of unbounded control and feedback operators with constants independent of the horizon T, except for the uniform estimate for the control. This is the statement of the following theorem.

Theorem 2.48. *Let Assumption 2.45 hold. Assume $(\delta x, \delta\lambda) \in C(0,T;X)^2$ solve (2.8) with $\varepsilon_1, \varepsilon_2 \in L_1(0,T;X)$. Let $\delta u = Q^{-1}B^*\delta\lambda$. Then there is a scaling factor $\mu > 0$ satisfying*

$$\mu < \frac{1}{\|M^{-1}\|_{L((L_2(0,T;X)\times X)^2, L_2(0,T;X)^2)}}$$

and a constant $c \ge 0$, both independent of T, such that defining

$$\rho := \left\|e^{-\mu t}\varepsilon_1(t)\right\|_S + \left\|e^{-T}\varepsilon_T\right\| + \left\|e^{-\mu t}\varepsilon_2(t)\right\|_S + \|\varepsilon_0\|$$

for $S := L_1(0,T;X)$ or $S := L_2(0,T;X)$, it holds that

$$\begin{aligned}\left\|e^{-\mu t}\delta x(t)\right\|_{L_2(0,T;X)} + \left\|e^{-\mu t}\delta u(t)\right\|_{L_2(0,T;U)} + \left\|e^{-\mu t}\delta\lambda(t)\right\|_{L_2(0,T;X)} &\le c\rho,\\ \left\|e^{-\mu t}\delta x(t)\right\|_{C(0,T;X)} + \left\|e^{-\mu t}\delta\lambda(t)\right\|_{C(0,T;X)} &\le c\rho.\end{aligned}$$

Proof. First, choosing μ, such that $\mu\|M^{-1}\|_{L((L_2(0,T;X)\times X)^2, L_2(0,T;X)^2)} < 1$, by the same reasoning as in the proof of Theorem 2.27 we conclude the estimates for the state and adjoint

$$\begin{aligned}\left\|e^{-\mu t}\delta x(t)\right\|_{L_2(0,T;X)} + \left\|e^{-\mu t}\delta\lambda(t)\right\|_{L_2(0,T;X)} &\le c\rho,\\ \left\|e^{-\mu t}\delta x(t)\right\|_{C(0,T;X)} + \left\|e^{-\mu t}\delta\lambda(t)\right\|_{C(0,T;X)} &\le c\rho,\end{aligned} \tag{2.48}$$

with $c, \mu \geq 0$ independent of T as the occurring operator norms can be bounded independently of T by Theorem 2.47. To estimate the control, we set $\widetilde{\delta x}(t) := e^{-\mu t}\delta x(t)$, $\widetilde{\delta\lambda}(t) := e^{-\mu t}\delta\lambda(t)$ and $\widetilde{\delta u}(t) := e^{-\mu t}\delta u$ and compute that

$$\|\widetilde{\delta u}\|^2_{L_2(0,T;U)} = \int_0^T \|Q^{-1}B^*\widetilde{\delta\lambda}(t)\|^2_U \, dt \leq c\int_0^T \|C\widetilde{\delta x}(t)\|^2_Y + \|R^{-*}B^*\widetilde{\delta\lambda}(t)\|^2_U \, dt.$$

Similarly to Lemma 2.37 we obtain that

$$\begin{aligned}
&\int_0^T \left\|C\widetilde{\delta x}(t)\right\|^2 + \left\|R^{-*}B^*\widetilde{\delta\lambda}(t)\right\|^2 dt \\
&= \langle\widetilde{\delta x}(T), \widetilde{\delta\lambda}(T)\rangle - \langle\widetilde{\delta x}(0), \widetilde{\delta\lambda}(0)\rangle \int_0^T -\langle\tilde{\varepsilon}_1(t), \widetilde{\delta x}(t)\rangle + \langle\tilde{\varepsilon}_2(t), \widetilde{\delta\lambda}(t)\rangle - 2\mu\langle\widetilde{\delta x}(t), \widetilde{\delta\lambda}(t)\rangle \, dt \\
&\leq \left(\|\tilde{\varepsilon}_1\|_{1\vee 2} + \|e^{-T}\varepsilon_T\| + \|\varepsilon_0\| + \|\tilde{\varepsilon}_2\|_{1\vee 2}\right)\left(\|\widetilde{\delta x}\|_{2\wedge\infty} + \|\widetilde{\delta\lambda}\|_{2\wedge\infty}\right) + 2\mu\|\widetilde{\delta x}\|_{L_2(0,T;X)}\|\widetilde{\delta\lambda}\|_{L_2(0,T;X)} \\
&\leq c\rho^2,
\end{aligned}$$

where we used Lemma 2.22 ii) for the state and adjoint equation, with $\tilde{\varepsilon}_i(t) = e^{-\mu t}\varepsilon_i(t)$, $i = 1, 2$ and the Hölder inequality of Lemma 2.31. In the last estimate we used the bounds on the scaled right-hand side and the estimate on the state and the adjoint (2.48). Taking the square root yields the result for the control. □

As a second consequence of the T-independent bound of Theorem 2.47, we obtain a generalization of the turnpike result of Theorem 2.30 to the case of unbounded control.

Theorem 2.49. *Let Assumption 2.45 hold. Assume (x, u, λ) solves (2.4). Moreover, let $(\bar{x}, \bar{u}, \bar{\lambda})$ solve the corresponding steady state problem (2.16) and set $(\delta x, \delta u, \delta\lambda) := (x - \bar{x}, u - \bar{u}, \lambda - \bar{\lambda})$. Then, there exist μ satisfying $0 < \mu < \frac{1}{\|M^{-1}\|_{L((L_2(0,T;X)\times X)^2, L_2(0,T;X)^2)}}$ and a constant $c \geq 0$, both independent of T, such that*

$$\begin{aligned}
\left\|\frac{1}{e^{-\mu t} + e^{-\mu(T-t)}}\delta x(t)\right\|_{L_2(0,T;X)} + \left\|\frac{1}{e^{-\mu t} + e^{-\mu(T-t)}}\delta u(t)\right\|_{L_2(0,T;U)} \\
+ \left\|\frac{1}{e^{-\mu t} + e^{-\mu(T-t)}}\delta\lambda(t)\right\|_{L_2(0,T;X)} \leq c\left(\|x_0 - \bar{x}\| + \|\bar{\lambda}\|\right), \\
\left\|\frac{1}{e^{-\mu t} + e^{-\mu(T-t)}}\delta x(t)\right\|_{C(0,T;X)} + \left\|\frac{1}{e^{-\mu t} + e^{-\mu(T-t)}}\delta\lambda(t)\right\|_{C(0,T;X)} \leq c\left(\|x_0 - \bar{x}\| + \|\bar{\lambda}\|\right).
\end{aligned}$$

Proof. Again, choosing μ, such that $\beta := \mu\|M^{-1}\|_{L((L_2(0,T;X)\times X)^2, L_2(0,T;X)^2)} < 1$ and completely analogous to the proof of Theorem 2.30, we conclude the estimates for state and adjoint, i.e., with $\tilde{z} := \frac{1}{e^{-\mu t}+e^{-\mu(T-t)}}(\delta x, \delta\lambda)$ we have

$$\begin{aligned}
\|\tilde{z}\|_{L_2(0,T;X)^2} &\leq \frac{\|M^{-1}\|_{L((L_2(0,T;X)\times X)^2, L_2(0,T;X)^2)}}{1-\beta}\|\tilde{r}\|_{(L_2(0,T;X)\times X)^2}, \\
\|\tilde{z}\|_{C(0,T;X)^2} &\leq \frac{\|M^{-1}\|_{L((L_2(0,T;X)\times X)^2, C(0,T;X)^2)}}{1-\beta}\|\tilde{r}\|_{(L_2(0,T;X)\times X)^2}
\end{aligned} \tag{2.49}$$

with $\tilde{r} := \frac{1}{1+e^{-\mu T}}(0, -\bar{\lambda}, 0, x_0 - \bar{x})$, cf. (2.25). The remainder of this proof consists of estimating the control to conclude (2.21). To this end, we recall the approach taken in the proof of Theorem 2.48, set $\widetilde{\delta u} := \frac{1}{e^{-\mu t}+e^{-\mu(T-t)}}\delta u$ and obtain that

$$\|\widetilde{\delta u}\|^2_{L_2(0,T;U)} = \int_0^T \left\|Q^{-1}B^*\widetilde{\delta\lambda}(t)\right\|^2 dt \leq c\int_0^T \left\|C\widetilde{\delta x}(t)\right\|^2 + \left\|R^{-*}B^*\widetilde{\delta\lambda}(t)\right\|^2 dt. \tag{2.50}$$

Again, similarly to Lemma 2.37, using Lemma 2.22 ii) for the scaled state and adjoint equation of system, we obtain that

$$\begin{aligned}
&\int_0^T \|C\widetilde{\delta x}(t)\|_Y^2 + \|B^*\widetilde{\delta\lambda}(t)\|_U^2\, dt \\
&\quad = \langle\widetilde{\delta x}(T), \widetilde{\delta\lambda}(T)\rangle - \langle\widetilde{\delta x}(0), \widetilde{\delta\lambda}(0)\rangle + \mu\int_0^T \langle F\widetilde{\delta x}(t), \widetilde{\delta\lambda}(t)\rangle + \langle\widetilde{\delta x}(t), F\widetilde{\delta\lambda}(t)\rangle\, dt \\
&\quad \leq \|\widetilde{\delta x}(T)\|\|\widetilde{\delta\lambda}(T)\| + \|\widetilde{\delta x}(0)\|\|\widetilde{\delta\lambda}(0)\| + 2\mu\|\widetilde{\delta x}\|_{L_2(0,T;X)}\|\widetilde{\delta\lambda}\|_{L_2(0,T;X)} \\
&\quad \leq \left(\|\widetilde{\delta x}(T)\| + \|\widetilde{\delta\lambda}(0)\|\right)\left(\|\widetilde{\delta x}(0)\| + \|\widetilde{\delta\lambda}(T)\|\right) + \mu\left(\|\widetilde{\delta x}\|^2_{L_2(0,T;X)} + \|\widetilde{\delta\lambda}\|^2_{L_2(0,T;X)}\right),
\end{aligned} \tag{2.51}$$

where we used that $\|F\|_{L(L_2(0,T;X),L_2(0,T;X))} \leq 1$. In order to estimate the end time value of the state and the initial value of the adjoint, we compute that

$$\begin{aligned}
\|\widetilde{\delta x}(T)\| + \|\widetilde{\delta\lambda}(0)\| &= \frac{1}{1+e^{-\mu T}}\left(\|\delta x(T)\| + \|\delta\lambda(0)\|\right) \leq \|\delta x\|_{C(0,T;X)} + \|\delta\lambda\|_{C(0,T;X)} \\
&\leq \|M^{-1}\|_{L((L_2(0,T;X)\times X)^2, C(0,T;X)^2)}\left(\|x_0 - \bar{x}\| + \|\bar{\lambda}\|\right).
\end{aligned}$$

Inserting this into (2.51) and (2.50) and using the estimate for the state and adjoint (2.25), we obtain that

$$\begin{aligned}
\|\widetilde{\delta u}\|^2_{L_2(0,T;U)} \leq c\Bigg(&\|M^{-1}\|^2_{L((L_2(0,T;X)\times X)^2, C(0,T;X)^2)}\left(\|x_0 - \bar{x}\|^2 + \|\bar{\lambda}\|^2\right) \\
&+ \mu\frac{\|M^{-1}\|^2_{L((L_2(0,T;X)\times X)^2, L_2(0,T;X)^2)}}{(1-\beta)^2}\left(\|x_0 - \bar{x}\|^2 + \|\bar{\lambda}\|^2\right)\Bigg),
\end{aligned}$$

where taking the square root and using $\mu\|M^{-1}\|_{L((L_2(0,T;X)\times X)^2, L_2(0,T;X)^2)} < 1$ yields

$$\begin{aligned}
\|\widetilde{\delta u}\|_{L_2(0,T;U)} \leq c\Bigg(&\|M^{-1}\|_{L((L_2(0,T;X)\times X)^2, C(0,T;X)^2)} \\
&+ \frac{\|M^{-1}\|_{L((L_2(0,T;X)\times X)^2, L_2(0,T;X)^2)}}{(1-\beta)^2}\Bigg)\left(\|x_0 - \bar{x}\| + \|\bar{\lambda}\|\right).
\end{aligned}$$

By Theorem 2.47 we obtain the T-independent bounds for the operator norms. Together with (2.25), the estimate (2.21) follows. □

2.4 The case of a terminal condition on the state

In this part, we will conclude another generalization of the approach taken in Section 2.2. More specifically, in addition to the initial condition, we will allow for a terminal condition on the state in Problem 2.23. In order to not hide the main ideas behind technical details, we will assume bounded control and feedback operators and discuss the case of an unbounded but admissible control operator in Remark 2.53. Intuitively, it is clear that in order to satisfy this constraint, the set of prescribed terminal values needs to be reachable in the sense that there is a control that steers the initial state to the specified terminal state. This concept is called controllability and we will briefly introduce it in the following subsection.

2.4.1 Observability and controllability

For an overview of controllability and observability of finite dimensional systems the interested reader is referred to the overview given in [83, Chapters 2-4] and [159, Chapter 2] or [130]. There exist many characterizations for controllability and observability, e.g., the Kalman rank condition, the Hautus test or observability estimates. A very important property is the duality of observability and controllability, similarly to the duality of detectability and stabilizability, cf. Assumption 2.32 or Assumption 2.45. Some concepts and properties of a finite dimensional setting carry over to an infinite dimensional setting; in particular, observability estimates or the duality mentioned above remain a very useful tool in the study of controllability and observability in infinite dimensions. However, as can be expected, there are some major differences in the infinite dimensional setting. First, there are several different concepts of controllability resp. observability, namely approximate and exact controllability resp. observability and additionally the notion of null controllability. For linear, time reversible systems, null controllability and exact controllability are equivalent, cf. [159, Remark 3.1 b)] or [139, Remark 6.1.2]. Moreover, other than in finite dimensions, controllability resp. observability at some time $t_c > 0$ does not imply controllability resp. observability for all times $t > 0$. The reader is referred to [34, 35, 139, 159] for an in-depth introduction to controllability and observability of infinite dimensional systems. For any $\tau \in [0, T]$, let us recall the input map $\phi_\tau \colon L_2(0, T; U) \to X$ with

$$\phi_\tau u := \int_0^\tau \mathcal{T}(\tau - s) B u(s)\, ds,$$

as defined in Definition 2.39 i). In the following, we will assume $(\mathcal{T}(t))_{t\geq 0}$ to always be the semigroup generated by $A \colon D(A) \subset X \to X$.

Definition 2.50. *(Exact and approximate controllability) We call (A, B) exactly controllable in time $t_c > 0$ if $\operatorname{ran} \phi_{t_c} = X$. Similarly, we call (A, B) approximately controllable in time t_c if $\overline{\operatorname{ran} \phi_{t_c}} = X$.*

It is clear that exact and approximate controllability coincide in finite dimensions. Moreover, it is obvious that exact controllability implies exponential stabilizability as defined in Assumption 2.32. An important characterization of controllability is the following observability inequality, which was proven first in the seminal paper [97] by the *Hilbert Uniqueness Method*.

Theorem 2.51 ([35, Theorem 4.1.7]). *(A, B) is exactly controllable in time $t_c > 0$ if and only if there is $\alpha_{t_c} > 0$ such that*

$$\int_0^{t_c} \|B^*\mathcal{T}^*(s)x_0\|_U^2 \, ds \geq \alpha_{t_c}\|x_0\|^2 \quad \forall\, x_0 \in X.$$

Using substitution in the previous estimate we immediately obtain that

$$\int_{T-t_c}^{T} \|B^*\mathcal{T}^*(T-s)\lambda_T\|_U^2 \, ds \geq \alpha_{t_c}\|\lambda_T\|^2 \quad \forall\, \lambda_T \in X. \tag{2.52}$$

2.4.2 Scaling results and T-independent bounds

In this section, we discuss an extension of our result to optimal control problems with a condition on the terminal state, i.e., adding a terminal condition $x(T) = x_T \in X$ in Problem 2.23. In the finite dimensional case, turnpike results for linear and nonlinear initial and terminal conditions were proven in [136]. In the Hilbert space setting, this problem was discussed in [135, Section 2.6], where, however, the lack of invertibility of the Lyapunov operator prohibited the derivation of a result for a terminal condition. To ensure existence of an optimal solution for arbitrary initial and terminal data, we assume (A, B) to be exactly controllable in time t_c where $0 < t_c \leq T$, i.e., for any initial datum $x_0 \in X$ and terminal state $x_T \in X$, we can find a control that drives the state from x_0 to x_T in any time $T \geq t_c$. This assumption excludes parabolic equations with control that does not act on the whole domain. The assumption is, however, fulfilled by many hyperbolic systems, cf. Section 2.5. For a discussion of controllability issues for PDEs, the reader is referred to the overview article [159].

Another crucial point in deriving Pontryagin Maximum Principles for problems including both initial and terminal conditions on the state in infinite dimensions is a codimensionality condition of the reachable set in X, cf. [95, Chapter 4]. This assumption is automatically satisfied if one assumes exact controllability, as the reachable set is the whole space X. For bounded control and observation operators, the optimal solutions satisfy the dynamics (2.4) with $x(0) = x_0$ and $x(T) = x_T$, where no terminal condition on the adjoint is imposed [95, Theorem 1.6]. Again, by eliminating the control, we obtain the extremal equations

$$\underbrace{\begin{pmatrix} C^*C & -\frac{d}{dt} - A^* \\ E_T & 0 \\ \frac{d}{dt} - A & -BQ^{-1}B^* \\ E_0 & 0 \end{pmatrix}}_{=:\widetilde{M}} \begin{pmatrix} x \\ \lambda \end{pmatrix} = \begin{pmatrix} C^*Cx_\mathrm{d} \\ x_T \\ Bu_\mathrm{d} + f \\ x_0 \end{pmatrix}. \tag{2.53}$$

We observe that in contrast to the initial condition on the state and the terminal condition on the adjoint equation in (2.27), the system (2.53) is subject to an initial and terminal condition on the state and no condition on the adjoint. As a consequence, the estimate presented in Lemma 2.36 contains the unknown value $\lambda(T)$. In order to bound this unknown quantity by the right hand side of (2.53), we will utilize the following observability estimate.

Proposition 2.52. *Let* $(x, \lambda) \in C(0, T; X)^2$ *solve*

$$\begin{pmatrix} C^*C & -\frac{d}{dt} - A^* \\ E_T & 0 \\ \frac{d}{dt} - A & -BQ^{-1}B^* \\ E_0 & 0 \end{pmatrix} \begin{pmatrix} x \\ \lambda \end{pmatrix} = \begin{pmatrix} l_1 \\ x_T \\ l_2 \\ x_0 \end{pmatrix} \tag{2.54}$$

and (A, B) *be exactly controllable in time* t_c. *Then there is* $c > 0$ *independent of* T, *such that*

$$\|\lambda(T)\|^2 \leq c \int_{T-t_c}^{T} \|B^*\lambda(s)\|_U^2 + \|Cx(s)\|_Y^2 + \|l_1(s)\|^2 \, ds.$$

Proof. The proof of this estimate is inspired by [113, Proof of Remark 2.1], where the finite dimensional case is considered. We decompose $\lambda = \lambda_1 + \lambda_2$, where

$$\begin{aligned} -\lambda_1' &= A^*\lambda_1, & \lambda_1(T) &= \lambda(T), \\ -\lambda_2' &= A^*\lambda_2 - C^*Cx + l_1, & \lambda_2(T) &= 0 \end{aligned} \quad ,$$

and apply the observability estimate (2.52) to $\lambda_1(t) = \mathcal{T}^*(T - t)\lambda(T)$, which yields

$$\alpha_{t_c}\|\lambda(T)\|^2 \leq \int_{T-t_c}^{T} \|B^*\lambda_1(s)\|_U^2 \, ds \leq \int_{T-t_c}^{T} \|B^*\lambda(s)\|_U^2 + \|B^*\lambda_2(s)\|_U^2 \, ds$$

and we conclude that

$$\begin{aligned} \int_{T-t_c}^{T} \|B^*\lambda_2(s)\|_U^2 \, ds &\leq \int_{T-t_c}^{T} \left\| B^* \int_s^T \mathcal{T}^*(\tau - s)(C^*Cx(\tau) + l_1(\tau)) \, d\tau \right\|_U^2 ds \\ &\leq c(t_c) \int_{T-t_c}^{T} \|Cx(s)\|_Y^2 + \|l_1(s)\|^2 \, ds. \end{aligned}$$

□

Similar to [113, Proof of Remark 2.1], it is important that we use integrals over time periods of length t_c, which yields the constants in the proof of Proposition 2.52, in particular α_{t_c} and $c(t_c)$, independent of T.

Having derived the desired estimate for the terminal state on the adjoint, we briefly comment on a possible extension to the unbounded case.

Remark 2.53. *Controllability or observability with an unbounded but admissible control resp. observation operator is discussed in, e.g., [34, Chapter 2] or [139, Chapter 6]. In that context,* (2.52) *is required for all* $x_0 \in D(A^*)$, *similarly to the admissibility inequality* (2.46), *cf. [34, Theorem 2.4.2] or [139, Definition 6.1.1]. In the above analysis, we observe that in the last estimate of the proof of Proposition 2.52, only an admissibility-like estimate of* B *is needed.*

Now, similarly to Theorem 2.38 and Theorem 2.47, we obtain a T-independent bound under an exact controllability assumption.

Theorem 2.54. *Let (A,B) be exactly controllable in time $0 < t_c \leq T$ and (A,C) be exponentially detectable. Then*

$$\|\widetilde{M}^{-1}\|_{L((L_2(0,T;X)\times X)^2,(C(0,T;X),\|\cdot\|_{2\wedge\infty})^2)} \leq c,$$

where $c \geq 0$ is independent of T.

Proof. We proceed analogously to Section 2.2.3. In order to estimate the unknown $\lambda(T)$ that appears on the right hand side of (2.42) and (2.44), we use the estimate obtained in Proposition 2.52. Thus, the statement (2.34) of Lemma 2.36 holds with an upper bound depending only on $\left(\|l_1\|_2^2 + \|l_2\|_{1\vee 2}^2 + \|x_0\|^2\right)$. Using Lemma 2.37 and the fact that $x(T) = x_T$ is a datum, a straightforward adaption of the proof of Theorem 2.38, where x_T plays the role of λ_T, yields the result. □

Completely analogously to Theorem 2.27, we can derive an estimate on the propagation of perturbations.

Theorem 2.55. *Let (A,B) be exactly controllable in time $0 < t_c \leq T$ and (A,C) be exponentially detectable. Let $(\varepsilon_1,\varepsilon_2) \in L_2(0,T;X)^2$ and $(\varepsilon_0,\varepsilon_T) \in X^2$. Assume $(\delta x,\delta\lambda) \in C(0,T;X)^2$ solve*

$$\begin{pmatrix} C^*C & -\frac{d}{dt} - A^* \\ E_T & 0 \\ \frac{d}{dt} - A & -BQ^{-1}B^* \\ E_0 & 0 \end{pmatrix} \begin{pmatrix} \delta x \\ \delta\lambda \end{pmatrix} = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_T \\ \varepsilon_2 \\ \varepsilon_0 \end{pmatrix}. \tag{2.55}$$

Then there is a scaling factor $\mu > 0$ satisfying

$$\mu < \frac{1}{\|\widetilde{M}^{-1}\|_{L((L_2(0,T;X)\times X)^2,L_2(0,T;X)^2)}}$$

and a constant $c \geq 0$, both independent of T, such that, defining

$$\rho := \left\|e^{-\mu t}\varepsilon_1(t)\right\|_{L_2(0,T;X)} + \left\|e^{-T}\varepsilon_T\right\| + \left\|e^{-\mu t}\varepsilon_2(t)\right\|_{L_2(0,T;X)} + \|\varepsilon_0\|,$$

we have

$$\begin{aligned} \left\|e^{-\mu t}\delta x(t)\right\|_{L_2(0,T;X)} + \left\|e^{-\mu t}\delta\lambda(t)\right\|_{L_2(0,T;X)} &\leq c\rho, \\ \left\|e^{-\mu t}\delta u(t)\right\|_{L_2(0,T;U)} &\leq c\rho \end{aligned} \tag{2.56}$$

and

$$\begin{aligned} \left\|e^{-\mu t}\delta x(t)\right\|_{C(0,T;X)} + \left\|e^{-\mu t}\delta\lambda(t)\right\|_{C(0,T;X)} &\leq c\rho, \\ \left\|e^{-\mu t}\delta u(t)\right\|_{L_\infty(0,T;U)} &\leq c\rho. \end{aligned} \tag{2.57}$$

Moreover, the corresponding counterpart of the turnpike result Theorem 2.30 in case of terminal conditions reads:

Theorem 2.56. *Let $(x,\lambda) \in C(0,T;X)^2$ solve (2.53) and $(\bar{x},\bar{\lambda}) \in X^2$ solve the corresponding steady state problem and set $u(t) = Q^{-1}B^*\lambda(t)$ for a.e. $t \in [0,T]$, $\bar{u} = Q^{-1}B^*\bar{\lambda}$ and $(\delta x, \delta u, \delta\lambda) := (x-\bar{x}, u-\bar{u}, \lambda-\bar{\lambda})$. Then, if (A,B) is exactly controllable and (A,C) is exponentially detectable, there exists a scaling factor $\mu > 0$ satisfying*

$$\mu < \frac{1}{\|\widetilde{M}^{-1}\|_{L((L_2(0,T;X)\times X)^2, L_2(0,T;X)^2)}}$$

and a constant $c \geq 0$, both independent of T, such that

$$\begin{aligned}
\left\|\frac{1}{e^{-\mu t}+e^{-\mu(T-t)}}\delta x(t)\right\|_{L_2(0,T;X)} &+ \left\|\frac{1}{e^{-\mu t}+e^{-\mu(T-t)}}\delta u(t)\right\|_{L_2(0,T;U)} \\
&+ \left\|\frac{1}{e^{-\mu t}+e^{-\mu(T-t)}}\delta \lambda(t)\right\|_{L_2(0,T;X)} \leq c\left(\|x_0-\bar{x}\|+\|x_T-\bar{x}\|\right), \\
\left\|\frac{1}{e^{-\mu t}+e^{-\mu(T-t)}}\delta x(t)\right\|_{C(0,T;X)} &+ \left\|\frac{1}{e^{-\mu t}+e^{-\mu(T-t)}}\delta u(t)\right\|_{L_\infty(0,T;U)} \\
&+ \left\|\frac{1}{e^{-\mu t}+e^{-\mu(T-t)}}\delta \lambda(t)\right\|_{C(0,T;X)} \leq c\left(\|x_0-\bar{x}\|+\|x_T-\bar{x}\|\right).
\end{aligned}$$

We conclude this part with a remark.

Remark 2.57. *Similarly to Section 2.3, a possible extension would be the case of an unbounded but admissible control operator, cf. Remark 2.53, analogously to Theorem 2.47* and *Theorem 2.48. As in Section 2.3, the price to pay will be the loss of the L_∞-estimate on the control in (2.57). Secondly, we only considered L_2-perturbations in this part for the sake of simplicity of exposition. All results, in particular Proposition 2.52 also hold with and L_1-norm on the perturbations. This can be proven using a straightforward adaption of the convolution inequality of Lemma 2.35 in the last estimate of the proof of Proposition 2.52. Note that here the semigroup is not assumed to be exponentially stable, however, the interval is bounded independently of T. This renders all constants independent of T.*

2.5 Examples

Finally, examples are provided to illustrate the sensitivity and turnpike results for bounded resp. unbounded control and observation with initial condition on the state, cf. Theorems 2.27, 2.30, 2.48 and 2.49, and for bounded control and observation with initial and terminal condition on the state, cf. Theorems 2.55 and 2.56. First, we consider the interior control of a heat equation and second, the boundary control of a wave equation.

Example 2.58. *(Interior control of an unstable heat equation) Let $\Omega \subset \mathbb{R}^n$, $n \in \{2,3\}$ be bounded, open and non-empty with smooth boundary and $\omega_c \subset \Omega$ be non-empty. For $T > 0$, we* *consider the heat equation*

$$\begin{aligned} \frac{\partial x}{\partial t} &= (\Delta + c^2 I)x + \chi_{\omega_c} u && \text{in } \Omega \times (0,T), \\ x &= 0 && \text{in } \partial\Omega \times (0,T), \\ x(0) &= x_0 && \text{in } \Omega, \end{aligned}$$

where χ_{ω_c} is the characteristic function of the control domain ω_c, $U = L_2(\omega_c)$ and $x_0 \in L_2(\Omega)$. Moreover, we consider an observation operator $C = \chi_{\omega_o}$ for non-empty $\omega_o \subset \Omega$. As Δ generates a strongly continuous semigroup on $X = L_2(\Omega)$ with domain $D(\Delta) = H^2(\Omega) \cap H_0^1(\Omega)$, $(\Delta + c^2 I)$ generates a strongly continuous semigroup with the same domain as Δ by classical perturbation results, cf. [44, Chapter III, Theorem 1.3]. If $c^2 > \lambda_1$, where λ_1 is the smallest eigenvalue of the negative Dirichlet Laplacian, the uncontrolled system is unstable. Defining $B\colon L_2(\omega_c) \to L_2(\Omega)$ via $Bu := \chi_{\omega_c} u$, the pair (Δ, B) is null controllable [12, 54] and hence exponentially stabilizable, cf. [154, Theorem 3.3]. Analogously, it follows that (Δ, C) is exponentially detectable. As $B \in L(U,X)$, one could apply the sensitivity result Theorem 2.27 or the turnpike result Theorem 2.30 to the prototypical optimal control problem Problem 2.23 governed by the operators defined above. Concerning the case of terminal constraints it is well known that, whenever $\omega_c \neq \Omega$, the system is not exactly controllable and the results of Section 2.4, in particular Theorems 2.55 and 2.56 can not be applied. This is due to the smoothing effect of the heat equation. For a discussion of this topic, the reader is referred to [159, Chapter 3].

Example 2.59 (Dirichlet control of a wave equation)**.** *Second, we provide an example of a hyperbolic PDE with unbounded but admissible control operator along the lines of [139, Section 10.9]. We consider the model of a vibrating membrane on $\Omega \subset \mathbb{R}^2$, where Ω is a bounded, non-empty and open domain with C^2-boundary. Further suppose that we can take action through Dirichlet boundary control on a part $\Gamma_c \subset \partial\Omega$ of the boundary. Moreover, we assume that (Ω, Γ_c, T) fulfills the* Geometric Control Condition (GCC)*, which ensures that all geometric optics have to enter the control domain in a time smaller than T. A consequence of this condition is exact controllability in time T, see [13, 117]. We consider the wave equation*

$$\begin{aligned} \frac{\partial^2 x}{\partial t^2} &= \Delta x && \text{in } \Omega \times (0,T), \\ x &= 0 && \text{in } \partial\Omega \setminus \Gamma_c \times (0,T), \\ x &= u && \text{in } \Gamma_c \times (0,T), \\ x(0) = f, \ \frac{\partial x}{\partial t}(0) &= g && \text{in } \Omega, \end{aligned}$$

where $f \in L_2(\Omega), g \in H^{-1}(\Omega)$ and $U := L_2(\Gamma_c)$. It was shown in [139, Proposition 10.9.1] that one can deduce a corresponding well-posed boundary control system on $X = L_2(\Omega) \times H^{-1}(\Omega)$ with generator $A = \begin{pmatrix} 0 & I \\ -A_0 & 0 \end{pmatrix}$, $D(A) = H_0^1(\Omega) \times L_2(\Omega)$, where A_0 is the Dirichlet Laplacian *and a control operator B defined by*

$$\begin{aligned} Bv &= \begin{pmatrix} 0 \\ A_0 Dv \end{pmatrix} && \forall v \in U = L_2(\Gamma_c), \\ B^* \begin{pmatrix} \varphi \\ \psi \end{pmatrix} &= -\frac{\partial}{\partial \nu}(A_0^{-1}\psi)_{|\Gamma_c} && \forall (\varphi, \psi) \in D(A), \end{aligned}$$

where D is the Dirichlet map and $\frac{\partial}{\partial \nu}$ the outward normal derivative. The reader is referred to [139, Section 10.6, Section 10.9] for details. In particular, the operator B is an admissible control operator for the semigroup generated by A. Moreover, as the GCC is satisfied, the pair (A, B) is exponentially stabilizable. If we now consider any observation region $\Omega_o \subset \Omega$ such that (Ω, Ω_o, T) satisfy the GCC, then the pair (A, C) is exponentially detectable. As in this example the control operator is unbounded on X, one cannot apply Theorem 2.27. However, Theorem 2.48 and Theorem 2.30 are applicable to the optimal control problems governed by this equation. In order to consider the case of a terminal condition on the state, we replace the boundary control by a distributed control on a subset $\omega_c \subset \Omega$, analogously to Example 2.58. Assuming that (Ω, ω_c, T) satisfy the GCC-condition, we obtain that the system is exactly controllable by distributed control on ω_c, see [159, Chapter 3]. Hence, we can impose a terminal condition on the state and the sensitivity and turnpike results of Theorems 2.55 and 2.56 apply.

2.6 The particular case of a parabolic equation

In the previous chapters, all estimates were posed in terms of integral or pointwise norms in X, i.e., the function space that the initial conditions belongs to. This is due to the fact that higher regularity can not be expected in this very general setting; in particular, only continuity in time with values in X is ensured by the definition of the semigroup, cf. Definition 2.1. In this chapter, we will present improved estimates by assuming more regularity of the solutions via *analyticity* of the underlying semigroup. The corresponding equations are often called *parabolic*, including the prominent example of the heat equation. For an in-depth treatment of analytic semigroups, the interested reader is referred to, e.g., [44, Section II.4.a], [109, Section 2.5] and [19, Part II-1, Section 2.7]. This class of semigroups shares favorable properties, i.e., e.g., that solutions satisfy the differential equation in an a.e. (temporal) sense in X. Recall that for general semigroups, differentiability of solutions in time is only given in a very weak sense in $D(A)^*$, cf. Definition 2.17. Before we head to the formal definition of analytic semigroups, we present a straightforward approach to refine the estimates of Theorem 2.38 and Theorem 2.47 a posteriori by bootstrapping arguments. We will discuss the connection to analytic semigroups in Remark 2.62.

2.6.1 Sharper estimates via direct bootstrapping

The bootstrapping analysis presented in this part uses the parabolic variational theory, cf. [49, 127, 138, 150]. We will thoroughly discuss this setting for non-autonomous systems in

Chapter 3. In this part, in order to keep the presentation concise, we will not delve into details and refer the interested reader to Section 3.1 or the literature cited above. Assume, there is an additional, more regular reflexive Banach space V with the continuous dense embedding $V \hookrightarrow X$ such that A can be extended to an operator $\bar{A} \in L(V, V^*)$ and we have a Gårding inequality for $x \in L_2(0,T;V)$:

$$\exists\, \omega \in \mathbb{R}, \alpha > 0: \quad \alpha \|x\|^2_{L_2(0,T;V)} \leq \int_0^T -\langle \bar{A}x(t), x(t)\rangle_{V^* \times V}\, dt + \omega \|x\|^2_{L_2(0,T;X)}. \tag{2.58}$$

By classical variational theory, cf. the references above, the state and adjoint equation yield the regularities $x, \lambda \in L_2(0,T;V)$, $x', \lambda' \in L_2(0,T;V^*)$ and the state and adjoint equation in (2.27) are satisfied in an $L_2(0,T;V^*)$-sense. Testing the second row of the optimality system (2.27) with x, using integration by parts on $\langle x', x\rangle_{V^*\times V}$, (2.45), and assuming boundedness of $\|B^*\|_{L(V,U)}$ (which is weaker than boundedness of $\|B^*\|_{L(X,U)}$), we compute that

$$\begin{aligned}\int_0^T -\langle \bar{A}x(t), x(t)\rangle_{V^*\times V}\, dt &= \int_0^T -\langle x'(t), x(t)\rangle_{V^*\times V} + \langle l_2(t) + B(R^*R)^{-1}B^*\lambda(t), x(t)\rangle\, dt \\ &\leq \frac{1}{2}\left(\|x(0)\|^2 - \|x(T)\|^2\right) + \|l_2\|_{1\vee 2}\|x\|_{2\wedge\infty} + \|R^{-*}B^*\lambda\|_{L_2(0,T;U)}\|R^{-*}B^*x\|_{L_2(0,T;U)} \\ &\leq \frac{1}{2}\|x_0\|^2 + \|r\|_{1\vee 2}\|z\|_{2\wedge\infty} + c(\|r\|_{1\vee 2} + \|z\|_{2\wedge\infty})\|R^{-*}B^*\|_{L(V,U)}\|x\|_{L_2(0,T;V)},\end{aligned}$$

where $z = (x, \lambda)$ and $r = (l_1, \lambda_T, l_2, x_0)$. By Theorem 2.27 we may bound $\|z\|_{2\wedge\infty}$ by $\|r\|_{1\vee 2}$ and substitute the result into (2.58). A short computation yields $\|x\|_{L_2(0,T;V)} \leq c\|r\|_{1\vee 2}$ and similarly $\|\lambda\|_{L_2(0,T;V)} \leq c\|r\|_{1\vee 2}$. Hence, there is a T-independent bound $c > 0$ such that

$$\|M^{-1}\|_{L((L_1(0,T;X), \|\cdot\|_{1\vee 2} \times X)^2, L_2(0,T;V)^2)} \leq c.$$

This is a refined version of Theorem 2.38, as $\|\cdot\|_V$ usually is stronger than $\|\cdot\|_X$, for example V could be a Sobolev space, whereas X is an L_2-space. By further bootstrapping via

$$x' = \bar{A}x + B(R^*R)^{-1}B^*\lambda + l_2 \qquad \text{in } L_2(0,T;V^*),$$

we obtain $\|x'\|_{L_2(0,T;V^*)} \leq c(\|r\|_{1\vee 2} + \|l_2\|_{L_2(0,T;X)})$ and similarly an estimate for $\|\lambda'\|_{L_2(0,T;V^*)}$. Thus, also for the parabolic space $W([0,T]) := \{v\colon [0,T] \to V \,|\, v \in L_2(0,T;V), v' \in L_2(0,T;V^*)\}$ equipped with the norm $\|v\|_{W([0,T])} = \|v\|_{L_2(0,T;V)} + \|v'\|_{L_2(0,T;V^*)}$, we get the T-independent bound $c > 0$ such that

$$\|M^{-1}\|_{L((L_2(0,T;X)\times X)^2, W([0,T])^2)} \leq c. \tag{2.59}$$

These additional estimates can be used to obtain results in Theorem 2.27 and Theorem 2.48 and also in Theorem 2.30 and Theorem 2.49 in stronger norms. However, despite the equation being well-posed for $l_1, l_2 \in L_2(0,T;V^*)$, the arguments presented here only allow for right hand sides bounded in $L_2(0,T;X)$, see (2.59). This means that we do not obtain stability estimates in the strongest possible norms in view of mere well-posedness. In order to get rid of this assumption on the perturbations, we will fully exploit the regularity theory of parabolic equations, or, in other words, the smoothing properties of analytic semigroups in the following part.

2.6.2 Sharper estimates via maximal parabolic regularity

As stated in the introduction of this part, analytic semigroups represent a very important subclass of strongly continuous semigroups. For example, the Laplace operator on $L_2(\Omega)$ gives rise to such an analytic semigroup. For further reading, the interested reader is referred to the respective parts of the monographs [19, Part II, Chapter 1], [133, Section 3.3] and [44, Section II.4.a]. The solutions described by analytic semigroups share very favorable properties, e.g., if the initial value lies in some interpolation space between $D(A)$ and X and if right-hand sides belong to $L_2(0,T;X)$, the solution lies in $D(A)$ for almost every point in time and a time derivative exists in $L_2(0,T;X)$, allowing the differential equation

$$x' = Ax + f$$

to be understood in an $L_2(0,T;X)$-sense. Note that compared to the previous subsection, $L_2(0,T;X)$ takes the role of $L_2(0,T;V^*)$.

We first define analytic semigroups and follow the presentation in [109, Section 2.5]. Note that there are many equivalent ways to define analytic semigroups, cf. [139, Definition 5.4.5], [19, Part II-1, Theorem 2.11] or [44, Definition 4.5].

Definition 2.60 ([109, Definition 5.1]). *Let $\varphi_1, \varphi_2 \in \mathbb{R}$ such that $\varphi_1 < 0 < \varphi_2$ and consider the sector $S := \{z \in \mathbb{C} \,|\, \varphi_1 < \arg z < \varphi_2\}$. Let $\mathcal{T}(z)$ be a bounded linear operator for all $z \in S$. Then we call $(\mathcal{T}(z))_{z\in S}$ an analytic semigroup if*

i) $z \mapsto \mathcal{T}(z)$ is analytic in S,

ii) $\mathcal{T}(0) = I$ and $\lim_{z\to 0, z\in S} \mathcal{T}(z)x = x$, for all $x \in X$,

iii) $\mathcal{T}(z_1 + z_2) = \mathcal{T}(z_1)\mathcal{T}(z_2)$ for all $z_1, z_2 \in S$.

We observe that ii) and iii) are very similar to the properties of a strongly continuous semigroup, cf. Definition 2.1. However analytic semigroups can be evaluated for elements of a sector around the positive real axis and the map $z \mapsto \mathcal{T}(z)$ is analytic. It can be shown that analytic semigroups are generated by sectorial operators, as stated in the following theorem. As common in the literature, we will assume $0 \in \rho(A)$, which can always be achieved by scaling the semigroup with $e^{-(\omega_0(A)+\varepsilon)}$, where $\omega_0(A)$ is the type of the semigroup and $\varepsilon > 0$, cf. the discussion in [109, p.61].

Theorem 2.61 ([109, Theorem 5.2]). *Let $(\mathcal{T}(t))_{t\geq 0}$ be a strongly continuous semigroup with $\|\mathcal{T}(t)\| \leq 1$, the operator A be its generator and $0 \in \rho(A)$. Then the following statements are equivalent:*

i) $(\mathcal{T}(t))_{t\geq 0}$ can be extended to an analytic semigroup on the sector $S_\delta := \{z \in \mathbb{C} \,|\, |\arg z| < \delta\}$ and $\|\mathcal{T}(z)\|$ is uniformly bounded on $\overline{S}_{\delta'}$ for all $\delta' < \delta$.

ii) There is $0 < \delta < \frac{\pi}{2}$ *and* $M > 0$ *such that*

$$\Sigma := \{0\} \cup \left\{ \lambda \in \mathbb{C} \,|\, |\arg \lambda| < \frac{\pi}{2} + \delta \right\} \subset \rho(A)$$

and

$$\|R(\lambda, A)\| \leq \frac{M}{|\lambda|} \qquad \text{for } \lambda \in \Sigma,\ \lambda \neq 0.$$

iii) $(\mathcal{T}(t))_{t\geq 0}$ *is differentiable for all* $t > 0$ *and there is* $C \geq 0$ *such that*

$$\|A\mathcal{T}(t)\| \leq \frac{C}{t} \qquad \text{for } t > 0.$$

Remark 2.62. *Any operator that satisfies the Gårding inequality (2.58) generates an analytic semigroup, cf. [19, Part II-1, Theorem 2.12] or [133, Section 3.6]. In particular, the Laplacian with Dirichlet or Neumann boundary conditions generates an analytic semigroup in, e.g.,* $L_2(\Omega)$*.*

After having defined analytic semigroups and obtaining a characterization of their generators, we would like to fully exploit maximal parabolic regularity to derive fine stability estimates.

Definition 2.63. *Let* $A : D(A) \subset X \to X$ *be the generator of a semigroup and denote*

$$W^{1,2}(0, T, D(A), X) := \{v \in L_2(0, T; D(A)) \,|\, v' \in L_2(0, T; X)\},$$
$$\|v\|_{W^{1,2}(0,T;D(A),X)} := \|v\|_{L_2(0,T;D(A))} + \|v'\|_{L_2(0,T;X)}.$$

Note that if A generates an exponentially stable analytic semigroup, then A is an isomorphism from $D(A)$ onto X and the norm $\|Ax\|$ is equivalent to the graph norm $\|x\| + \|Ax\|$ on $D(A)$, cf. [19, Part II-1, Section 3.6.2]. Moreover, by [19, Part II-1, Remark 4.2], we have the T-independent embedding $W^{1,2}(0, T, D(A), X) \hookrightarrow C(0, T; (D(A), X)_{\frac{1}{2}})$, where $(D(A), X)_{\frac{1}{2}}$ denotes a real interpolation space as defined in [19, Part II-1, Section 4.3]. Note that in the Hilbert space setting and as $D(A) \subset X$, the complex and real interpolation spaces with exponent 2 coincide, [19, Part II-1, Section 4.7] and [137, Remark 3 and Remark 4]. Interpolation spaces are a very involved subject and for the sake of clarity of presentation, we will not define them here. The interested reader is referred to the literature cited above, in particular to the monographs [137] and [152]. As we will see later, interpolation spaces between $D(A)$ and X can, in some cases, be shown to be isomorphic to domains of fractional powers of A, cf. Example 2.71 and [90, Section 0.2.1].

Lemma 2.64. *Let* A_{cl} *generate an exponentially stable analytic semigroup on* X *and* $x_0 \in (D(A_{cl}), X)_{\frac{1}{2}}$ *and* $f \in L_2(0, T; X)$*. Then, if* $x \in C(0, T; X)$ *is the mild solution of*

$$x' = A_{cl}x + f, \qquad x(0) = x_0,$$

we have the improved regularity $x \in W^{1,2}(0, T, D(A_{cl}), X)$*. Moreover, for* $c \geq 0$ *independent of* T*, we have*

$$\|x\|_{W^{1,2}(0,T,D(A_{cl}),X)} \leq c \left(\|f\|_{L_2(0,T;X)} + \|x_0\|_{(D(A_{cl}),X)_{\frac{1}{2}}} \right).$$

Proof. See [19, Part II-1, Theorem 3.1]. □

The main assumption in this part is the following.

Assumption 2.65 (Standing assumptions)**.** *Let X_1 be the space introduced in the beginning of Section 2.3.1, i.e. $D(A)$ equipped with the norm $\|\cdot\|_1 := \|(\beta I - A)\cdot\|$ for $\beta \in \rho(A)$. Moreover, assume that*

i) A generates an analytic semigroup on X,

ii) $B \in L(U, X)$, $C \in L(X, Y)$,

iii) (A, B) is exponentially stabilizable and (A, C) is exponentially detectable.

Proposition 2.66. *Let (A, B, C) satisfy Assumption 2.65 and consider K_C and K_B to be stabilizing feedback operators. Then*

i) $A + K_C C$ and $A^* + K_B^* B^*$ generate analytic semigroups on X.

ii) The graph norms of $A + K_C C$ and A, resp. $A^* + K_B^* B^*$ and A^*, are equivalent.

Proof. Boundedness of B, C and the feedback operators immediately implies A-*boundedness* of the perturbations, cf. [44, Chapter III, Lemma 2.16]. This in turn ensures the analyticity of the perturbed semigroup, cf. [44, Chapter III, Theorem 2.10]. For ii), see [44, Chapter III, Lemma 2.4]. □

Theorem 2.67. *Let Assumption 2.65 hold and $(\delta x, \delta\lambda)$ solve the system*

$$\begin{pmatrix} C^*C & -\frac{d}{dt} - A^* \\ 0 & E_T \\ \frac{d}{dt} - A & -BQ^{-1}B^* \\ E_0 & 0 \end{pmatrix} \begin{pmatrix} x \\ \lambda \end{pmatrix} = \begin{pmatrix} l_1 \\ \lambda_T \\ l_2 \\ x_0 \end{pmatrix}.$$

where $x_0 \in (D(A), X)_{\frac{1}{2}}$, $\lambda_T \in (D(A^*), X)_{\frac{1}{2}}$ and $l_1, l_2 \in L_2(0, T; X)$. Then

$$\|x\|^2_{W^{1,2}(0,T,D(A),X)} \le c\left(\|Cx\|^2_{L_2(0,T;Y)} + \|R^{-*}B^*\lambda\|^2_{L_2(0,T;U)} + \|x_0\|^2_{(D(A),X)_{\frac{1}{2}}} + \|l_2\|^2_{L_2(0,T;X)}\right),$$

$$\|\lambda\|^2_{W^{1,2}(0,T,D(A^*),X)} \le c\left(\|Cx\|^2_{L_2(0,T;Y)} + \|R^{-*}B^*\lambda\|^2_{L_2(0,T;U)} + \|\lambda_T\|^2_{(D(A^*),X)_{\frac{1}{2}}} + \|l_1\|^2_{L_2(0,T;X)}\right),$$

Proof. We add the exponentially stabilizing feedback and get

$$x' = (A + K_C C)x - K_C Cx - BQ^{-1}B^*\lambda + l_2$$

and, using Proposition 2.66 and exponential stability, we conclude

$$\begin{aligned} &\|x\|_{W^{1,2}(0,T,D(A+K_CC),X)} \\ &\quad \le c\left(\|Cx\|^2_{L_2(0,T;Y)} + \|R^{-*}B^*\lambda\|^2_{L_2(0,T;U)} + \|x_0\|^2_{(D(A+K_CC),X)_{\frac{1}{2}}} + \|l_2\|^2_{L_2(0;T;X)}\right). \end{aligned}$$

By the equivalence of the graph norms of A and $A + K_C C$ we obtain

$$\|x\|_{W^{1,2}(0,T,D(A),X)} \leq c \left(\|Cx\|^2_{L_2(0,T;Y)} + \|R^{-*}B^*\lambda\|^2_{L_2(0,T;U)} + \|x_0\|^2_{(D(A),X)_{\frac{1}{2}}} + \|l_2\|^2_{L_2(0;T;X)} \right).$$

Proceeding analogously for the adjoint equation, we conclude the result. □

Theorem 2.68. *Let Assumption 2.65 hold. Then, there is a constant $c \geq 0$ independent of T such that*

$$\|M^{-1}\|_{L\left(\left(L_2(0,T;X)\times(D(A^*),X)_{\frac{1}{2}}\times L_2(0,T;X)\times(D(A),X)_{\frac{1}{2}}\right),(W^{1,2}(0,T,D(A),X)\times W^{1,2}(0,T,D(A^*),X))\right)} \leq c.$$

Proof. Analogously to the case of a general semigroup, Theorem 2.67 together with Lemma 2.37 yields the result. □

Thus, a perturbation result and turnpike result are now a direct consequence by straightforward adaptation of the proof of Theorem 2.48 and Theorem 2.30.

Corollary 2.69. *Let Assumption 2.65 hold. Then there are constants $c, \mu > 0$, independent of T, such that the perturbation estimate (2.10) of Theorem 2.27 for the case $E = L_2(0,T;X)$ and the turnpike result (2.21) of Theorem 2.30 still hold when replacing the $L_2(0,T;X)$-norms on the left hand side of the estimate with the (stronger) $W^{1,2}(0,T,D(A),X)$ resp. $W^{1,2}(0,T,D(A^*),X)$-norm and the X-norm on the initial and terminal state by the $(D(A),X)_{\frac{1}{2}}$ resp. $(D(A^*),X)_{\frac{1}{2}}$-norm.*

We briefly discuss an extension to unbounded control or observation.

Remark 2.70. *The notion of A-boundedness allows for perturbations of the semigroup by unbounded operators, cf. [44, Chapter III]. In case of, e.g., an unbounded observation operator, $A + K_C C$ still generates an analytic semigroup if $K_C C$ is compact as linear operator from $D(A)$ to X.*

2.6.3 Example of heat equation revisited

We will now recall the example of a heat equation and show, how the refined analysis of Section 2.6.1 and Section 2.6.2 leads to sharper estimates.

Example 2.71 (Example 2.58 revisited)**.** *Consider the system*

$$\begin{aligned} \frac{\partial x}{\partial t} &= (\Delta + c^2 I)x + \chi_{\omega_c} u && \text{in } \Omega \times (0,T), \\ x &= 0 && \text{in } \partial\Omega \times (0,T), \\ x(0) &= x_0 && \text{in } \Omega, \end{aligned}$$

where $c \in \mathbb{R}$, χ_{ω_c} is the characteristic function of the control domain $\omega_c \subset \Omega$, $U = L_2(\omega_c)$ and $x_0 \in L_2(\Omega)$. Moreover, we set the observation operator $C = \chi_{\omega_o}$ for non-empty $\omega_o \subset \Omega$. Not using the parabolic structure of the equation, the results of Section 2.2 would yield $L_2(0,T;L_2(\Omega))$- or $C(0,T;L_2(\Omega))$-estimates on state and adjoint for perturbations in $L_2(0,T;L_2(\Omega))$ and initial values in $L_2(\Omega)$. However, these estimates can be improved.

Application of direct bootstrapping (Section 2.6.1):

Choosing $V = H_0^1(\Omega)$, $X = L_2(\Omega)$, *the method presented in Section 2.6.1 yields* $W([0,T])$ *sensitivity and turnpike estimates for perturbations in* $L_2(0,T;L_2(\Omega))$ *and initial values in* $L_2(\Omega)$.

Application of maximal parabolic regularity (Section 2.6.2):

We set $X = L_2(\Omega)$ *and hence* $D(\Delta) = H^2(\Omega) \cap H_0^1(\Omega)$. *In this case,* $(D(\Delta), X)_{1/2} \cong D(-\Delta_\omega^{\frac{1}{2}}) \overset{(*)}{\cong} H^1(\Omega)$, *where* ω *is large enough such that* $\Delta_\omega = \Delta - \omega I$ *generates a semigroup with negative type and* $\cdot^{\frac{1}{2}}$ *denotes a fractional power of a positive operator, cf. [90, Section 0.2.1] or [19, Part II-1, Section 1.4]. For the last relation* $(*)$, *see [139, Section 10.7]. Thus, Corollary 2.69 yields* $W^{1,2}(0,T,H^2(\Omega) \cap H_0^1(\Omega), X)$*-estimates for state and adjoint for perturbations in* $L_2(0,T;L_2(\Omega))$ *if* $x_0 \in H^1(\Omega)$.

2.7 Outlook

We briefly present several extensions of the approach presented in this chapter.

- We assumed the state, the input and the output space to be Hilbert spaces. A natural question that arises is if a generalization to, e.g., reflexive Banach spaces is possible. To this end, the reader is also referred to Section 4.4, where we derive $L_{p_1}(0,T;L_{p_2}(\Omega))$-estimates with $2 < p_1, p_2$ for parabolic equations. Considering the semigroup on other L_p-spaces can be useful to allow for control and observation that are not admissible for the semigroup on, e.g., $L_2(\Omega)$ but on higher order $L_p(\Omega)$ spaces, cf. the heat equation with Dirichlet boundary control, cf. [129, Section 4 4].

- In Assumption 2.45 we assumed that either the control operator B or the observation operator C is unbounded. It would be interesting to allow for a fully unbounded but admissible setting, i.e., control, observation and corresponding feedback operators to be unbounded with a suitable well-posedness assumption.

- Another open problem is the rigorous derivation of the optimality conditions in case of an unbounded control operator. An inspection of the proof of Theorem 2.24 shows that the critical point is to interchange the application of B^* and integration in time.

Chapter 3

Sensitivity analysis for linear quadratic optimal control of non-autonomous parabolic equations

We now analyze how the results of the previous chapter carry over to the case of non-autonomous parabolic equations in a variational setting. To this end, we aim to quantify how perturbations of the right hand side of the extremal equations influence the solution if the underlying spatial differential operator or the cost function is time-dependent. We recall that our main motivation to consider perturbations is to estimate the influence of discretization errors, i.e., the perturbations may represent the residual of a discretization scheme. Similar to the autonomous setting considered in Chapter 2, we show that perturbations that increase exponentially in time only influence the initial part of the solution negligibly. As indicated in the previous chapter, this feature can be used to construct very efficient discretization schemes for MPC. Again, we stress that locality of discretization errors is a priori unclear, as the backwards-in-time adjoint equation could propagate perturbations from close to the end time T to the initial part. As in the autonomous setting of the previous chapter, we split the proof into an abstract scaling result provided in Section 3.2.1 and the derivation of T-independent bounds on the extremal equations' solution operator's norm in Section 3.2.3. Additionally, we will provide a turnpike result under the assumption that the problem is autonomous and relate the result to those obtained in Chapter 2. Finally, we provide numerical examples and evaluate the performance of a priori space and time grids that are specialized for MPC.

Although there exists an extension of the semigroup concept to non-autonomous Cauchy problems—so called *evolution families*—their analysis in a general setting is rather involved. The interested reader is referred to the respective parts in the monographs [44, Section VI.9], [109, Chapter 5] or [133, Chapter 5]. Additionally, for non-autonomous equations, stability results are difficult to establish (even in the finite dimensional case) due to the lack of a characterization of stability via the spectrum of the (time-dependent) generator, cf. [151]. To this end, we consider a particular notion of stability defined via an uniform ellipticity condition. Despite being stronger than classical exponential stability as introduced in Definition 2.12, this particular notion has several advantages and is particularly well suited for non-autonomous parabolic equations: First and foremost, it straightforwardly allows for stability estimates in the context of non-autonomous equations and considerably facilitates the proof of such. Second, this particular notion of stability allows us to derive estimates in Sobolev norms with boundary control and observation. Last, this stabilizability assumption can very easily be verified as we will illustrate by means of several examples.

Non-autonomous optimal control problems are an interesting subject in their own right, however, their analysis will be particularly useful when we approach nonlinear optimal control problems in Chapter 4, where we will linearize the nonlinear extremal equations around a time-dependent trajectory, which directly leads to a non-autonomous system.

Structure. In Section 3.1, we introduce the function spaces involved, the weak time derivative, and the resulting formulation of the PDE. Moreover, we define a linear quadratic optimal control problem with dynamics governed by a parabolic PDE and derive optimality conditions. Section 3.2 contains two central results in view of MPC. Under the assumption of a T-independent bound on the extremal equations' solution operator, we obtain an estimate in Theorem 3.14 that proves exponential decay of perturbations for non-autonomous optimal control problems. Under the same assumption and assuming that the system is autonomous, we will draw a link to the previous chapter and present an exponential turnpike result in Theorem 3.16. In Corollary 3.30, we prove that under a particular stabilizability assumption, the extremal equations' solution operator can indeed be bounded independently of T.

The majority of the results in this chapter have been published in [69] and [70].

3.1 Setting and preliminaries

We will briefly introduce the generalized time derivative, the formulation of the parabolic equations and the linear quadratic optimal control problem of interest. To this end, we recall some fundamental results on variational parabolic equations from the literature, cf. [49, 55, 96, 150, 158].

3.1.1 Gelfand triples and generalized time derivatives

Let $(V, \|\cdot\|_V)$ be a separable and reflexive Banach space and H a separable and real Hilbert space with scalar product $\langle \cdot, \cdot \rangle$ and induced norm $\|\cdot\|$. Further, assume that $V \hookrightarrow H$ continuously and densely. Thus, $V \hookrightarrow H \cong H^* \hookrightarrow V^*$ continuously and densely, where V^* is the topological dual of V. Such an ensemble of spaces is often called Gelfand triple or evolution triple. In the following, we will identify H with its dual via the Riesz isomorphism.

Definition 3.1 (Generalized time derivative, [158, Definition 23.15])**.** *Let Y and Z be Banach spaces. Consider $x \in L_1(0, T; Y)$ and $w \in L_1(0, T; Z)$. Then w is called the generalized time* *derivative of x on $(0,T)$ if*

$$\int_0^T \varphi'(t)x(t)\,dt = -\int_0^T \varphi(t)w(t)\,dt \qquad \forall\,\varphi \in C_0^\infty(0,T), \tag{3.1}$$

where $C_0^\infty(0,T) = \{\varphi \in C^\infty(0,T)\,|\,\varphi(0) = \varphi(T) = 0\}$.

In general, the left and right hand side of (3.1) lie in different spaces, i.e., Y and Z, respectively and validity of the formula implies that both lie in $Y \cap Z$. However, by density of the embeddings in the Gelfand triple, one obtains the following characterization.

Lemma 3.2 (Characterization of the generalized time derivative, [158, Proposition 23.20])**.** *Let $V \hookrightarrow H \hookrightarrow V^*$ form a Gelfand triple and $\frac{1}{p} + \frac{1}{q} = 1$. Then, for any $x \in L_p(0,T;V)$, the function $w \in L_q(0,T;V^*)$ is the generalized time derivative of x if and only if*

$$\int_0^T \varphi'(t)\langle x(t), v\rangle\,dt = -\int_0^T \varphi(t)\langle w(t), v\rangle_{V^*\times V}\,dt \qquad \forall v \in V,\ \varphi \in C_0^\infty(0,T).$$

In this case, the generalized time derivative w is denoted $\frac{d}{dt}x$ or x'.

We further will need a product rule for generalized time derivatives.

Lemma 3.3 (Product rule for generalized time derivatives)**.** *Let $x \in L_2(0,T;V)$ with generalized time derivative $x' \in L_2(0,T;V^*)$ and $s \in C^\infty(0,T)$. Then,*

$$(sx)' = s'x + sx'.$$

Proof. The proof follows directly from the defining equation (3.1) and the product rule for functions in $C^\infty(0,T)$. ☐

We now define the space of functions in $L_2(0,T;V)$ with generalized time derivative in $L_2(0,T;V^*)$ and recall well-known properties. From now on, we will assume that $V \hookrightarrow H \hookrightarrow V^*$ form a Gelfand triple.

Lemma 3.4 (Solution space and important properties)**.** *Define the function space*

$$W([0,T]) := \{v\colon [0,T] \to V\,|\,v \in L_2(0,T;V), v' \in L_2(0,T;V^*)\}$$

endowed with the norm $\|v\|_{W([0,T])} := \|v\|_{L_2(0,T;V)} + \|v'\|_{L_2(0,T;V^*)}$. Then,

i) $W([0,T]) \hookrightarrow C(0,T;H)$ continuously with embedding constant independent of T.

ii) For $v, w \in W([0,T])$ and $0 \le s \le t \le T$, we have the integration by parts formula

$$\langle v(t), w(t)\rangle - \langle v(s), w(s)\rangle = \int_s^t \langle v'(\tau), w(\tau)\rangle_{V^*\times V} + \langle w'(\tau), v(\tau)\rangle_{V^*\times V}\,d\tau.$$

iii) $\left(W([0,T]), \|\cdot\|_{W([0,T])}\right)$ *is a Banach space.*

iv) For $w \in W([0,T])$*, it holds that*

$$\langle w', w\rangle_{L_2(0,T;V^*)\times L_2(0,T;V)} = \frac{1}{2}\left(\|w(T)\|^2 - \|w(0)\|^2\right).$$

Proof. See [158, Proposition 23.23, Problem 23.10d] and [127, Section 2.3]. □

3.1.2 Parabolic PDEs in variational form

In this part, we introduce a solution concept for parabolic PDEs in variational form. To this end, after having defined the generalized time derivative, we will consider a spatial differential operator A satisfying the following assumptions.

Assumption 3.5.

i) $A \in L(L_2(0,T;V), L_2(0,T;V^*))$.

ii) A is local in time, i.e., for any $s\colon \mathbb{R} \to \mathbb{R}$ *and* $x \in L_2(0,T;V)$*, we have* $A(sx) = sAx$.

iii) A satisfies the Gårding inequality:

$$\exists \omega \in \mathbb{R},\ \alpha > 0: \quad \alpha\|x\|^2_{L_2(0,T;V)} \le -\langle Ax, x\rangle_{L_2(0,T;V^*)\times L_2(0,T;V)} + \omega\|x\|^2_{L_2(0,T;H)}. \tag{3.2}$$

The inequality (3.2) also occurred in the bootstrapping arguments of the previous chapter, particularly Section 2.6.1. We briefly give an example of an operator satisfying these assumptions.

Example 3.6. *Consider* $\kappa(t,\omega)\colon [0,T]\times\Omega \to \mathbb{R}$*, continuous and uniformly bounded from below in both arguments, i.e.,* $v\cdot\kappa(t,\omega)v \ge \alpha|v|^2$ *for* $\alpha > 0$*. Moreover, let* $V = H^1_0(\Omega)$ *or* $V = H^1(\Omega)$ *and* $H = L_2(\Omega)$*. Then, for any constant* $c \in \mathbb{R}$*, the linear operator* $A\colon L_2(0,T;V) \to L_2(0,T;V^*)$ *defined by*

$$\langle Ax_1, x_2\rangle_{L_2(0,T;V^*)\times L_2(0,T;V)} := \int_0^T \int_\Omega -\nabla x_1(t,\omega)\cdot\kappa(t,\omega)\nabla x_2(t,\omega) + c x_1(t,\omega)x_2(t,\omega)\, d\omega\, dt$$

for $x_1, x_2 \in L_2(0,T;V)$ *satisfies Assumption 3.5.*

Using the time evaluation operator $E_t x = x(t)$ for $t \in [0,T]$ and $x \in C(0,T;H)$ as introduced in Definition 2.26, we define an operator corresponding to a parabolic PDE in weak form via

$$\Lambda := \begin{pmatrix} \frac{d}{dt} - A \\ E_0 \end{pmatrix} : W([0,T]) \to (L_2(0,T;V)\times H)^*,$$

where for $x \in W([0,T])$ and test functions $(v, v_0) \in L_2(0,T;V) \times H$

$$\langle \Lambda x, (v, v_0)\rangle_{(L_2(0,T;V)\times H)^* \times (L_2(0,T;V)\times H)} := \left\langle \left(\frac{d}{dt} + A\right) x, v \right\rangle_{L_2(0,T;V^*)\times L_2(0,T;V)} + \langle x(0), v_0\rangle. \tag{3.3}$$

A parabolic problem in variational form is to find $x \in W([0,T])$ such that

$$\langle \Lambda x, (v, v_0)\rangle_{(L_2(0,T;V)\times H)^* \times (L_2(0,T;V)\times H)} = \langle f, v\rangle_{L_2(0,T;V^*)\times L_2(0,T;V)} + \langle x_0, v_0\rangle \tag{3.4}$$

for all $(v, v_0) \in L_2(0,T;V) \times H$, where $f \in L_2(0,T;V^*)$ is a source term and $x_0 \in H$ an initial datum. Whenever we call a function a solution to a variational parabolic problem, we mean it in the sense of (3.4).

Remark 3.7. *Solvability of problems of type (3.4) is a classical issue and the interested reader is referred to [49, 55, 96, 127, 150, 158]. If A satisfies Assumption 3.5, it can be shown that $\Lambda\colon W([0,T]) \to L_2(0,T;V^*) \times H$ is an isomorphism, cf. [127, Theorem 3.4]. Additionally, if $x \in W([0,T])$ solves (3.4), then the terms x', Ax and f share the same temporal and spatial regularity, i.e., they belong to $L_2(0,T;V^*)$. This feature is known as maximal parabolic regularity, cf. Section 2.6.2 where we discussed this topic in a semigroup framework. For hyperbolic equations, where A in particular does not satisfy (3.2) due to skew-adjointness, the corresponding operator Λ can not be shown to be an isomorphism due to lack of surjectivity, cf. [19, Part II-1, Remark 3.5].*

The following proposition is a central result in view of optimal control and characterizes the regularity of solutions of adjoint equations with particular right hand sides.

Proposition 3.8 (Higher regularity of adjoint state, [127, Proposition 3.8])**.** *Let $(\lambda, \lambda_0) \in L_2(0,T;V) \times H$ be given. Then, the following assertions are equivalent:*

i) There exist $(l, l_0, l_T) \in L_2(0,T;V^*) \times H \times H$ such that for all $w \in W([0,T])$ it holds that

$$\langle \Lambda^*(\lambda, \lambda_0), w\rangle_{W([0,T])^* \times W([0,T])} = \int_0^T \langle l(t), w(t)\rangle_{V^*\times V}\, dt + \langle l_0, w(0)\rangle + \langle l_T, w(T)\rangle.$$

ii) $\lambda \in W([0,T])$.

If these conditions are satisfied, then $\lambda(T) = l_T$ and $\lambda_0 - \lambda(0) = l_0$.

Therefore, for right hand sides of the adjoint equation in $L_2(0,T;V^*) \times H$, we obtain an adjoint state $\lambda \in W([0,T])$ with a prescribed terminal value. As a consequence, the rule of integration by parts holds for the adjoint state and the adjoint equation can be interpreted as a backwards-in-time equation, cf. [127, Section 3.2].

3.1.3 Optimization problems with parabolic PDEs

In this section, we will move to optimization problems governed by parabolic PDEs. In contrast to the previous chapter, where we considered autonomous optimal control problems governed by general semigroups in Problem 2.23, we will consider the following non-autonomous optimal control problem governed by a parabolic PDE. In order to formulate the problem of interest, we will consider the following standing assumptions.

Assumption 3.9.

i) $T > 0$ *is a fixed time horizon,*

ii) A *satisfies Assumption 3.5,* $f \in L_2(0,T;V^*)$ *and* $x_0 \in H$*,*

iii) $u_d \in L_2(0,T;U)$*,* U *is a real Hilbert space with scalar product* $\langle \cdot,\cdot \rangle_U$ *and induced norm* $\|\cdot\|_U$*,* $B \in L(L_2(0,T;U), L_2(0,T;V^*))$*,*

iv) $R \in L(L_2(0,T;U))$ *with* $\|Ru\|^2_{L_2(0,T;U)} \geq \alpha \|u\|^2_{L_2(0,T;U)}$ *for* $\alpha > 0$ *and all* $u \in L_2(0,T;U)$*,*

v) $x_d \in L_2(0,T;V)$*,* Y *is a real Hilbert space with scalar product* $\langle \cdot,\cdot \rangle_Y$ *and induced norm* $\|\cdot\|_Y$*,* $C \in L(L_2(0,T;V), L_2(0,T;Y))$*,*

vi) C, B, R *are local in time in the sense of Assumption 3.5 ii),*

vii) $\|A\|_{L(L_2(0,T;V),L_2(0,T;V^*))}$*,* $\|B\|_{L(L_2(0,T;U),L_2(0,T;V^*))}$*,* $\|C\|_{L(L_2(0,T;V),L_2(0,T;Y))}$*, and* $\|R\|_{L(L_2(0,T;U))}$ *can be bounded independently of* T*.*

With these assumptions at hand, we aim to analyze solutions to the following optimal control problem.

Problem 3.10. *Find* $(x,u) \in W([0,T]) \times L_2(0,T;U)$ *solving*

$$\begin{aligned} \min_{(x,u)} \frac{1}{2} \int_0^T \|C(x(t) - x_d(t))\|_Y^2 + \|R(u(t) - u_d(t))\|_U^2 \, dt \\ \text{s.t. } x' = Ax + Bu + f, \\ x(0) = x_0. \end{aligned} \tag{3.5}$$

Remark 3.11. *We will briefly comment on this problem and Assumption 3.9.*

- *In the case* $V = H^1(\Omega)$*, the above setting naturally incorporates the case of boundary control or observation, i.e.,* $\langle Bu, v\rangle_{L_2(0,T;V^*)\times L_2(0,T;V)} := \int_0^T \int_{\partial\Omega} u \operatorname{tr} v \, d\gamma \, dt$ *and* $Cx := \operatorname{tr} x$ *with* $Y = U = L_2(\partial\Omega)$*, where* $\operatorname{tr}\colon H^1(\Omega) \to L_2(\partial\Omega)$ *is the Dirichlet trace operator.*

- *Assumption 3.9 vi) is merely needed for the scaling approach applied in, e.g., Theorem 2.27, to permute the application of the operator and the multiplication with a scalar scaling function. In particular, existence and uniqueness of solutions can be deduced for operators* B*,* C*, and* R *that are non-local time, cf. [127].*

- *Assumption 3.9 vii) is crucial to derive T-independent bounds. It is, however, not too restrictive in terms of applications. Two examples of operators $L\colon L_2(0,T;V) \to L_2(0,T;V*)$ bounded independently of T are*
  - $\langle Lv, w\rangle_{L_2(0,T;V^*)\times L_2(0,T;V)} = \int_0^T \langle \underline{L}v, w\rangle_{V^*\times V}\, dt$, *where* $\underline{L} \in L(V,V^*)$, *as clearly* $\|L\|_{L(L_2(0,T;V),L(L_2(0,T;V^*))} \leq \|\underline{L}\|_{L(V,V^*)}$,
  - $\langle Lv, w\rangle_{L_2(0,T;V^*)\times L_2(0,T;V)} = \int_0^T \langle \underline{L}(t)v, w\rangle_{V^*\times V}\, dt$, *where* $\underline{L}\colon [0,\infty[\to L(V,V^*)$ *is continuous and* $\sup_{t\in[0,\infty[} \|\underline{L}(t)\|_{L(V,V^*)} < \infty$.

The analysis in this non-autonomous parabolic setting, again, is based on the characterization of minimizers via the first-order necessary optimality conditions. Due to convexity of the problem, these conditions are also sufficient. The following proposition states the optimality conditions at a minimizer.

Proposition 3.12 (First order optimality conditions)**.** *Let $(x,u) \in W([0,T]) \times L_2(0,T;U)$ be a minimizer of problem Problem 3.10. Then there is an adjoint state $\lambda \in W([0,T])$ such that*

$$\begin{aligned}
C^*Cx - \left(\frac{d}{dt} + A^*\right)\lambda &= C^*Cx_d && \text{in } L_2(0,T;V^*),\\
\lambda(T) &= 0 && \text{in } H,\\
R^*Ru - B^*\lambda &= R^*Ru_d && \text{in } L_2(0,T;U),\\
\left(\frac{d}{dt} - A\right)x - Bu &= f && \text{in } L_2(0,T;V^*),\\
x(0) &= x_0 && \text{in } H.
\end{aligned}$$

Proof. See [127, Theorem 1.1, Remark 1.2]. □

Defining $Q := R^*R$ and using coercivity of R, we can eliminate the control via $u = Q^{-1}B^*\lambda + u_\mathrm{d}$ and obtain that the optimal state and corresponding adjoint satisfy

$$\underbrace{\begin{pmatrix} C^*C & -\frac{d}{dt} - A^* \\ 0 & E_T \\ \frac{d}{dt} - A & -BQ^{-1}B^* \\ E_0 & 0 \end{pmatrix}}_{=:M} \begin{pmatrix} x \\ \lambda \end{pmatrix} = \begin{pmatrix} C^*Cx_\mathrm{d} \\ 0 \\ Bu_\mathrm{d} + f \\ x_0 \end{pmatrix}, \tag{3.6}$$

where E_t is time evaluation operator defined by $E_t x = x(t)$ for $t \in [0,T]$ in Definition 2.26 and

$$M\colon W([0,T])^2 \to (L_2(0,T;V^*) \times H)^2.$$

In the following, we will refer to (3.6) as the extremal equations. We first formulate a scaling result analogous to Lemma 2.22, where we proposed a similar lemma in a semigroup framework.

Lemma 3.13. *Let A satisfy Assumption 3.5, $x_0, x_T \in H$ and $f_1, f_2 \in L_2(0,T;V^*)$. Assume that $x_1, x_2 \in W([0,T])$ solve*

$$\begin{aligned} x_1' &= Ax_1 + f_1, & x_1(0) &= x_0, \\ -x_2' &= A^*x_2 + f_2, & x_2(T) &= x_T. \end{aligned}$$

Then, the following holds.

i) For any $\mu \in \mathbb{R}$, the operators $A - \mu I$ and $A + \mu I$ satisfy Assumption 3.5. Moreover, $\widetilde{x}_1(t) := e^{-\mu t}x_1(t)$ and $\widetilde{x}_2(t) := e^{-\mu t}x_2(t)$ solve

$$\begin{aligned} \widetilde{x}_1' &= (A - \mu I)\widetilde{x}_1 + e^{-\mu t} f_1, & \widetilde{x}_1(0) &= x_0, \\ -\widetilde{x}_2' &= (A + \mu I)^*\widetilde{x}_2 + e^{-\mu t} f_2, & \widetilde{x}_2(T) &= e^{-\mu T} x_T. \end{aligned} \tag{3.7}$$

ii) For all $0 \le s \le t \le T$ it holds that

$$\langle x_1(t), x_2(t)\rangle - \langle x_1(s), x_2(s)\rangle = \int_s^t \langle f_1(\tau), x_2(\tau)\rangle_{V^*\times V} - \langle f_2(\tau), x_1(\tau)\rangle_{V^*\times V}\, d\tau.$$

Proof. It is easily checked that $A - \mu I$ and $A + \mu I$ satisfy Assumption 3.5. The second part of i) immediately follows by the product rule for generalized time derivatives, cf. Lemma 3.3. Note that locality of A in time as defined in Assumption 3.5 ii) is important here in order to permute application of A and multiplication by the scaling term $e^{-\mu t}$. The formula in ii) results from testing the first equation with x_2, the second equation with x_1, subtracting both equations and integrating by parts in time in the sense of Lemma 3.4. □

3.2 Exponential sensitivity analysis

This section constitutes the main part of this chapter. We first present two abstract scaling results under the assumption of a T-independent bound on the solution operator of (3.6). After that, we derive the desired bound on the solution operator under a particular stabilizability assumption.

3.2.1 An abstract exponential sensitivity result

We will refer to the solution $(x, \lambda) \in W([0,T])^2$ of (3.6) as the exact solution. Assume that there is a second pair of variables $(\tilde{x}, \tilde{\lambda}) \in W([0,T])^2$ solving the perturbed system

$$M \begin{pmatrix} \tilde{x} \\ \tilde{\lambda} \end{pmatrix} = \begin{pmatrix} C^*Cx_{\mathrm{d}} \\ 0 \\ Bu_{\mathrm{d}} + f \\ x_0 \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_T \\ \varepsilon_2 \\ \varepsilon_0 \end{pmatrix}$$

for $\varepsilon_1, \varepsilon_2 \in L_2(0,T;V^*)$ and $\varepsilon_0, \varepsilon_T \in H$. This solution will be referred to as the perturbed solution. The terms $\varepsilon_1, \varepsilon_2 \in L_2(0,T;V^*)$ are perturbations of the state and adjoint equation that could stem from discretization errors in time or space, whereas $\varepsilon_0, \varepsilon_T \in H$ describe a perturbation of the initial and terminal condition by space discretization errors. In this subsection, we will give an estimate on the absolute error, i.e., the difference of $(\tilde{x}, \tilde{\lambda})$ and (x, λ). It follows by linearity that this difference $\big(\delta x, \delta\lambda\big) := \big(\tilde{x} - x, \tilde{\lambda} - \lambda\big)$ between exact and perturbed solution satisfies the system of equations

$$M \begin{pmatrix} \delta x \\ \delta\lambda \end{pmatrix} = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_T \\ \varepsilon_2 \\ \varepsilon_0 \end{pmatrix}. \tag{3.8}$$

Analogous to the autonomous counterpart analyzed in Section 2.2, the main question here is the following: How does the behavior of the perturbations ε_1 and ε_2 over time and ε_0 resp. ε_T influence the behavior of the error δx and $\delta\lambda$? To answer this question, we adapt the scaling approach introduced in Theorem 2.27 to the parabolic variational setting.

Theorem 3.14. *Assume that $(\delta x, \delta\lambda) \in W([0,T])^2$ solves (3.8), where $\varepsilon_1, \varepsilon_2 \in L_2(0,T;V^*)$ and $\varepsilon_0, \varepsilon_T \in H$. Let $\delta u = Q^{-1}B^*\delta\lambda$. Suppose that $\|M^{-1}\|_{L((L_2(0,T;V^*)\times H)^2, W([0,T])^2)}$ can be bounded independently of T. Then, for any scaling factor $\mu > 0$ satisfying*

$$\mu < \frac{1}{\|M^{-1}\|_{L((L_2(0,T;V^*)\times H)^2, W([0,T])^2)}}$$

there is a constant $c \geq 0$ independent of T such that defining

$$\rho := \left\|e^{-\mu t}\varepsilon_1(t)\right\|_{L_2(0,T;V^*)} + \left\|e^{-T}\varepsilon_T\right\| + \left\|e^{-\mu t}\varepsilon_2(t)\right\|_{L_2(0,T;V^*)} + \|\varepsilon_0\|$$

we have the estimate

$$\left\|e^{-\mu t}\delta x(t)\right\|_{W([0,T])} + \left\|e^{-\mu t}\delta u(t)\right\|_{L_2(0,T;U)} + \left\|e^{-\mu t}\delta\lambda(t)\right\|_{W([0,T])} \leq c\rho.$$

Proof. We proceed completely analogously to the L_2-case in the proof of Theorem 2.27. We define scaled variables $\widetilde{\delta x} := e^{-\mu t}\delta x \in W([0,T])$ and $\widetilde{\delta\lambda} := e^{-\mu t}\delta\lambda \in W([0,T])$ and conclude with Lemma 3.13 that $\left(\widetilde{\delta x}, \widetilde{\delta\lambda}\right)$ solves

$$\left(\begin{pmatrix} C^*C & -\frac{d}{dt} - A^* \\ 0 & E_T \\ \frac{d}{dt} - A & -BQ^{-1}B^* \\ E_0 & 0 \end{pmatrix} + \mu \underbrace{\begin{pmatrix} 0 & -I \\ 0 & 0 \\ I & 0 \\ 0 & 0 \end{pmatrix}}_{=:P} \right) \begin{pmatrix} \widetilde{\delta x} \\ \widetilde{\delta\lambda} \end{pmatrix} = \begin{pmatrix} e^{-\mu t}\varepsilon_1 \\ e^{-T}\varepsilon_T \\ e^{-\mu t}\varepsilon_2 \\ \varepsilon_0 \end{pmatrix}.$$

Setting $\tilde{z} := \left(\widetilde{\delta x}, \widetilde{\delta \lambda}\right)$ and $\tilde{\varepsilon} := \left(e^{-\mu t}\varepsilon_1, e^{-\mu T}, e^{-\mu t}\varepsilon_2, \varepsilon_0\right)$ we compute

$$\begin{aligned} (M + \mu P)\tilde{z} &= \tilde{\varepsilon} && \text{in } (L_2(0,T;V^*) \times H)^2, \\ (I + \mu M^{-1}P)\tilde{z} &= M^{-1}\tilde{\varepsilon} && \text{in } W([0,T])^2. \end{aligned}$$

Clearly we have the bound $\|P\|_{L(W([0,T])^2,(L_2(0,T;V^*)\times H)^2)} \leq 1$. Thus, by choosing $\mu > 0$ such that $\beta := \mu\|M^{-1}\|_{L((L_2(0,T;V^*)\times H)^2,W([0,T])^2)} < 1$, we get invertibility of $(I + \mu M^{-1}P)$ as an operator from $W([0,T])^2$ to $W([0,T])^2$ by a standard Neumann argument, cf. [85, Theorem 2.14]. Moreover, the Neumann series representation of $(I + \mu M^{-1}P)^{-1}$ yields

$$\|(I + \mu M^{-1}P)^{-1}\|_{L(W([0,T])^2,W([0,T])^2)} \leq \sum_{i=0}^{\infty} \|(\mu M^{-1}P)^i\|_{L(W([0,T])^2,W([0,T])^2)} \leq \sum_{i=0}^{\infty} \beta^i = \frac{1}{1-\beta}.$$

Hence, we conclude

$$\tilde{z} = (I - \mu M^{-1}P)^{-1}M^{-1}\tilde{\varepsilon},$$

which implies the estimate

$$\|\tilde{z}\|_{W([0,T])^2} \leq \frac{\|M^{-1}\|_{L((L_2(0,T;V^*)\times H)^2,W([0,T])^2)}}{1-\beta}\|\tilde{\varepsilon}\|_{(L_2(0,T;V^*)\times H)^2} \leq c\rho.$$

Writing this in the original variables yields the result. For the control, we compute

$$\begin{aligned} \left\|e^{-\mu t}\delta u\right\|_{L_2(0,T;U)} &= \left\|e^{-\mu t}R^{-1}B^*\delta\lambda\right\|_{L_2(0,T;U)} = \left\|R^{-1}B^*e^{-\mu t}\delta\lambda\right\|_{L_2(0,T;U)} \\ &\leq \|R^{-1}B^*\|_{L(L_2(0,T;V),L_2(0,T;U))}c\rho, \end{aligned}$$

where we used that R and B are local in time, i.e., application of the operators and multiplication with the scaling term commute. □

Corollary 3.15. *Let the assumptions of Theorem 3.14 hold. Then there exist $\mu, c > 0$ independent of T such that*

$$\left\|e^{-\mu t}\delta x(t)\right\|_{C(0,T;H)} + \left\|e^{-\mu t}\delta\lambda(t)\right\|_{C(0,T;H)} \leq c\rho.$$

If additionally $B \in L(L_2(0,T;U), L_2(0,T;H))$ with $\langle Bu, v\rangle_{L_2(0,T;V^*)\times L_2(0,T;V)} = \int_0^T \langle \underline{B}(t)u(t), v(t)\rangle$ for $\underline{B} \in L_\infty(0,T;L(U,H))$ and $\langle Ru, v\rangle_{L_2(0,T;U)} = \int_0^T \langle \underline{R}(t)u(t), v(t)\rangle_U$ for $\underline{R} \in L_\infty(0,T;L(U))$ continuously invertible, and if $\underline{B}$ and $\underline{R}$ are bounded independently of T, then we have

$$\left\|e^{-\mu t}\delta u\right\|_{L_\infty(0,T;U)} \leq c\rho$$

with a constant $c \geq 0$ independent of T.

Proof. The bound on state and adjoint state follows from $W([0,T]) \hookrightarrow C(0,T;H)$, cf. Lemma 3.4. By the assumptions on B and R, we have

$$e^{-\mu t}\delta u(t) = \underline{R}^{-1}(t)\underline{B}(t)^* e^{-\mu T}\delta\lambda(t)$$

for a.e. $t \in [0,T]$. Hence, the pointwise bound on the control follows by the pointwise bound on the adjoint state. □

In Theorem 3.14 and Corollary 3.15, we assumed a T-independent bound on M^{-1}. In Section 3.2.3 we will derive this bound under a stabilizability assumption. Before that, however, we briefly consider the case of an autonomous system to deduce a turnpike result similar to Theorem 2.30.

3.2.2 An exponential turnpike result

In this section, in order to define a steady state problem corresponding to Problem 3.10 we will restrict ourselves to an autonomous version of Problem 3.10, where the involved operators are induced by time-independent ones, e.g., $\langle Ax, v\rangle_{L_2(0,T;V^*)\times L_2(0,T;V)} = \int_0^T \langle \bar{A}x(t), v(t)\rangle_{V^*\times V}\, dt$, for $\bar{A} \in L(V,V^*)$. Similarly, we assume B to be given by a time-independent $\bar{B}$, C by $\bar{C}$ and R by $\bar{R}$. Moreover, we assume constant references $x_\mathrm{d} \equiv \bar{x}_\mathrm{d} \in V$ and $u_\mathrm{d} \equiv \bar{u}_\mathrm{d} \in U$ and a constant source term $f \equiv \bar{f} \in V^*$. The first pair of variables $(x,\lambda) \in W([0,T])^2$ we consider is the solution of the extremal equations (3.6). Secondly, we introduce the solution of a steady state optimization problem, namely $(\bar{x},\bar{u}) \in V \times U$ being the minimizer of

$$\min_{\bar{x},\bar{u}} \ \frac{1}{2}\|\bar{C}(\bar{x}-\bar{x}_\mathrm{d})\|_Y^2 + \|\bar{R}(\bar{u}-\bar{u}_\mathrm{d})\|_U^2$$
$$\text{s.t.} -\bar{A}\bar{x} - \bar{B}\bar{u} = \bar{f}.$$

This problem has a unique solution if, e.g., A is continuously invertible. This, for example can be ensured by classical elliptic theory if (3.2) holds with $\omega = 0$. Then, there is an adjoint state $\bar{\lambda} \in V$ such that $(\bar{x},\bar{\lambda})$ is a solution of the corresponding first-order conditions

$$\begin{pmatrix} \bar{C}^*\bar{C} & -\bar{A}^* \\ -\bar{A} & -\bar{B}\bar{Q}^{-1}\bar{B}^* \end{pmatrix} \begin{pmatrix} \bar{x} \\ \bar{\lambda} \end{pmatrix} = \begin{pmatrix} \bar{C}^*\bar{C}x_\mathrm{d} \\ \bar{B}u_\mathrm{d} + \bar{f} \end{pmatrix}, \tag{3.9}$$

where $\bar{Q} := \bar{R}^*\bar{R}$ and $\bar{u} = Q^{-1}B^*\lambda + \bar{u}_\mathrm{d}$. Similar to Theorem 2.30, we present the following turnpike result.

Theorem 3.16. *Let (x,u,λ) solve Problem 3.10 and let $(\bar{x},\bar{u},\bar{\lambda})$ solve the corresponding steady state problem (3.9). Assume $\|M^{-1}\|_{L((L_2(0,T;V^*)\times H)^2, W([0,T])^2)}$ can be bounded independently of T and set $(\delta x, \delta u, \delta\lambda) := (x-\bar{x}, u-\bar{u}, \lambda-\bar{\lambda})$. Then, for any $\mu > 0$ satisfying*

$$\mu < \frac{1}{\|M^{-1}\|_{L((L_2(0,T;V^*)\times H)^2, W([0,T])^2)}},$$

there exists a constant $c \geq 0$ independent of T such that

$$\begin{aligned}\left\|\frac{1}{e^{-\mu t}+e^{-\mu(T-t)}}\delta x(t)\right\|_{W([0,T])} + \left\|\frac{1}{e^{-\mu t}+e^{-\mu(T-t)}}\delta u(t)\right\|_{L_2(0,T;U)} & \\ + \left\|\frac{1}{e^{-\mu t}+e^{-\mu(T-t)}}\delta \lambda(t)\right\|_{W([0,T])} &\leq c\left(\|x_0-\bar{x}\| + \|\bar{\lambda}\|\right), \end{aligned} \tag{3.10}$$

$$\left\|\frac{1}{e^{-\mu t}+e^{-\mu(T-t)}}\delta x(t)\right\|_{C(0,T;H)} + \left\|\frac{1}{e^{-\mu t}e^{-\mu(T-t)}}\delta \lambda(t)\right\|_{C(0,T;H)} \leq c\left(\|x_0-\bar{x}\| + \|\bar{\lambda}\|\,|\right). \tag{3.11}$$

Proof. First, we integrate (3.9) over $[0,T]$, add $\frac{d}{dt}\bar{x} = \frac{d}{dt}\bar{\lambda} = 0$, and $E_t\bar{x} = \bar{x}$, $E_t\bar{\lambda} = \bar{\lambda}$ and conclude that $(\delta x, \delta\lambda) := (x-\bar{x}, \lambda - \bar{\lambda})$ solves

$$\begin{pmatrix} C^*C & -\frac{d}{dt} - A^* \\ 0 & E_T \\ \frac{d}{dt} - A & -BQ^{-1}B^* \\ E_0 & 0 \end{pmatrix} \begin{pmatrix} \delta x \\ \delta\lambda \end{pmatrix} = \begin{pmatrix} 0 \\ -\bar{\lambda} \\ 0 \\ x_0 - \bar{x} \end{pmatrix}$$

We introduce scaled variables $\widetilde{\delta x} := \frac{1}{(e^{-\mu t}+e^{-\mu(T-t)})}\delta x$ and $\widetilde{\delta \lambda} := \frac{1}{(e^{-\mu t}+e^{-\mu(T-t)})}\delta \lambda$. Then, by the product rule for generalized time derivatives, cf. Lemma 3.3, we obtain

$$\left(\begin{pmatrix} C^*C & \frac{d}{dt} - A^* \\ 0 & E_T \\ \frac{d}{dt} - A & -BQ^{-1}B^* \end{pmatrix} - \mu \underbrace{\begin{pmatrix} 0 & F \\ 0 & 0 \\ -F & 0 \\ 0 & 0 \end{pmatrix}}_{=:P}\right) \begin{pmatrix} \widetilde{\delta x} \\ \widetilde{\delta\lambda} \end{pmatrix} = \begin{pmatrix} 0 \\ -\frac{1}{1+e^{-\mu T}}\bar{\lambda} \\ 0 \\ \frac{1}{1+e^{-\mu T}}(x_0 - \bar{x}) \end{pmatrix},$$

where $\langle Fv, w\rangle_{V^*\times V} = \int_0^T \frac{(e^{-\mu(T-t)}-e^{-\mu t})}{(e^{-\mu t}+e^{-\mu(T-t)})}\langle v(t), w(t)\rangle_{V^*\times V}\, dt$. Defining $\tilde{z} := \left(\widetilde{\delta x}, \widetilde{\delta\lambda}\right)$ and $\tilde{l} := \frac{1}{1+e^{-\mu T}}\left(0, -\frac{1}{1+e^{-\mu T}}\bar{\lambda}, 0, \frac{1}{1+e^{-\mu T}}(x_0-\bar{x})\right)$, we compute that

$$(M - \mu P)\tilde{z} = \tilde{l}$$

if and only if

$$(I - \mu M^{-1}P)z = M^{-1}\tilde{l}. \tag{3.12}$$

We proceed to show the bound $\|P\|_{L(W([0,T])^2,(L_2(0,T;V^*)\times H)^2)} \leq 1$. This follows immediately by the definition of P and the estimate

$$|\langle Fv, w\rangle_{V^*\times V}| = \left|\int_0^T \underbrace{\frac{(e^{-\mu(T-t)}-e^{-\mu t})}{(e^{-\mu t}+e^{-\mu(T-t)})}}_{|\cdot|<1}\langle v(t), w(t)\rangle_{V^*\times V}\, dt\right| \leq \|v\|_{W([0,T])}\|w\|_{L_2(0,T;V)}.$$

Hence, we choose $\mu > 0$ such that $\beta = \mu\|M^{-1}\|_{L((L_2(0,T;V^*)\times H)^2,W([0,T])^2)} < 1$ and obtain invertibility of $(I - \mu M^{-1}P)$ by [85, Theorem 2.14]. Using the Neumann series representation of $(I - \mu M^{-1}P)^{-1}$ yields

$$\|(I - \mu M^{-1}P)^{-1}\|_{L(W([0,T])^2,W([0,T])^2)} \leq \sum_{i=0}^{\infty} \|(\mu M^{-1}P)^i\|_{L(W([0,T])^2,W([0,T])^2)} \leq \sum_{i=0}^{\infty} \beta^i = \frac{1}{1-\beta},$$

and together with (3.12), we get

$$\|\tilde{z}\|_{W([0,T])^2} \leq \frac{\|M^{-1}\|_{L((L_2(0,T;V^*)\times H)^2,W([0,T])^2)}}{1-\beta}\|\tilde{l}\|_{(L_2(0,T;V^*)\times H)^2} \leq c\left(\|\bar{\lambda}\| + \|x_0 - \bar{x}\|\right).$$

Writing this in the original variables yields the result. The L_2-estimate for the control follows immediately analogous to the respective part of the proof of Theorem 3.14. The pointwise bounds can be proven completely analogously to the proof of Corollary 3.15.

□

We immediately obtain the following corollary, where the bound for the control follows analogously to the proof of Corollary 3.15.

Corollary 3.17. *Let the assumptions of Theorem 3.16 hold. Then there exist $c, \mu > 0$ independent of T such that for all $t \in [0, T]$*

$$\|x(t) - \bar{x}\| + \left\|\lambda(t) - \bar{\lambda}\right\| \leq c(e^{-\mu T} + e^{-\mu(T-t)})\left(\|x_0 - \bar{x}\| + \left\|\bar{\lambda}\right\|\right).$$

If additionally $\bar{B} \in L(U, H)$, we get following the pointwise bound on the control

$$\|u(t) - \bar{u}\|_U \leq c(e^{-\mu T} + e^{-\mu(T-t)})\left(\|x_0 - \bar{x}\| + \left\|\bar{\lambda}\right\|\right).$$

We will now show that the T-independent bounds on $\|M^{-1}\|_{L((L_2(0,T;V^*)\times H)^2,W([0,T])^2)}$ assumed in Theorem 3.14 and Theorem 3.16 can be proven under suitable stabilizability assumptions.

3.2.3 T-independent bounds for the solution operator

By $c > 0$ we will denote a generic T-independent constant, which will be redefined as necessary in the proofs.

In order to derive the desired bound on the solution operator, let $l_1, l_2 \in L_2(0, T; V^*)$ and $x_0, \lambda_T \in H$, and consider the system

$$\begin{pmatrix} C^*C & -\frac{d}{dt} - A \\ 0 & E_T \\ \frac{d}{dt} + A & -BQ^{-1}B^* \\ E_0 & 0 \end{pmatrix} \begin{pmatrix} x \\ \lambda \end{pmatrix} = \begin{pmatrix} l_1 \\ \lambda_T \\ l_2 \\ x_0 \end{pmatrix}. \tag{3.13}$$

In the following, we will bound (x, λ) by means of $(l_1, \lambda_T, l_2, x_0)$ uniformly in T. In the autonomous case, we assumed exponential stabilizability and exponential detectability to derive the desired bound, cf. Assumption 2.32. In this part, we introduce a particular notion of stability characterized by an ellipticity condition, which is especially well-suited for non-autonomous parabolic equations.

Definition 3.18. *An operator $S\colon L_2(0,T;V) \to L_2(0,T;V^*)$ is called $L_2(0,T;V)$-elliptic if there exists $\alpha > 0$ independent of T such that*

$$\langle Sv, v\rangle_{L_2(0,T;V^*)\times L_2(0,T;V)} \geq \alpha\|v\|^2_{L_2(0,T;V)} \qquad \forall v \in L_2(0,T;V). \tag{3.14}$$

Remark 3.19. *We briefly comment on $L_2(0,T;V)$-ellipticity in the context of evolution equations.*

- *It is clear that S is $L_2(0,T;V)$-elliptic if and only if S^* is.*
- *If an operator S satisfies Assumption 3.5 and the $L_2(0,T;V)$-ellipticity condition (3.14) for $\alpha > 0$, then it can be shown by a simple scaling argument, cf. Lemma 3.13, that the solution of*

$$v' = -Sv, \qquad v(0) = v_0$$

  with $v_0 \in H$ fulfills $\left\|e^{\mu t}v\right\|_{L_2(0,T;V)} \leq \frac{1}{\sqrt{\alpha-\mu}}\|v_0\|$ for any $\mu < \alpha$ and also $\|v(t)\| \leq e^{-\alpha t}\|v(0)\|$ for $t \geq 0$. Thus, if $\|\cdot\|$ represents an energy, the latter estimate yields immediate energy dissipation.
- *An example of an $L_2(0,T;H^1_0(\Omega))$-elliptic operator is the Laplacian in weak form, i.e.,*

$$\langle Sv, v\rangle_{L_2(0,T;H^{-1}(\Omega))\times L_2(0,T;H^1_0(\Omega))} := \int_0^T \int_\Omega \nabla v \cdot \kappa(t,\omega)\nabla v \, d\omega dt,$$

  where $\kappa(t,\omega)$ is a uniformly bounded measurable function from $[0,T]$ into the set of real matrices, satisfying the uniform ellipticity condition $v \cdot \kappa(t,\omega)v \geq \alpha|v|^2$ for $\alpha > 0$.

The exponential estimates in the previous remark motivate the following definition of V-exponential stabilizability.

Definition 3.20. *Let A, B, C be defined as in Assumption 3.9. We call (A, B) V-exponentially stabilizable if there exists a feedback operator $K_B \in L(L_2(0,T;V), L_2(0,T;U))$ such that $-(A + BK_B)$ is $L_2(0,T;V)$-elliptic, i.e., fulfills (3.14). Similarly, we call (A, C) V-exponentially stabilizable if there exists $K_C \in L(L_2(0,T;Y), L_2(0,T;V^*))$ such that $-(A + K_C C)$ is $L_2(0,T;V)$-elliptic, i.e., fulfills (3.14).*

We briefly illustrate this stabilizability property by means of two examples of an unstable heat equation with distributed control and Neumann boundary control, respectively.

Example 3.21. *Let $\Omega \subset \mathbb{R}^n$, $n \in \{2,3\}$, be a non-empty, open, bounded domain with smooth boundary and set $V = H^1(\Omega)$ and $H = L_2(\Omega)$. For $\gamma > 0$ and $x, v \in L_2(0,T;H^1(\Omega))$ let A be defined by*

$$\langle Ax, v\rangle_{L_2(0,T;H^1(\Omega)^*)\times L_2(0,T;H^1(\Omega))} := \int_0^T \int_\Omega -\nabla x \cdot \kappa(t,\omega)\nabla v + \gamma x v \, d\omega dt,$$

where $\kappa(t,\omega)$ is a uniformly bounded measurable function from $[0,T]$ into the set of real matrices, satisfying the uniform ellipticity condition $v \cdot \kappa(t,\omega)v \geq \alpha|v|^2$ for $\alpha > 0$. First, we observe that for small $c > 0$ (depending on the uniform ellipticity constant of κ) we have the lower bound

$$-\langle Ax, x\rangle_{L_2(0,T;(H^1(\Omega))^*)\times L_2(0,T;H^1(\Omega))} + (\gamma + c)\|x\|^2_{L_2(0,T;L_2(\Omega))} \geq c\|x\|^2_{L_2(0,T;H^1(\Omega))}$$

and hence A satisfies Assumption 3.5. However, for any $\gamma > 0$, the solutions of

$$x' - Ax = 0, \qquad x(0) = x_0$$

are not stable. This can easily be seen by inserting a spatially constant function into the PDE.

To stabilize the system $H^1(\Omega)$-exponentially, we consider a subset $\Omega_c \subset \Omega$ with positive measure and set $U = L_2(\Omega_c)$. The control operator will be given by

$$\langle Bu, v\rangle_{L_2(0,T;H^1(\Omega)^*)\times L_2(0,T;H^1(\Omega))} := \int_0^T \int_{\Omega_c} uv \, d\omega dt.$$

Defining the feedback operator $K_B\colon L_2(0,T;H^1(\Omega)) \to L_2(0,T;L_2(\Omega_c))$ by $K_B x(t) := -Kx(t)_{|\Omega_c}$ for $K > 0$ and a.e. $t \in [0,T]$, we conclude

$$\begin{aligned}-\langle (A + BK_B)x, x\rangle_{L_2(0,T;H^1(\Omega)^*)\times L_2(0,T;H^1(\Omega))} &\geq \int_0^T \int_\Omega c\|\nabla x\|^2 - \gamma x^2 + \chi_{\Omega_c} K x^2 \, d\omega dt \\ &\geq C(\gamma, c, K, \Omega)\|x\|^2_{L_2(0,T;H^1(\Omega))}.\end{aligned}$$

For given (c, Ω, Ω_c), positivity of $C(\gamma, c, K, \Omega)$ can be ensured if $\gamma > 0$ is small enough. This follows by the generalized Poincaré inequality, cf. [138, Lemma 2.5]. If $\Omega_c = \Omega$, we note that by choosing $K > \gamma$, the feedback operator K_B defined above is $H^1(\Omega)$-exponentially stabilizing for every $\gamma > 0$. If we assume an observation on the whole domain, i.e., C is the embedding of $L_2(0,T;H^1(\Omega))$ into $L_2(0,T;L_2(\Omega))$ and $Y = L_2(\Omega)$, we can choose $K_C = -KE$ where $K > \gamma$ and E is the embedding of $L_2(0,T;L_2(\Omega))$ into $L_2(0,T;(H^1(\Omega)^*))$, which yields the $L_2(0,T;H^1(\Omega))$-ellipticity of $-(A + K_C C)$ for all $\gamma > 0$.

Example 3.22. *A similar result holds if we replace the distributed control in Example 3.21 by Neumann boundary control of the form $\frac{\partial}{\partial \nu}x = u$ on a subset $\Gamma_c \subset \partial\Omega$ of positive measure. In this case, $V = H^1(\Omega)$, $H = L_2(\Omega)$ and $\langle Bu, v\rangle_{L_2(0,T;V^*)\times L_2(0,T;V)} := \int_0^T \int_{\Gamma_c} u \operatorname{tr}(v) \, ds dt$ where $\operatorname{tr}\colon H^1(\Omega) \to L_2(\partial\Omega)$ is the Dirichlet trace operator. A stabilizing feedback operator K_B can be defined via $K_B x = -K \operatorname{tr} x$ for $K > 0$, leading to $L_2(0,T;H^1(\Omega))$-ellipticity of $-(A + BK_B)$ if γ is moderate. This follows by the generalized Friedrichs inequality, cf. [138, Lemma 2.5].*

Remark 3.23. *In the above examples, the instability constant γ has to be moderate to show V-exponential stabilizability using the Friedrichs and Poincaré inequality if the control and observation region is not the whole domain. Classical exponential stabilizability, i.e., such that the closed-loop solution satisfies $\|x(t)\| \leq Me^{-\mu t}\|x_0\|$ for $M \geq 1$ and $\mu > 0$, can be shown, however, for arbitrary γ, cf. [3, Section 3.4.1]. We recall that if an operator is $L_2(0,T;V)$-elliptic, the solutions satisfy $\|x(t)\| \leq Me^{-\mu t}\|x_0\|$ with $M = 1$. In the case of Neumann boundary control, it was shown that the equation is exponentially stabilizable for arbitrary $\gamma > 0$ with overshoot constant $M = 1$ for the case where γ is smaller than the constant of the generalized Friedrichs or Poincaré inequality and with $M > 1$ for arbitrary large γ. For this fact, we again refer to [3, Section 3.4.1]. This illustrates that in the case of an autonomous equation, there are cases where classical stabilizability holds, whereas V-exponential stabilizability can not be established via the straightforward approach of Examples 3.21 and 3.22.*

The introduced stabilizability assumption will be the main tool to obtain a T-independent bound on the solution operator.

Assumption 3.24. *Let Assumption 3.9 hold and additionally assume that*

i) (A, B) and (A, C) are V-exponentially stabilizable,

ii) the stabilizing feedbacks K_B and K_C can be bounded independently of T in the sense of Assumption 3.9 vii).

Under these assumptions, we can conclude a preliminary stability estimate.

Lemma 3.25. *Assume $(y, \lambda) \in W([0,T])^2$ solves (3.13) and let Assumption 3.24 hold. Then there is a constant $c > 0$ independent of T such that*

$$\|x(T)\|^2 + \|x\|^2_{L_2(0,T;V)} \leq c\left(\|Cx\|^2_{L_2(0,T;Y)} + \|R^{-*}B^*\lambda\|^2_{L_2(0,T;U)} + \|l_2\|^2_{L_2(0,T;V^*)} + \|x_0\|^2\right),$$
$$\|\lambda(0)\|^2 + \|\lambda\|^2_{L_2(0,T;V)} \leq c\left(\|Cx\|^2_{L_2(0,T;Y)} + \|R^{-*}B^*\lambda\|^2_{L_2(0,T;U)} + \|l_1\|^2_{L_2(0,T;V^*)} + \|\lambda_T\|^2\right).$$

Proof. For the result on the state, we test the state equation of (3.13) with x and get

$$\left\langle \left(\frac{d}{dt} - A\right) x, x \right\rangle_{L_2(0,T;V^*)\times L_2(0,T;V)} = \langle BQ^{-1}B^*\lambda + l_2, x\rangle_{L_2(0,T;V^*)\times L_2(0,T;V)}.$$

Let K_C be a stabilizing feedback for (A, C) in the sense of Definition 3.20. Applying Lemma 3.4 iv) and adding $-\langle K_C C x, x\rangle_{L_2(0,T;V^*)\times L_2(0,T;V)}$ on both sides yields

$$\begin{aligned} &\frac{1}{2}\|x(T)\|^2 - \langle\, (A + K_C C)\, x, x\rangle_{L_2(0,T;V^*)\times L_2(0,T;V)} \\ &\qquad = \langle -K_C C x + BQ^{-1}B^*\lambda + l_2, x\rangle_{L_2(0,T;V^*)\times L_2(0,T;V)} + \frac{1}{2}\|x_0\|^2. \end{aligned} \tag{3.15}$$

Using the $L_2(0,T;V)$-ellipticity of $-(A + K_C C)$, we get that

$$\frac{1}{2}\|x(T)\|^2 + \alpha_1\|x\|^2_{L_2(0,T;V)} \leq \| - K_C C x + BQ^{-1}B^*\lambda + l_2\|_{L_2(0,T;V^*)}\|x\|_{L_2(0,T;V)} + \frac{1}{2}\|x_0\|^2.$$

for $\alpha > 0$. The left-hand side can be bounded from below by $\min\{\frac{1}{2}, \alpha\}(\|x(T)\|^2 + \|x\|^2_{L_2(0,T;V)})$. Then, using the estimate

$$\begin{aligned}\| - K_C Cx + BQ^{-1}B^*\lambda + l_2\|_{L_2(0,T;V^*)}\|x\|_{L_2(0,T;V)} \\ \leq \frac{1}{2}\left(\frac{\| - K_C Cx + BQ^{-1}B^*\lambda + l_2\|^2_{L_2(0,T;V^*)}}{c} + c\|x\|^2_{L_2(0,T;V)}\right)\end{aligned}$$

for $c = \min\{\frac{1}{2}, \alpha\}$ and applying the triangle inequality, we conclude the result for the state. The result for the adjoint follows analogously, testing the first equation of (3.13) with λ and subtracting the term $\langle BK_B\lambda, \lambda\rangle_{L_2(0,T;V^*)\times L_2(0,T;V)}$ on both sides. □

Remark 3.26. *If the stabilized operators are $L_2(0,t;V)$-elliptic for all $t > 0$, as it is the case in Examples 3.21 and 3.22, then one could also deduce a bound on $\|x\|_{C(0,T;H)} + \|x\|_{L_2(0,T;V)}$ by deriving (3.15) on $[0,t]$ for arbitrary $t \in [0,T]$. Similarly, a pointwise estimate for the adjoint follows by considering (3.15) on $[t,T]$ for any $t \in [0,T]$. We will conclude this estimate a posteriori after having obtained a bound in the $W([0,T])$-norm via the T-independent embedding $W([0,T]) \hookrightarrow C(0,T;H)$, cf. Lemma 3.4 i).*

The bounds in Lemma 3.25 still depend on x and λ. This dependence can be eliminated with the following lemma.

Lemma 3.27. *Let $(x,\lambda) \in W([0,T])^2$ solve (3.13). Then*

$$\begin{aligned}&\|Cx\|^2_{L_2(0,T;Y)} + \|R^{-*}B^*\lambda\|^2_{L_2(0,T;U)} \\ &\quad = -\langle x_0, \lambda(0)\rangle + \langle \lambda_T, x(T)\rangle - \langle l_2, x\rangle_{L_2(0,T;V^*)\times L_2(0,T;V)} + \langle l_1, \lambda\rangle_{L_2(0,T;V^*)\times L_2(0,T;V)} \\ &\quad \leq \frac{1}{2}\big(a\|z\|^2_{(L_2(0,T;V)\times H)^2} + \frac{\|l\|^2_{(L_2(0,T;V^*)\times H)^2}}{a}\big)\end{aligned}$$

for arbitrary $a > 0$, where $z := (x, x(T), \lambda, \lambda(0))$ and $l := (l_1, \lambda_T, l_2, x_0)$.

Proof. Testing the first equation of (3.13) with x and the third equation of (3.13) with λ and subtracting the former from the latter yields the result, cf. also Lemma 3.13 ii). The second estimate follows from the classical estimate $yz \leq ay^2 + \frac{z^2}{a}$ for all $y, z \in \mathbb{R}$ and $a > 0$. □

Eventually, we obtain the following stability estimate.

Proposition 3.28. *Assume $(x,\lambda) \in W([0,T])^2$ solves (3.13) and let Assumption 3.24 hold. Then, there exists a constant $c > 0$ independent of T such that*

$$\begin{aligned}\|\lambda(0)\|^2 + \|x(T)\|^2 + \|x\|^2_{L_2(0,T;V)} + \|\lambda\|^2_{L_2(0,T;V)} \\ \leq c(\|l_1\|^2_{L_2(0,T;V^*)} + \|l_2\|^2_{L_2(0,T;V^*)} + \|x_0\|^2 + \|\lambda_T\|^2).\end{aligned}$$

Proof. Adding the two stability estimates from Lemma 3.25 and using the bound of Lemma 3.27 yields the result. □

The stability estimate for the variables in the $L_2(0,T;V)$-norm allows for the deduction of an estimate in the $W([0,T])$-norm.

Theorem 3.29. *Assume* $(x,\lambda) \in W([0,T])^2$ *solve* (3.13) *and let Assumption 3.24 hold. Then, there exists a constant* $c > 0$ *independent of* T *such that*

$$\|x\|^2_{W([0,T])} + \|\lambda\|^2_{W([0,T])} \leq c(\|l_1\|^2_{L_2(0,T;V^*)} + \|l_2\|^2_{L_2(0,T;V^*)} + \|x_0\|^2 + \|\lambda_T\|^2).$$

Proof. Rewriting the third equation of (3.13), i.e.,

$$x' = Ax - BQ^{-1}B^*\lambda + l_2 \quad \text{in } L_2(0,T;V^*),$$

and taking norms on both sides implies

$$\begin{aligned}\|x'\|_{L_2(0,T;V^*)} \leq & \|A\|_{L(L_2(0,T;V),L_2(0,T;V^*))}\|x\|_{L_2(0,T;V)} \\ & + \|BQ^{-1}B^*\|_{L(L_2(0,T;V),L_2(0,T;V^*))}\|\lambda\|_{L_2(0,T;V)} + \|l_2\|_{L_2(0,T;V^*)}.\end{aligned}$$

Proceeding analogously for the adjoint and using the estimate on $\|x\|_{L_2(0,T;V)} + \|\lambda\|_{L_2(0,T;V)}$ of Proposition 3.28 yields the result, as $\|v\|_{W([0,T])} := \|v\|_{L_2(0,T;V)} + \|v'\|_{L_2(0,T;V^*)}$. □

The desired bound on the norm of the extremal equations' solution operator now follows immediately.

Corollary 3.30. *Let Assumption 3.24 hold. Then, there is a constant* $c > 0$ *independent of* T *such that*

$$\|M^{-1}\|_{L((L_2(0,T;V^*)\times H)^2, W([0,T])^2)} \leq c.$$

To conclude this part, we recall the most important results. Using the abstract scaling result Theorem 3.14 together with Corollary 3.30, we showed, for systems fulfilling Assumption 3.24, and thus in particular Examples 3.21 and 3.22, that perturbations of the extremal equations' right hand side growing exponentially in time only lead to errors in the variables that are growing exponentially in time. In particular, perturbations that are small on the initial part of a long time interval lead to errors in the solutions that are small on the initial part. This result is of particular interest in the context of a Model Predictive Controller, as the MPC feedback consists of the control on an initial part of the time interval. Moreover, in Theorem 3.16 we showed that, if the system is indeed autonomous, the solution of the dynamic optimal control problem resides close to the solution of the corresponding steady state problem for the majority of the time.

3.3 Numerical results

We will illustrate the theoretical results by means of two numerical examples of optimal control with a heat equation. First, in Section 3.3.1, we will depict the turnpike property established in Theorem 3.16 for the distributed control of an autonomous problem. Second, we will use the theoretical results of Theorem 3.14 to construct efficient grids that are tailored to an MPC context. The sensitivity result of Theorem 3.14 suggests that in order to obtain a high quality MPC feedback, i.e., the discretization error being small on $[0, \tau]$, the space and time grid should be predominantly refined on $[0, \tau]$. We will investigate the performance of different a priori time discretization schemes following this idea. To this end, we consider the distributed control of an autonomous problem in Section 3.3.1 and the boundary control of a non-autonomous problem in Section 3.3.2. Finally, in Section 3.3.3 we will apply an extension of this approach to a priori space refinement and discuss limitations of these techniques, motivating the use of more sophisticated a posteriori methods, which will be discussed in Chapter 5.

The spatial discretization was performed via standard finite elements, and the temporal discretization is (loosely speaking) an implicit Euler scheme. We will not go into detail here, as the implementation details will be provided in Chapter 5.

A priori time discretization strategies. Motivated by the sensitivity result of Theorem 3.14, we present two methods to generate an a priori time grid tailored to MPC. First, we suggest a construction of an exponential grid as described in Algorithm 2, motivated by the exponential estimates of the previous sections.

Algorithm 2 Exponential grid generation

Given: Rate $c > 0$, $t_0 = 0$

Set $I := \frac{1}{N-1} \int_0^T e^{-ct}\, dt = \frac{1}{(N-1)c}(1 - e^{-cT})$

for $i \in \{0, \ldots, N-2\}$ **do**

  $t_{i+1} = -\frac{1}{c} \log(-cI + e^{-ct_i})$

end for

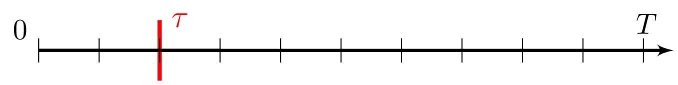
The algorithm above computes vertices t_i, $i \in \{0, \ldots, N-1\}$ such that

$$\int_{t_i}^{t_{i+1}} e^{-ct}\, dt = I \qquad \forall i \in \{0, \ldots, N-2\}. \tag{3.16}$$

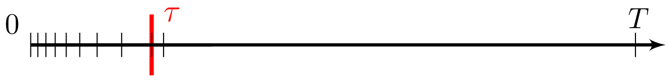
In our particular case, we chose $c = 1$. As a second strategy we use the same number of grid points in $[0, \tau]$ as in $[\tau, T]$. If $\tau \ll T$, this naturally induces a clustering of discretization points in $[0, \tau]$. We will compare both specialized discretization schemes with a standard uniform grid.

We briefly illustrate the three different discretization schemes for 11 time discretization points. In all cases we ensure to have one time grid point at the implementation horizon τ.

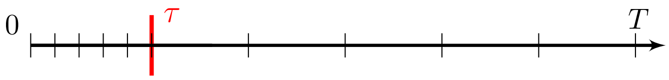
i) Uniform:



ii) Exponential:



iii) Piecewise uniform:



For each of these grids we run the MPC Algorithm 1, where the optimal control problem in each step is solved on the respective a priori grid. The simulation of the model in step 5 of Algorithm 1 is always performed on a very fine temporal mesh with 51 grid points on the interval $[k\tau, (k+1)\tau]$ for every iteration index k. The spatial mesh was kept constant in time and we used 768 triangles and correspondingly 417 vertices at every time step for the optimal control problem as well as the simulation.

3.3.1 Distributed control with static reference

We first consider a linear quadratic optimal control problem with distributed control on a rectangular domain $\Omega := [0,3] \times [0,1]$. We choose the time horizon $T = 10$, $Y = U = L_2(\Omega)$, and consider the cost functional

$$J(x,u) := \frac{1}{2}\|(x(t) - x_{\mathrm{d}}^{\mathrm{stat}})\|^2_{L_2(0,T;L_2(\Omega))} + \frac{\alpha}{2}\|u\|^2_{L_2(0,T;L_2(\Omega))},$$

where $\alpha > 0$ is a Tikhonov parameter and $x_{\mathrm{d}}^{\mathrm{stat}}$ is a time-independent reference defined by

$$\begin{aligned} x_{\mathrm{d}}^{\mathrm{stat}}(\omega) &:= g\left(\frac{10}{3}\left\|\omega - \begin{pmatrix}1.5\\0.5\end{pmatrix}\right\|\right), \\ \text{where} \qquad g(s) &:= \begin{cases} 10e^{1-\frac{1}{1-s^2}} & s < 1 \\ 0 & \text{else.} \end{cases} \end{aligned} \tag{3.17}$$

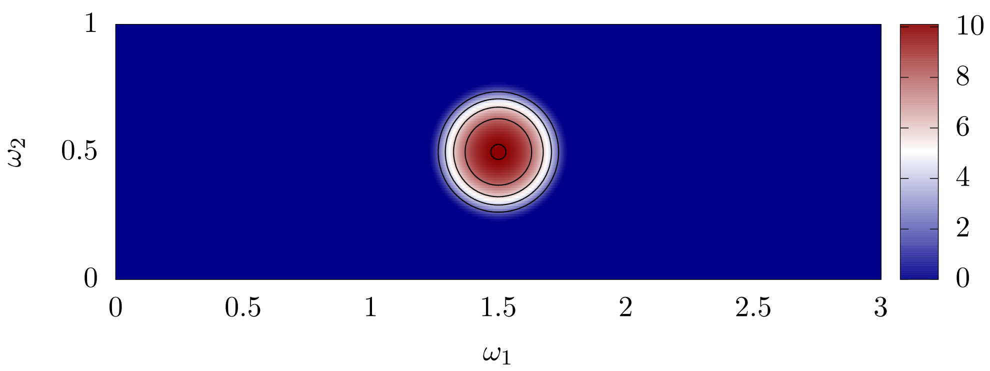


Figure 3.1: Static reference trajectory $x_{\mathrm{d}}^{\mathrm{stat}}(\omega_1, \omega_2)$.

As introduced in Example 3.21, we consider dynamics governed by the parabolic PDE in classical form

$$\begin{aligned} x' - 0.1\Delta x - sx &= u && \text{in } \Omega \times (0,T), \\ x &= 0 && \text{in } \partial\Omega \times (0,T), \\ x(0) &= 0 && \text{in } \Omega, \end{aligned}$$

where $s \geq 0$ is an instability parameter. The natural space in this example is $V = H_0^1(\Omega)$, which yields the equation in weak form

$$\begin{aligned} x' - Ax - Bu &= 0 && \text{in } L_2(0,T;V^*), \\ x(0) &= 0 && \text{in } L_2(\Omega), \end{aligned}$$

where

$$\begin{aligned} \langle Ax, v\rangle_{L_2(0,T;V^*)\times L_2(0,T;V)} &:= \int_0^T \int_\Omega -0.1\nabla x \cdot \nabla v + sxv \, d\omega \, dt, \\ \langle Bu, v\rangle_{L_2(0,T;V^*)\times L_2(0,T;V)} &:= \int_0^T \int_\Omega uv \, d\omega \, dt. \end{aligned}$$

We note that for $s = 0$ this equation models a linear heat equation with zero initial condition that is stable due to the ellipticity of the negative Laplacian in $H_0^1(\Omega)$. The parameter s allows us to reduce the stability, or to render the problem unstable. It can be easily seen that (A, B) is $H_0^1(\Omega)$-exponentially stabilizable for arbitrary s by choosing the feedback $K_B x = -(s + 0.1)x$, leading to $-\langle (A + BK_B)x, x\rangle_{L_2(0,T;V^*)\times L_2(0,T;V)} = 0.1\|x\|_{L_2(0,T;H_0^1(\Omega))}$. Similarly, (A, C) can be shown to be $H_0^1(\Omega)$-exponentially stabilizable.

We apply four steps of Algorithm 1, where the MPC implementation horizon is chosen as $\tau = 0.5$ for two different choices of the instability parameter s. We observe that the solution of the optimal control problem in every MPC step satisfies the turnpike property as proven in Theorem 3.16. In both configurations the turnpike is approximately reached after one MPC iteration. We see that the leaving arc is still present for the open-loop trajectories of MPC iterations two to four. For the problem configuration depicted on the right, we observe that switching off the control action towards the end of the optimization horizon leads to an increasing norm of the state. This is due to the instability of the uncontrolled state equation. In contrast, in the stable problem depicted on the left, switching off the control leads to a decay of the state norm. Additionally, the leaving arc is less pronounced for smaller values of the Tikhonov parameter α. Loosely speaking, this occurs because the control can be held longer at the turnpike due to the lower control costs.

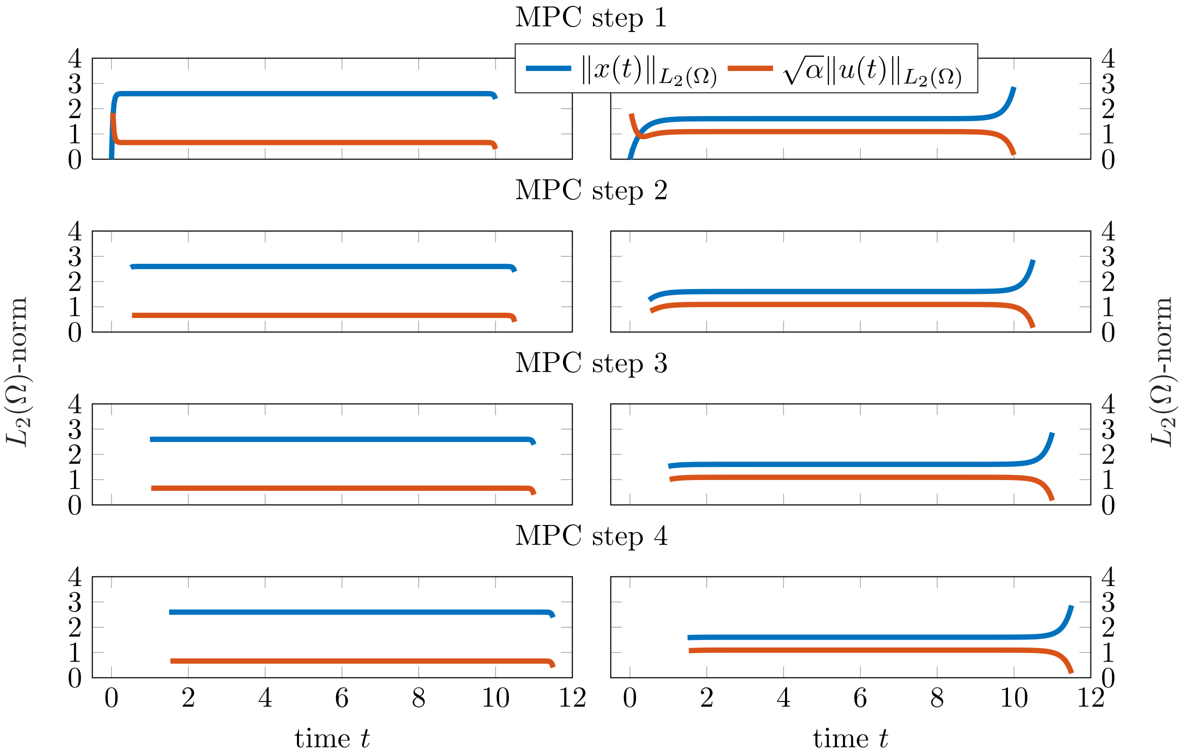


Figure 3.2: Spatial norm of open-loop control and state over time for every MPC step. Left: Stable problem with parameters $s = 0$ and $\alpha = 10^{-3}$. Right: Unstable problem with parameters $s = 4$ and $\alpha = 10^{-1}$.

In Figure 3.3 the cost functional value of the closed-loop trajectory is depicted. The specialized a priori grids show a better performance than the classical uniform grid in all cases. While the closed-loop cost is moderately lower for the stable problem shown on the left, the difference is significant in case of unstable dynamics depicted on the right. The choice of a concentrated time grid towards $[0, \tau]$ leads to a closed-loop cost that is lower by almost one order of magnitude, despite identical numerical effort. In the unstable problem setting, the closed-loop cost of approximately 4.5 which was achieved using 21 grid points in an exponential or piecewise uniform grid was reached not before using 101 uniformly distributed grid points. Due to our particular choice of the scaling parameter in (3.16), on $[0, \tau]$ the exponential grid is coarser than the piecewise uniform grid. This might serve as an explanation, why the piecewise uniform grid performs better than the exponential grid. A more aggressive exponential refinement on $[0, \tau]$ by increasing the parameter $c > 0$ in (3.16) could increase the performance of the exponential grid for these particular examples. In the nonlinear case, more precisely in Figure 4.4, we will observe a better performance of the exponential grid compared to the piecewise uniform grid.

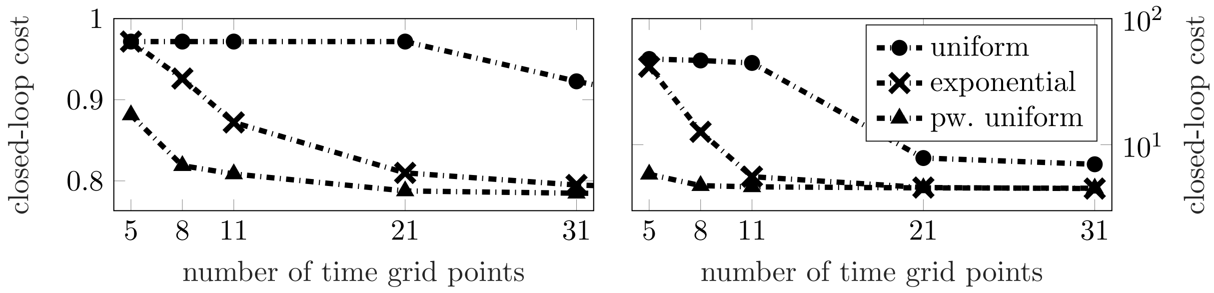


Figure 3.3: Comparison of MPC closed-loop cost for an autonomous problem with different priori time discretization schemes. Left: Stable problem with parameters $s = 0$ and $\alpha = 10^{-3}$. Right: Unstable problem with parameters $s = 4$ and $\alpha = 10^{-1}$.

3.3.2 Boundary control with dynamic reference

We now consider the case of Neumann boundary control with distributed observation, cf. Example 3.22. In this case, $Y = L_2(\Omega)$ and $U = L_2(\partial\Omega)$ and the cost functional is given by

$$J(x,u) = \frac{1}{2}\|(x - x_{\mathrm{d}}^{\mathrm{dyn}})\|^2_{L_2(0,T;L_2(\Omega))} + \frac{\alpha}{2}\|u\|^2_{L_2(0,T;L_2(\partial\Omega))},$$

where again $\alpha > 0$. The time-dependent reference trajectory $x_{\mathrm{d}}^{\mathrm{dyn}}$ is defined by

$$x_{\mathrm{d}}^{\mathrm{dyn}}(t,\omega) := g\left(\frac{10}{3}\left\|\omega - \begin{pmatrix}\omega_{1,\mathrm{peak}}(t)\\ \omega_{2,\mathrm{peak}}(t)\end{pmatrix}\right\|\right), \tag{3.18}$$

$$\text{where} \qquad g(s) := \begin{cases} 10e^{1-\frac{1}{1-s^2}} & s < 1 \\ 0 & \text{else} \end{cases} \tag{3.19}$$

and

$$\omega_{1,\mathrm{peak}}(t) := 1.5 - \cos\left(\pi\left(\frac{t}{10}\right)\right), \qquad w_{2,\mathrm{peak}}(t) := \left|\cos\left(\pi\left(\frac{t}{10}\right)\right)\right|,$$

as depicted in Figure 3.4.

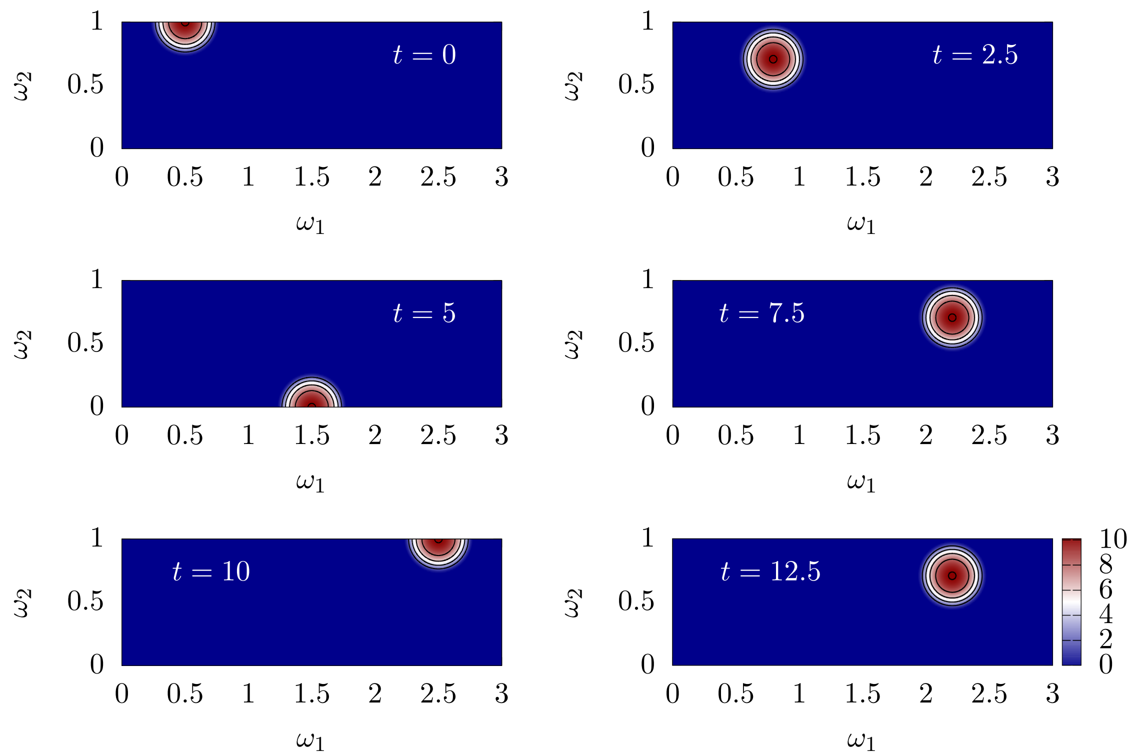


Figure 3.4: Snapshots of the dynamic reference trajectory $x_{\mathrm{d}}^{\mathrm{dyn}}(t, \omega_1, \omega_2)$ at different time instances.

The dynamics in classical form are given by

$$\begin{aligned} x' - 0.1\Delta x &= 0 && \text{in } \Omega \times (0,T), \\ 0.1\frac{\partial x}{\partial \nu} &= u && \text{in } \partial\Omega \times (0,T), \\ x(0) &= 0 && \text{in } \Omega. \end{aligned}$$

In this case, we choose $V = H^1(\Omega)$ and the equation in weak form reads

$$\begin{aligned} x' - Ax - Bu &= 0 && \text{in } L_2(0,T;V^*), \\ x(0) &= 0, \end{aligned}$$

where

$$\begin{aligned} \langle Ax, v\rangle_{L_2(0,T;V^*)\times L_2(0,T;V)} &:= \int_0^T \int_\Omega -0.1\nabla x \cdot \nabla v \, d\omega \, dt, \\ \langle Bu, v\rangle_{L_2(0,T;V^*)\times L_2(0,T;V)} &:= \int_0^T \int_{\partial\Omega} u \, \mathrm{tr}\, v \, d\omega \, dt. \end{aligned}$$

Analogously to Example 3.22, $H^1(\Omega)$-exponential stabilizability of (A, C) follows immediately, whereas $H^1(\Omega)$-exponential stabilizability of (A, B) can be shown with the generalized Friedrichs inequality, cf. [138, Lemma 2.5]. We apply four steps of Algorithm 1 with the implementation horizon $\tau = 1$. As the reference trajectory is time dependent, we do not have a steady state turnpike property, cf. Figure 3.5. However, it can be observed that the open-loop trajectories for MPC step two to four are very similar. This is due to the fact that even in a non-autonomous setting, turnpike properties are often present, cf. [68]. In that case, the solutions to the dynamic optimal control problem can be shown to be close to the solution of an infinite horizon optimal control problem with free initial data for the majority of the time. In the autonomous case, this reduces to an optimal steady state, cf. [24, Section 6].

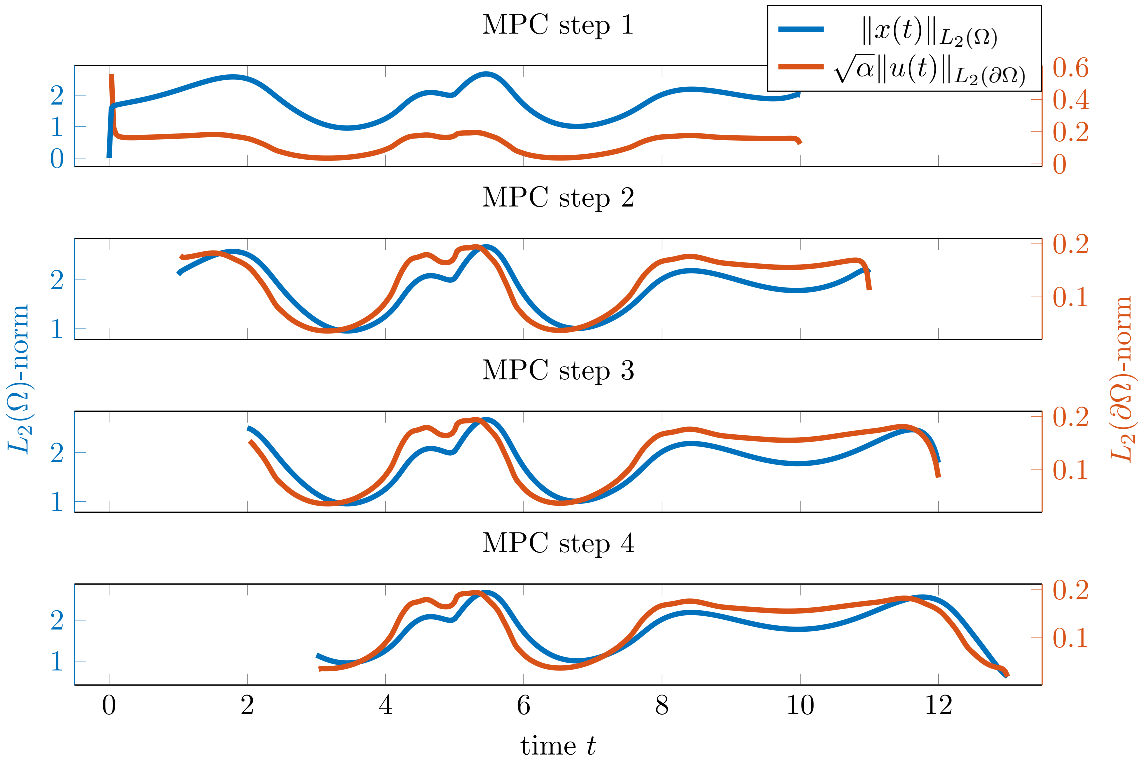


Figure 3.5: Spatial norm of open-loop state and control over time for every MPC step for non-autonomous problem with $\alpha = 10^{-3}$.

Despite the lack of a steady state turnpike, Theorem 3.14 still suggests that meshes concentrated on $[0, \tau]$ should perform better than uniform ones. Figure 3.6 depicts the closed-loop performance of different a priori grids. We observe that again, the exponential and piecewise uniform a priori grids perform better than the uniform grid.

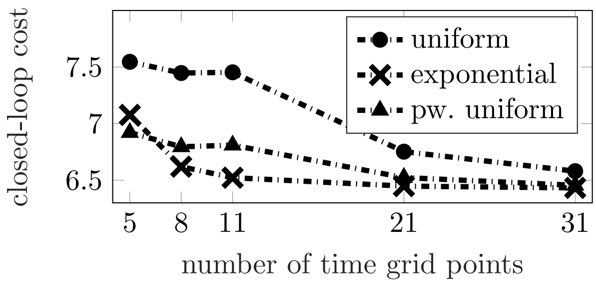


Figure 3.6: Comparison of cost functional values of MPC closed-loop trajectory for different a priori space discretization schemes with Tikhonov parameter $\alpha = 10^{-3}$.

3.3.3 Discussion

We will briefly discuss a possible extension of the a priori time grid generation presented above to a priori space refinement. In this context, it is important to note that the space discretization is allowed to be time dependent, i.e., to every time grid point, a space grid that is independent of the neighboring space grids is assigned. This has the advantage of allowing full flexibility in the grid refinement. The price to pay is that the implementation of the time stepping scheme requires particular attention, as the finite element spaces can change every time discretization point. We will discuss an efficient remedy of this in detail in Chapter 5. In this part, we compare the MPC closed-loop performance of the following a priori space discretization schemes. For time discretization, we use the piecewise uniform refinement as described in Section 3.3 with 11 time grid points.

A priori space discretization strategies. We will compare the performance of the following a priori space discretization strategies:

i) Uniform: All space grids share the same number of degrees of freedom.

ii) Piecewise uniform:

   (a) We apply **one additional uniform refinement** to the grids on $[0, \tau]$ compared to the space grids on $(\tau, T]$.

   (b) We apply **two additional uniform refinements** to the grids on $[0, \tau]$ compared to the space grids on $(\tau, T]$.

We evaluate the performance of these approaches by means of both examples introduced in Sections 3.3.1 and 3.3.2. We observe that in particular for a lower number of space discretization points, the closed-loop cost achieved with the specialized grids ii)(a) and ii)(b) is better than the closed-loop cost achieved with a uniform space discretization. However, for higher total space grid points, a saturation effect takes place and no difference of the three approaches can be observed.

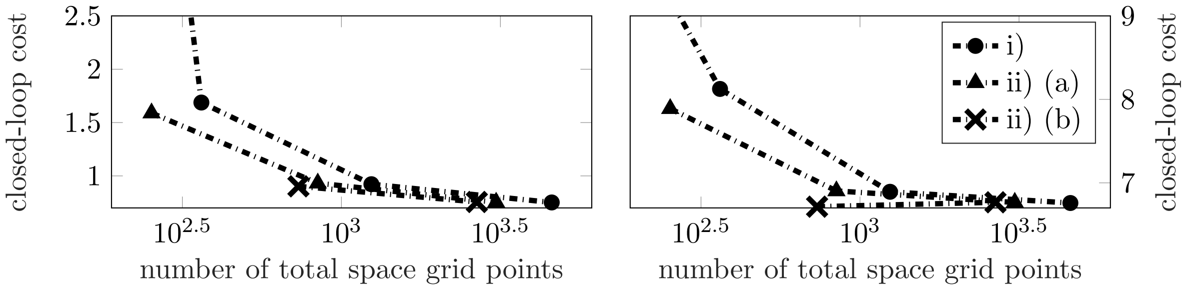


Figure 3.7: Comparison of cost functional values of MPC closed-loop trajectory for different a priori space discretization schemes. Left: Stable autonomous problem of Section 3.3.1 with parameters $s = 0$, $\alpha = 10^{-3}$, and $\tau = 0.5$. Right: Stable non-autonomous problem of Section 3.3.2 with parameters $\alpha = 10^{-3}$ and $\tau = 1$.

We conclude that a priori grid refinement tailored to MPC can be efficient in increasing the closed-loop performance. In particular, ignoring the treatment of dynamic space grids for the moment, no additional computational effort is required to perform the specialized discretization schemes. Nonetheless, as to be expected, there are limitations to these a priori approaches. First, no error estimator is at hand, i.e., it is not clear how much time or space grid points are needed to ensure a particular accuracy. Thus, the suggested a priori discretization schemes are only of qualitative nature. In particular, only uniform refinements in space were performed without considering any structure of the optimal triple. Additionally, despite the exponential decay of perturbations established in Theorem 3.14, it could be worth to refine the space or time grid outside of $[0, \tau]$. This is due to the fact that the precise decay parameter is not known beforehand and if it is very small, large perturbations could affect the optimal triple on the initial part. For these reasons, we will inspect a goal oriented a posteriori grid refinement technique in Chapter 5 that is particularly well suited to address these drawbacks.

3.4 Outlook

We briefly outline possible extensions of the work presented in this chapter.

- Future research could be focused on an extension of the results of this chapter to non-autonomous hyperbolic problems. To this end, a framework with evolution families could turn out useful, cf. the discussion at the beginning of this chapter.
- Similar to the analysis performed in Section 3.2.2, one could try to deduce a turnpike result for non-autonomous systems. For that matter, one could compare optimal solutions for the problem on $[0, T]$ with the optimal solutions for the problem on $[0, \infty]$ with free initial data. However, the core of the analysis performed in this work is a comparison of the optimality conditions and the derivation of such for problems on an infinite horizon is a very delicate issue, even for finite dimensional systems.

- Recently, an approach to deduce turnpike results for shape optimization problems was presented in, cf. [88]. Motivated by the close connection of decay of perturbations and the turnpike property established in this part, one could try to prove locality of discretization errors for shape optimization problems.

Chapter 4

Sensitivity and turnpike analysis for nonlinear optimal control problems

In this chapter, we will extend the sensitivity and turnpike analysis of Chapters 2 and 3 to nonlinear parabolic problems. To this end, we briefly provide a different interpretation of the sensitivity result of Theorem 3.14 and the turnpike result of Theorem 3.16: If the solution operator of the extremal equations is bounded independently of T in unscaled spaces, then there is a scaling parameter $\mu > 0$ independent of T such that the solution operator is also bounded independently of T in scaled spaces with scaling function $e^{-\mu t}$ for the sensitivity result and with scaling function $\frac{1}{e^{-\mu t}+e^{-\mu(T-t)}}$ for the turnpike result, respectively. The analysis in this chapter will use this methodology to employ an implicit function theorem. We thus formulate the nonlinear extremal equations as a nonlinear operator equation. In order to apply the implicit function theorem, we have to perform two main steps. The first is to choose a functional analytic framework in which we can establish continuity and differentiability of the nonlinear operator equation. This step will heavily rely on the theory of superposition operators. As a second step, we show a T-independent bound on the solution operator corresponding to the linearized extremal equations, similarly to, e.g., the approach in Sections 2.2.3 and 3.2.3. In this context, the extremal equations linearized at a solution trajectory can be non-autonomous and we benefit from the analysis of non-autonomous problems in the previous chapter. By the nature of the implicit function theorem, all results will be local, i.e., they hold for small perturbations of the extremal equations. In case of the turnpike result, this means that the initial and terminal datum of the dynamic problem need to be sufficiently close to the turnpike, and, in the case of the sensitivity result, the perturbation of the dynamics by, e.g., discretization errors is required to be sufficiently small.

We briefly recall existing work on nonlinear turnpike theory. Nonlinear finite dimensional problems were considered in [136], including the case of nonlinear initial and terminal conditions. This was extended to a Hilbert space setting in [113, 114, 135]. A turnpike result for the two-dimensional Navier-Stokes equations was obtained in [155]. These works analyze the turnpike property via the extremal equations and are of local nature, i.e., the initial resp. terminal datum for state and adjoint need to be close to the turnpike. In [110], a semi-global turnpike result for a semilinear heat equation with initial datum of arbitrary size is given, under the assumption that either the state reference trajectory is small or that the control acts everywhere. A geometric approach to tackle nonlinear problems was presented in [123]. A different approach that leads to global turnpike properties is stability analysis based on a dissipativity concept. Motivated by the seminal papers by Willems [148, 149], a notion of dissipativity for optimal control problems can be defined, where the supply rate is related to the cost functional. Assuming this dissipativity property, a global turnpike result for states and controls was deduced in, e.g., [53, 65, 68] or [66, Proposition 8.15]. Under the assumption of a global turnpike property of states and controls, a global turnpike property for the corresponding adjoint states was derived in [52]. The connection of dissipativity and the turnpike property is discussed in [62, 63, 65, 134]. The difficult task remaining is to indeed verify this dissipativity notion in particular applications. The reader is referred to [62, 63] for a construction of storage functions in a linear quadratic finite dimensional setting, which could also offer a promising strategy for problems with monotone nonlinearities. Recently, the connection of turnpike properties and long-time behavior of the Hamilton-Jacobi equation was analyzed in [48].

Structure. In Section 4.1, the optimal control problem of interest, the corresponding first-order optimality conditions, and the implicit function theorem are introduced. Moreover, we present the concept of superposition operators and discuss T-dependence of continuity and differentiability. For specific problems, two main steps are necessary to apply the implicit function theorem: The first one is to show T-independent invertibility of the linearization corresponding to the first-order optimality system. The second one is to verify a T-uniform differentiability condition of the superposition operators corresponding to the nonlinearities. In Section 4.3, we will analyze the case of optimal control with an ordinary differential equation to illustrate the main steps without functional analytic technicalities. After that, we address the case of a semilinear parabolic equation in Section 4.4. We analyze the case where the data is sufficiently smooth—i.e., the initial datum lies in $H^1(\Omega)$ resp. $H_0^1(\Omega)$, depending on the boundary conditions and the right-hand sides of the dynamics are supposed to be in $L_2(0,T;L_2(\Omega))$. In that case, continuity and differentiability of the superposition operators follows straightforwardly. However, the T-independent bound on the solution operator's norm requires a refined approach. Last, in Section 4.5, we present numerical examples of distributed control of a semilinear heat equation and boundary control of a quasilinear heat equation.

This chapter comprises the results of [72, 73].

4.1 Setting and preliminaries

We briefly define the nonlinear optimal control problem of interest and formally derive the optimality conditions. Assume that $(V, \|\cdot\|_V)$ is a separable and reflexive Banach space and H is a separable and real Hilbert space with scalar product $\langle \cdot, \cdot \rangle$ and corresponding norm $\|\cdot\|$.

Further suppose that $V \hookrightarrow H \cong H^* \hookrightarrow V^*$ form a Gelfand triple, cf. Section 3.1.1. The control space U will be assumed to be a real Hilbert space with scalar product denoted by $\langle \cdot, \cdot \rangle_U$ and induced norm $\| \cdot \|_U$.

We consider the following parabolic nonlinear optimal control problem.

$$\begin{aligned} \min_{(x,u)} J(x,u) &:= \int_0^T \bar{J}(t, x(t)) + \frac{1}{2}\|R(u(t) - u_{\mathrm{d}}(t))\|_U^2 \, dt \\ \text{s.t. } x'(t) &= \bar{A}(x(t)) + \bar{B}u(t) + f(t), \\ x(0) &= x_0, \end{aligned} \tag{4.1}$$

where $x_0 \in H$, $f \in L_2(0,T;V^*)$, $u_{\mathrm{d}} \in L_2(0,T;U)$, $J(x,u)\colon L_2(0,T;V) \times L_2(0,T;U) \to \mathbb{R}$ is a sufficiently smooth functional, $\bar{B}\colon U \to V^*$ is a continuous and linear operator, $\bar{A}\colon V \to V^*$ is a sufficiently smooth operator, $R \in L(L_2(0,T;U), L_2(0,T;U))$ such that $\|Ru\|^2_{L_2(0,T;U)} \geq \alpha \|u\|^2$ for $\alpha > 0$.

Similarly to (3.3) we define an operator corresponding to the PDE with initial condition denoted by $\Lambda\colon W([0,T]) \to L_2(0,T;V^*) \times H$ via

$$\langle \Lambda(x), (\lambda, \lambda_0)\rangle_{(L_2(0,T;V)\times H)^* \times (L_2(0,T;V)\times H)} := \int_0^T \langle x'(t) - \bar{A}(x(t)), \lambda(t)\rangle_{V^* \times V} \, dt + \langle x(0), \lambda_0 \rangle$$

for $(\lambda, \lambda_0) \in L_2(0,T;V^*) \times H$ and $B\colon L_2(0,T;U) \to L_2(0,T;V^*)$ via

$$\langle Bu, \lambda \rangle_{L_2(0,T;V^*)\times L_2(0,T;V)} := \int_0^T \langle \bar{B}u(t), \lambda(t)\rangle_{V^*\times V} \, dt$$

for $\lambda \in L_2(0,T;V)$. This allows us to briefly write the nonlinear PDE in variational form

$$\begin{aligned} \langle \Lambda(x), (\lambda, \lambda_0)\rangle_{(L_2(0,T;V)\times H)^* \times (L_2(0,T;V)\times H)} &- \langle Bu, \lambda\rangle_{L_2(0,T;V^*)\times L_2(0,T;V)} \\ &= \langle f, \lambda \rangle_{L_2(0,T;V^*)\times L_2(0,T;V)} + \langle x_0, \lambda_0 \rangle \end{aligned}$$

for all $(\lambda, \lambda_0) \in L_2(0,T;V) \times H$. We further define $A\colon L_2(0,T;V) \to L_2(0,T;V^*)$ by

$$\langle A(x), \lambda \rangle_{L_2(0,T;V^*)\times L_2(0,T;V)} := \int_0^T \langle \bar{A}(x(t)), \lambda(t) \rangle_{V^*\times V} \, dt.$$

We will assume that the optimal control problem (4.1) has a solution $(x,u) \in W([0,T]) \times L_2(0,T;U)$. One important ingredient for establishing this property are the classical lower semicontinuity and coercivity properties of the cost functional. A second factor can be to establish continuous invertibility of $\Lambda(x)$, i.e., the existence of a continuous control to state map. In the linear case , i.e., if $A(x) = Ax$, we ensured this by assuming $-A$ to satisfy a Gårding inequality, cf. (3.2):

$$\exists \omega \in \mathbb{R}, \alpha > 0: \quad \alpha \|x\|^2_{L_2(0,T;V)} \leq -\langle Ax, x\rangle_{L_2(0,T;V^*)\times L_2(0,T;V)} + \omega \|x\|^2_{L_2(0,T;H)}.$$

For solvability of semilinear equations with globally Lipschitz semilinearities, we refer to [109, Chapter 6] and [138, Chapter 5]. Locally Lipschitz semilinearities are treated in [119], where global existence of solutions was ensured by sufficiently regular data $(Bu + f, x_0)$, such that the solution is bounded, i.e., $x \in L_\infty(0,T;L_\infty(\Omega))$. For a in-depth analysis of optimal control problems governed by quasilinear parabolic equations, the interested reader is referred to [21, 31, 87, 104, 108].

We introduce a Lagrange multiplier $(\lambda, \lambda_0) \in L_2(0,T;V)\times H$ and define the Lagrange function via

$$\begin{aligned} &L(x,u,(\lambda,\lambda_0)) := \\ &\quad J(x,u) + \langle \Lambda(x), (\lambda,\lambda_0)\rangle_{(L_2(0,T;V)\times H)^*\times(L_2(0,T;V)\times H)} - \langle Bu+f,\lambda\rangle_{L_2(0,T;V^*)\times L_2(0,T;V)}. \end{aligned} \tag{4.2}$$

Proceeding formally, the first-order optimality conditions of (4.1) are characterized by the stationarity conditions of the Lagrange function at a minimizer $(x,u,(\lambda,\lambda_0))$, i.e.,

$$0 = \begin{pmatrix} L_x(x,u,(\lambda,\lambda_0)) \\ L_u(x,u,(\lambda,\lambda_0)) \\ L_{(\lambda,\lambda_0)}(x,u,(\lambda,\lambda_0)) \end{pmatrix} = \begin{pmatrix} J_x(x,u) + \Lambda'(x)^*(\lambda,\lambda_0) \\ R^*R(u-u_\mathrm{d}) - B^*\lambda \\ \Lambda(x) - Bu - f \end{pmatrix}.$$

If $A'(x)$ satisfies the Gårding inequality (3.2), we get the improved regularity $\lambda \in W([0,T])$ and $\lambda(0) = \lambda_0$, cf. [127, Proposition 3.8] as in the linear quadratic setting considered in Section 3.1.2. This allows us to write the adjoint equation as a backwards-in-time equation in the variable λ. Further separating the initial and terminal conditions from the dynamics leads to the extremal system

$$L'(x,u,\lambda) = \begin{pmatrix} J_x(x,u) - \lambda' - A'(x)^*\lambda \\ \lambda(T) \\ R^*R(u-u_\mathrm{d}) - B^*\lambda \\ x' - A(x) - Bu - f \\ x(0) - x_0 \end{pmatrix} = 0. \tag{4.3}$$

Remark 4.1. *The quadratic dependence of the objective function on the control allows for an elimination of the control analogously to the linear quadratic case. For more general problem settings, in order to represent the optimal control by the adjoint state arising in the first-order necessary conditions, a standard assumption is the existence of $\alpha > 0$ such that $J_{uu}(x,u)(\delta u, \delta u) \geq \alpha\|\delta u\|^2_{L_2(0,T;U)}$ for all $\delta u \in L_2(0,T;U)$. This property is sometimes referred to as the strengthened Legendre-Clebsch condition, cf. [25, Chapter 6]. A second aspect if one allows for a general nonlinear control dependence is that improved regularity of the optimal control might be needed to conclude an implicit function argument. In particular cases, this improved regularity can be established by classical bootstrapping in the optimality system.*

Setting $Q := R^*R$ and $L_r(x,\lambda) := L(x, Q^{-1}B^*\lambda + u_\mathrm{d}, \lambda)$, the reduced extremal equations read

$$L_r'(x,\lambda) = \begin{pmatrix} J_x(x,u) - \lambda' - A'(x)^*\lambda \\ \lambda(T) \\ x' - A(x) - BQ^{-1}B^*\lambda - Bu_\mathrm{d} - f \\ x(0) - x_0 \end{pmatrix} = 0. \tag{4.4}$$

This nonlinear system will be the starting point of our subsequent analysis. We present two perturbations of the extremal equations (4.4) that we aim to analyze in this chapter.

First, in order to obtain a sensitivity result, we will consider $(\tilde{x}, \tilde{u}, \tilde{\lambda}) \in W([0,T]) \times L_2(0,T;U) \times W([0,T])$ that solves a perturbed version of (4.4), i.e.,

$$L_r'(\tilde{x}, \tilde{\lambda}) = \begin{pmatrix} J_x(\tilde{x}, \tilde{u}) - \tilde{\lambda}' - A'(\tilde{x})^* \tilde{\lambda} \\ \tilde{\lambda}(T) \\ \tilde{x}' - A(\tilde{x}) - BQ^{-1}B^* \tilde{\lambda} - Bu_{\mathrm{d}} - f \\ \tilde{x}(0) - x_0 \end{pmatrix} = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_T \\ \varepsilon_2 \\ \varepsilon_0 \end{pmatrix}, \tag{4.5}$$

where $\tilde{u} = Q^{-1}B^*\tilde{\lambda} + u_{\mathrm{d}}$ and $\varepsilon = (\varepsilon_1, \varepsilon_T, \varepsilon_2, \varepsilon_0) \in (L_2(0,T;V^*) \times H)^2$.

It is important to note that in order to derive a sensitivity result, we do not assume the existence of a corresponding steady state problem, i.e., in particular $\bar{J}$ can explicitly depend on time which is the case for, e.g., tracking-type cost functionals with time-dependent reference trajectories.

Second, to derive a turnpike result, we will consider the first-order necessary optimality conditions of the steady state problem as a perturbation of the first-order optimality conditions of the dynamic problem. In that context, we will always assume that $\bar{J}(t,x) \equiv \bar{J}(x)$, i.e., $\bar{J}$ does not explicitly depend on time and $u_{\mathrm{d}} \in U$, $R \in L(U,U)$, $f \in V^*$ are independent of time. To indicate this time-independence, we denote $\bar{R} := R$, $\bar{u}_{\mathrm{d}} := u_d$ and $\bar{f} := f$. We thus formulate the corresponding steady state problem

$$\begin{aligned} \min_{(\bar{x},\bar{u})} \; & \bar{J}(\bar{x}) + \frac{1}{2}\|\bar{R}(\bar{u} - \bar{u}_{\mathrm{d}})\|_U^2 \\ \text{s.t.} \; & -\bar{A}(\bar{x}) = \bar{B}\bar{u} + \bar{f}, \end{aligned} \tag{4.6}$$

where we again assume that there is a minimizer $(\bar{x}, \bar{u}) \in V \times U$. For $(\bar{x}, \bar{u}, \bar{\lambda}) \in V \times U \times V$, we define the Lagrange function of the steady state system as $\bar{L}(\bar{x}, \bar{u}, \bar{\lambda}) := \bar{J}(\bar{x}, \bar{u}) - \langle \bar{A}(\bar{x}) - \bar{B}\bar{u}, \lambda \rangle_{V^* \times V}$, which leads to the first-order conditions

$$\bar{L}'(\bar{x}, \bar{u}, \bar{\lambda}) = \begin{pmatrix} \bar{J}_x(\bar{x}) - \bar{A}'(\bar{x})^* \bar{\lambda} \\ \bar{R}^* \bar{R}(\bar{u} - \bar{u}_d) - \bar{B}^* \bar{\lambda} \\ -\bar{A}(\bar{x}) - \bar{B}\bar{u} - \bar{f} \end{pmatrix} = 0. \tag{4.7}$$

Setting $\bar{Q} = \bar{R}^*\bar{R}$, eliminating the control via $\bar{u} = \bar{Q}^{-1}\bar{B}^*\bar{\lambda} + u_{\mathrm{d}}$ and defining the reduced static Lagrangian $\bar{L}_r(\bar{x}, \bar{\lambda}) := \bar{L}(\bar{x}, \bar{Q}^{-1}\bar{B}^*\bar{\lambda} + u_{\mathrm{d}}, \bar{\lambda})$. By the same argumentation as in the linear quadratic case, cf. Lemma 2.29 or the proof of Theorem 3.16, this steady state system can be written as a perturbation of the dynamic extremal equations by interpreting $\bar{\lambda}$ and $\bar{x}$ as functions constant in time and adding $\bar{\lambda}' = \bar{x}' = 0$ and initial resp. terminal values to the equations, i.e.,

$$L_r'(\bar{x}, \bar{\lambda}) = \begin{pmatrix} J_x(\bar{x}, \bar{u}) - \bar{\lambda}' - A'(\bar{x})^* \bar{\lambda} \\ \bar{\lambda}(T) \\ \bar{x}' - A(\bar{x}) - BQ^{-1}B^* \bar{x} - Bu_{\mathrm{d}} - f \\ \bar{x}(0) - x_0 \end{pmatrix} = \begin{pmatrix} 0 \\ \bar{\lambda} \\ 0 \\ \bar{x} - x_0 \end{pmatrix}. \tag{4.8}$$

The main results of this chapter will be the following: On the one hand we will deduce a turnpike result stating that the solution of the dynamic problem (4.1) is close to the solution of the static problem (4.6) for the majority of the time. On the other hand, we establish a sensitivity result for $(\tilde{x}, \tilde{u}, \tilde{\lambda})$ solving (4.5) stating that the behavior of the perturbations $(\varepsilon_1, \varepsilon_T, \varepsilon_2, \varepsilon_0)$ towards T influences the MPC feedback, i.e., the optimal control on $[0, \tau]$ only negligibly, if $\tau \ll T$.

We will perform this analysis by means of an implicit function theorem. In that context, the derivative of the nonlinear first-order optimality condition will be needed, which in our case (formally) reads

$$L_r''(x, \lambda) = \begin{pmatrix} J_{xx}(x,u) - A''(x)^*\lambda & -\frac{d}{dt} - A'(x)^* \\ 0 & E_T \\ \frac{d}{dt} - A'(x) & -BQ^{-1}B^* \\ E_0 & 0 \end{pmatrix}, \tag{4.9}$$

where $E_t x := x(t)$ for $t \in [0, T]$ and $x \in C(0, T; H)$, cf. Definition 2.26.

To obtain localized estimates in time, we consider a smooth scaling function $s \colon \mathbb{R}^{\geq 0} \to \mathbb{R}$ with $s(t) > 0$ for all $t \in \mathbb{R}^{\geq 0}$. If X is a Banach space, we will make use of the scaled norm

$$\|x\|_{L_p^s(0,T;X)} := \|sx\|_{L_p(0,T;X)} \tag{4.10}$$

for any $1 \leq p \leq \infty$. The equivalence of this norm to the standard $L_p(0, T; X)$-norm follows from the positivity of s as we get for $1 \leq p < \infty$ that

$$\min_{t\in[0,T]} s(t) \left(\int_0^T \|x(t)\|_X^p \, dt \right)^{\frac{1}{p}} \leq \left(\int_0^T \|s(t)x(t)\|_X^p \, dt \right)^{\frac{1}{p}} \leq \max_{t\in[0,T]} s(t) \left(\int_0^T \|x(t)\|_X^p \, dt \right)^{\frac{1}{p}} \tag{4.11}$$

and

$$\min_{t\in[0,T]} s(t) \operatorname*{ess\,sup}_{t\in[0,T]} \|x(t)\|_X \leq \operatorname*{ess\,sup}_{t\in[0,T]} \|s(t)x(t)\|_X \leq \max_{t\in[0,T]} s(t) \operatorname*{ess\,sup}_{t\in[0,T]} \|x(t)\|_X. \tag{4.12}$$

As $L_p(0, T; X)$ with the standard norm is a Banach space, by the equivalence of the norms above, $\left(L_p(0; T; X), \|\cdot\|_{L_p^s(0,T;X)}\right)$ is also a Banach space. Note that the equivalence of norms can (and in our case with, e.g., $s(t) = e^{-\mu t}$ will) deteriorate for $T \to \infty$. The dynamics we will inspect in Section 4.4 will be described by a closed operator $\mathcal{A}_2 : D(\mathcal{A}_2) \subset L_2(\Omega) \to L_2(\Omega)$ that is a generator of an analytic semigroup in $L_2(\Omega)$, where $D(\mathcal{A}_2)$ is the domain of $\mathcal{A}_2$ endowed with the graph norm $\|\cdot\| + \|\mathcal{A}_2 \cdot\|$. More precisely, $\mathcal{A}_2$ will be a second order elliptic differential operator. Further, we will impose either homogeneous Neumann or homogeneous Dirichlet boundary conditions and hence $D(\mathcal{A}_2) = \{v \in H^2(\Omega) \,|\, \frac{\partial}{\partial \nu_A} v = 0\}$, where $\frac{\partial}{\partial \nu_A}$ is a conormal derivative corresponding to $\mathcal{A}_2$, or $D(\mathcal{A}_2) = H^2(\Omega) \cap H_0^1(\Omega)$. We will denote

$$\begin{aligned} W^{1,2}(0, T, D(\mathcal{A}_2), L_2(\Omega)) &:= \{v \in L_2(0, T; D(\mathcal{A}_2)) \,|\, v' \in L_2(0, T; L_2(\Omega))\}, \\ \|v\|_{W^{1,2}(0,T;D(\mathcal{A}_2),L_2(\Omega))} &:= \|v\|_{L_2(0,T;D(\mathcal{A}_2))} + \|v'\|_{L_2(0,T;L_2(\Omega))}, \end{aligned}$$

where the time derivative is meant in a weak sense. For this vector-valued Sobolev space, we will also utilize a scaled norm, i.e.,

$$\|v\|_{W_s^{1,2}(0,T,D(\mathcal{A}_2),L_2(\Omega))} := \|sv\|_{W^{1,2}(0,T,D(\mathcal{A}_2),L_2(\Omega))}.$$

For the scaling terms we have in mind, i.e., exponential functions, one can straightforwardly show that the norm $\|v\|_{W_s^{1,2}(0,T,D(\mathcal{A}_2),L_2(\Omega))} = \|sv\|_{L_2(0,T;D(\mathcal{A}_2))} + \|(sv)'\|_{L_2(0,T;L_2(\Omega))}$ is equivalent to $\|sv\|_{L_2(0,T;D(\mathcal{A}_2))} + \|sv'\|_{L_2(0,T;L_2(\Omega))} = \|v\|_{L_2^s(0,T;D(\mathcal{A}_2))} + \|v'\|_{L_2^s(0,T;L_2(\Omega))}$ if $\mu < 1$. We show in Remark 4.2 that this choice of μ does not constitute a real restriction. By the equivalence of scaled and unscaled L_2-norms shown above, $\|\cdot\|_{W_s^{1,2}(0,T,D(\mathcal{A}_2),L_2(\Omega))}$ is equivalent to the standard norm $\|\cdot\|_{W^{1,2}(0,T,D(\mathcal{A}_2),L_2(\Omega))}$ with constants strongly depending on T. Hence,

$$\left(W^{1,2}(0,T,D(\mathcal{A}_2),L_2(\Omega)), \|\cdot\|_{W_s^{1,2}(0,T,D(\mathcal{A}_2),L_2(\Omega))}\right)$$

is a Banach space.

As the semigroup generated by $\mathcal{A}_2$ is analytic, we have the T-independent continuous embedding $W^{1,2}(0,T,D(\mathcal{A}_2),L_2(\Omega)) \hookrightarrow C(0,T;V)$, cf. [19, Part II-1, Remark 4.1, Remark 4.2], where $V = H^1(\Omega)$ or $V = H_0^1(\Omega)$ depending on the choice of boundary conditions, i.e. $V = D((-\mathcal{A}_2)^{-\frac{1}{2}})$ is the domain of a fractional power of $-\mathcal{A}_2$, cf. Example 2.71 and [90, Section 0.2.1].

Finally, whenever we write $V^{s(t)}$ for $t \in [0,T]$, we mean V endowed with the equivalent norm $s(t)\|\cdot\|_V$. This notation will be used to indicate a scaling of the initial resp. terminal datum.

Remark 4.2. *We briefly show that in the linear case of Chapter 2, we implicitly assumed that $\mu < 1$. In the sensitivity and turnpike results of Theorems 2.27 and 2.30 we chose $\mu > 0$ such that*

$$\mu < \frac{1}{\|M^{-1}\|_{L(L_2(0,T;X)\times X)^2, C(0,T;X)^2)}},$$

where M^{-1} is the solution operator of the extremal equations (2.27). We claim that this directly yields $\mu < 1$. To prove this, we assume that $c_M := \|M^{-1}\|_{L(L_2(0,T;X)\times X)^2, C(0,T;X)^2)} < 1$. Then, setting $l_1 = l_2 = 0$ and for arbitrary $x_0, \lambda_T \in X$ we get the estimate

$$\|\lambda\|_{C(0,T;X)} + \|x\|_{C(0,T;X)} \le c_M\left(\|x_0\| + \|\lambda_T\|\right) < \|x_0\| + \|\lambda_T\|,$$

where $(x,\lambda) \in C(0,T;X)^2$ solve the corresponding extremal equations (2.27). With the simple estimate $\|x_0\| + \|\lambda_T\| = \|x(0)\| + \|\lambda(T)\| \le \|x\|_{C(0,T;X)} + \|\lambda\|_{C(0,T;X)}$ we obtain a contradiction. Hence, we always have $\|M^{-1}\|_{L(L_2(0,T;X)\times X)^2, C(0,T;X)^2)} \ge 1$.

4.2 An abstract framework for sensitivity analysis

In view of the first-order optimality conditions (4.4) and the perturbations (4.5) resp. (4.8), the question we aim to answer is the following: How do $\tilde{z} = (\tilde{x}, \tilde{\lambda})$ and $\bar{z} = (\bar{x}, \bar{\lambda})$ differ from

$z = (x, \lambda)$ depending on $(\varepsilon_1, \varepsilon_T, \varepsilon_2, \varepsilon_0)$ and $(0, \bar{\lambda}, 0, \bar{x} - x_0)$, respectively? In particular, we aim to obtain results localized in time, i.e., estimates in scaled norms as in Theorems 2.27, 2.30, 3.14 and 3.16. In the linear quadratic framework of Chapters 2 and 3, we concluded localized estimates by subtracting the perturbed and unperturbed extremal equations, a scaling result and a bound on the solution operator. In the nonlinear case, we will conclude a local result by means of an implicit function theorem. For its application, we denote the solution space by Z and the perturbation space by E. These spaces will contain the solutions and right hand sides of, e.g., (4.4), i.e., Z contains (x, λ), $(\bar{x}, \bar{\lambda})$ and $(\tilde{x}, \tilde{\lambda})$ and E contains $\big(0, \bar{\lambda}, 0, \bar{x} - x_0\big)$ and $(\varepsilon_1, \varepsilon_T, \varepsilon_2, \varepsilon_0)$. We introduce a nonlinear operator

$$G\colon Z \times E \to E$$

defined by

$$G(z, \varepsilon) := L_r'(z) - \varepsilon \qquad \forall (z, \varepsilon) \in Z \times E. \tag{4.13}$$

It is clear that

- $G(z, 0) = 0$ for any solution $z = (x, \lambda)$ of the dynamic problem (4.4),
- $G(\bar{z}, (0, \bar{\lambda}, 0, \bar{x} - x_0)) = 0$ for any solution $\bar{z} = (\bar{x}, \bar{\lambda})$ of the static problem (4.7),
- $G(\tilde{z}, (\varepsilon_1, \varepsilon_T, \varepsilon_2, \varepsilon_0)) = 0$ for any solution $\tilde{z} = (\tilde{x}, \tilde{\lambda})$ of the perturbed dynamic problem (4.5).

First, in Section 4.3 we apply the abstract approach of this section to finite dimensional problems to highlight the main steps without too many functional analytic overhead. In that context, we will have $V = H = \mathbb{R}^n$ and we will deduce estimates in the scaled spaces

$$\begin{aligned} Z_s &= \big(H^1(0, T; \mathbb{R}^n), \|\cdot\|_{W^s([0,T])}\big)^2, \\ E_s &= \Big(L_2(0, T; \mathbb{R}^n), \|\cdot\|_{L_2^s(0,T;\mathbb{R}^n)}\Big) \times (\mathbb{R}^n)^{s(T)} \times \Big(L_2(0, T; \mathbb{R}^n), \|\cdot\|_{L_2^s(0,T;\mathbb{R}^n)}\Big) \times (\mathbb{R}^n)^{s(0)}. \end{aligned}$$

where $H^1(0, T; \mathbb{R}^n)$ contains all functions $v \in L_2(0, T; \mathbb{R}^n)$ with weak derivative $v' \in L_2(0, T; \mathbb{R}^n)$. This space obviously coincides with $W([0, T])$ if one sets $V = H = \mathbb{R}^n$ in Lemma 3.4.

In Section 4.4 we consider optimal control problems governed by semilinear parabolic evolution equations. In that context, we will utilize the smoothing effect of parabolic equations and obtain sensitivity and turnpike estimates in the scaled spaces

$$\begin{aligned} Z_s &= \Big(L_p(0, T; L_p(\Omega)) \cap W^{1,2}(0, T; D(\mathcal{A}_2), L_2(\Omega)), \|\cdot\|_{L_p^s(0,T;L_p(\Omega)) \cap W_s^{1,2}(0,T;D(\mathcal{A}_2),L_2(\Omega))}\Big)^2, \\ E_s &= \Big(L_2(0; T; L_2(\Omega)), \|\cdot\|_{L_2^s(0,T;L_2(\Omega))}\Big) \times V^{s(T)} \\ &\qquad\qquad \times \Big(L_2(0; T; L_2(\Omega)), \|\cdot\|_{L_2^s(0,T;L_2(\Omega))}\Big) \times V^{s(0)}. \end{aligned}$$

The perturbations of the dynamics are assumed to belong to an L_2-space, whereas the perturbations of the initial values have to belong to V, which will be a H^1-space. This regularity of the data leads to solutions with values a.e. in $D(\mathcal{A}_2)$, a H^2-space, that have a weak time derivative with values a.e. in $L_2(\Omega)$ by maximal parabolic regularity, cf. [19, Part II-1, Section 3].

4.2.1 An implicit function theorem

We now present an implicit function theorem that allows for estimates in scaled norms in a general setting. A particular feature of the following implicit function theorem is the tracking of dependencies of the chosen neighborhoods for perturbations and solutions in scaled and unscaled norms. This allows us to formulate criteria that render these neighborhoods independent of T, namely a T-uniform continuity condition on the linearization and T-uniform invertibility of the operator corresponding to the linearized first-order necessary conditions. This uniformity in T is crucial to derive meaningful turnpike and sensitivity results. The assumption of T-independence of the solution operators norm is also a central assumption in the linear quadratic setting and in that case can be achieved under stabilizability and detectability assumptions, cf. Theorem 5.5 and Corollary 3.30. We will derive a similar property for the linearized system in Section 4.4. Finally we emphasize that even though the scaled and unscaled norms are equivalent, the involved constants in case of exponential scalings strongly depend on T. Thus, this equivalence of norms can not be directly used to derive estimates, motivating a refined analysis as carried out in the following theorem.

Theorem 4.3. *Let $(Z, \|\cdot\|_Z)$ and $(E, \|\cdot\|_E)$ be Banach spaces, let $\|v\|_{Z_s}$ resp. $\|v\|_{E_s}$ be equivalent norms on Z resp. E and set $Z_s := (Z, \|\cdot\|_{Z_s})$ and $E_s := (E, \|\cdot\|_{E_s})$. Consider the mapping $G\colon Z \times E \to E$ defined in (4.13) with $G(z^0, \varepsilon^0) = 0$ for $(z^0, \varepsilon^0) \in Z \times E$. Assume the following:*

i) $G_z(z^0, \varepsilon^0)$ is continuously invertible in $L(Z, E)$.

ii) It holds that

$$\delta_\varepsilon(z^1, z^2) := \frac{\|G(z^1, \varepsilon) - G(z^2, \varepsilon) - G_z(z^0, \varepsilon^0)(z^1 - z^2)\|_E}{\|z^1 - z^2\|_Z} \to 0,$$

if $z^1, z^2 \to z^0$ in Z and $\varepsilon \to \varepsilon^0$ in E.

iii) It holds that

$$\delta^s_\varepsilon(z^1, z^0) := \frac{\|G(z^1, \varepsilon) - G(z^0, \varepsilon) - G_z(z^0, \varepsilon^0)(z^1 - z^0)\|_{E_s}}{\|z^1 - z^0\|_{Z_s}} \to 0,$$

if $z^1 \to z^0$ in Z and $\varepsilon \to \varepsilon^0$ in E.

Then there is $r_E, r_Z \geq 0$, such that for every $\varepsilon \in E$ satisfying $\|\varepsilon - \varepsilon^0\|_E \leq r_E$ there exists $z^*(\varepsilon) \in Z$ such that $\|z^*(\varepsilon) - z^0\|_Z \leq r_Z$ and $G(z^*, \varepsilon) = 0$. Further, we have the estimate

$$\|z^*(\varepsilon) - z^0\|_{Z_s} \leq c\|\varepsilon - \varepsilon^0\|_{E_s}. \tag{4.14}$$

Moreover we have the following T-uniformity:

- *If the convergence in ii) is uniform in T and $\|G_z(z^0, \varepsilon^0)^{-1}\|_{L(E,Z)}$ is bounded independently of T, then r_Z and r_E can be chosen independently of T.*

- *If, additionally, the convergence of iii) is uniform in T and $\|G_z(z^0,\varepsilon^0)^{-1}\|_{L(E_s,Z_s)}$ is bounded independently of T, then the constant in (4.14) is independent of T.*

Proof. Throughout this proof, we denote $B^Z_{r_Z}(z^0) := \{z \in Z \mid \|z - z^0\|_Z \leq r_Z\}$ and $B^E_{r_E}(\varepsilon^0) := \{\varepsilon \in E \mid \|\varepsilon - \varepsilon^0\|_E \leq r_E\}$. For $k \in \mathbb{N}^0$ let $\delta z^k := -G_z(z^0,\varepsilon^0)^{-1}G(z^k,\varepsilon)$ and $z^{k+1} = z^k + \delta z^k$. As

$$\delta z^{k+1} = -G_z(z^0,\varepsilon^0)^{-1}\left(G(z^{k+1},\varepsilon) - G(z^k,\varepsilon) - G_z(z^0,\varepsilon^0)(z^{k+1} - z^k)\right)$$

we have with ii) that

$$\|\delta z^{k+1}\|_Z \leq \|G_z(z^0,\varepsilon^0)^{-1}\|_{L(E,Z)}\delta_\varepsilon(z^{k+1},z^k)\|\delta z^k\|_Z \tag{4.15}$$

where $\delta_\varepsilon(z^{k+1},z^k) \to 0$, if $z^{k+1}, z^k \to z^0$ in Z and $\varepsilon \to \varepsilon^0$ in E. We now choose $r_Z > 0$ and $r_E > 0$ such that $\delta_\varepsilon(z_1,z_2) \leq \frac{1}{2\|G_z(z^0,\varepsilon^0)^{-1}\|_{L(E,Z)}}$ for all $z_1, z_2 \in B^Z_{r_Z}(z^0)$ and $\varepsilon \in B^E_{r_E}(\varepsilon^0)$. Further, by continuity of $G(z^0,\varepsilon)$ in ε, continuous invertibility of $G(z^0,\varepsilon^0)$ and as $G(z^0,\varepsilon^0) = 0$, we can further decrease r_E such that

$$\|\delta z^0\|_Z = \|G_z(z^0,\varepsilon^0)^{-1}G(z^0,\varepsilon)\|_Z \leq \frac{r_Z}{2}$$

for all $\varepsilon \in B^E_{r_E}(\varepsilon^0)$. Thus, we get

$$\|\delta z^{k+1}\|_Z \leq \|G_z(z^0,\varepsilon^0)^{-1}\|_{L(E,Z)}\delta_\varepsilon(z^{k+1},z^k)\|\delta z^k\|_Z \leq \left(\frac{1}{2}\right)^k \|\delta z^0\|_Z. \tag{4.16}$$

Hence,

$$\|z^{k+1} - z^0\|_Z \leq \sum_{i=0}^{k} \|\delta z^k\|_Z \leq \frac{1}{1-\frac{1}{2}}\|\delta z^0\|_Z \leq r_Z \tag{4.17}$$

and hence inductively, $z^k \in B^Z_{r_Z}(z^0)$ for all $k \in \mathbb{N}$. Thus, by completeness of Z, the iteration $z^k = z^0 + \sum_{i=0}^k \delta z^k$ converges to an element $z^* \in B^Z_{r_Z}(z^0)$ and as $\delta_\varepsilon(z^*,z^k) \leq \frac{1}{2\|G_z(z^0,\varepsilon^0)^{-1}\|_{L(E,Z)}}$ we get

$$\|G_z(z^0,\varepsilon^0)^{-1}\left(G(z^*,\varepsilon) - G(z^k,\varepsilon) - G_z(z^0,\varepsilon^0)(z^* - z^k)\right)\|_Z \leq \frac{1}{2}\|z^* - z^k\|_Z.$$

Hence, by the reverse triangle inequality, we get

$$\|G_z(z^0,\varepsilon^0)^{-1}G(z^*,\varepsilon)\|_Z \leq \|\delta z^k + z^* - z^k\|_Z + \frac{1}{2}\|z^* - z^k\|_Z \to 0$$

for $k \to \infty$ and thus $G(z^*,\varepsilon) = 0$. To obtain an estimate in the scaled norms, we compute, using $z^1 = z^0 + \delta z^0$ that

$$\|z^* - z^0\|_{Z_s} \leq \|z^* - z^1\|_{Z_s} + \|\delta z^0\|_{Z_s}.$$

We further estimate with $\delta z^0 = -G_z(z^0,\varepsilon^0)^{-1}G(z^0,\varepsilon)$, by *iii)* and $z^* \in B^Z_{r_Z}(z^0)$ after possibly further decreasing r_Z and r_E such that $\delta^s_\varepsilon(z,z^0) \leq \frac{1}{2\|G_z(z^0,\varepsilon^0)^{-1}\|_{L(E_s,Z_s)}}$ for all $z \in B^Z_{r_Z}(z^0)$ and $\varepsilon \in B^E_{r_E}(\varepsilon^0)$ that

$$\begin{aligned}\|z^* - z^1\|_{Z_s} &= \|G_z(z^0,\varepsilon^0)^{-1}(G_z(z^0,\varepsilon^0)(z^* - z^1))\|_{Z_s}\\ &= \|G_z(z^0,\varepsilon^0)^{-1}(G_z(z^0,\varepsilon^0)(z^* - z^0) - (G(z^*,\varepsilon) - G(z^0,\varepsilon)))\|_{Z_s}\\ &\leq \frac{1}{2}\|z^* - z^0\|_{Z_s}.\end{aligned}$$

Hence by the particular structure of G, i.e., $G(z^0,\varepsilon) = G(z^0,\varepsilon^0) + \varepsilon - \varepsilon^0 = \varepsilon - \varepsilon^0$ we obtain

$$\frac{1}{2}\|z^* - z^0\|_{Z_s} \leq \|\delta z^0\|_{Z_s} = \|G_z(z^0,\varepsilon^0)^{-1}G(z^0,\varepsilon)\|_{Z_s} \leq \|G_z(z^0,\varepsilon^0)^{-1}\|_{L(E_s,Z_s)}\|\varepsilon - \varepsilon^0\|_{E_s}$$

which concludes the proof. □

We have two particular applications of Theorem 4.3 in mind. First, to derive a turnpike result, we set $z^0 = (\bar{x},\bar{\lambda})$ solving the static extremal equations (4.8), $\varepsilon^0 = (0,\bar{\lambda},0,\bar{x} - x_0)$, and $\varepsilon = 0$ to derive an estimate on the difference of (x,λ) and $(\bar{x},\bar{\lambda})$ in scaled norms with scaling function $s(t) = \frac{1}{e^{-\mu t} + e^{-\mu(T-t)}}$. Second, in order to obtain a sensitivity result, we set $z^0 = (x,\lambda)$ solving the exact dynamic extremal equations (4.4), $\varepsilon^0 = (0,0)$, and $\varepsilon = (\varepsilon_1,\varepsilon_T,\varepsilon_2,\varepsilon_0)$ to derive an estimate on the difference of (x,λ) and $(\tilde{x},\tilde{\lambda})$ solving the perturbed extremal equations (4.5) in scaled norms with scaling function $s(t) = e^{-\mu t}$.

Remark 4.4. *Due to its generality, Theorem 4.3 can also be applied to general evolution equations, i.e., hyperbolic equations. Moreover, we could apply it to elliptic PDEs to prove an exponential decay property of the influence of right-hand sides in space, a well-known property for elliptic equations, without knowledge of Green's function.*

A crucial point in the proof of the implicit function theorem, i.e., Theorem 4.3, is to ensure that the series generated by $G_z(z^0,\varepsilon^0)^{-1}G(z^k,\varepsilon)$ converges in Z. In the assumptions of the theorem, this is ensured by i) and ii), i.e., differentiability of the nonlinear operator and continuous invertibility of the linearization. As we will see in the following section, in general, the image of a nonlinear map, e.g., $G(z^k,\varepsilon)$ has lower integrability than its argument z^k. Thus, it is necessary to prove a smoothing effect of the solution operator to the linearized problem, e.g., $G_z(z^0,\varepsilon^0)^{-1}$ to make up for this loss of regularity.

4.2.2 Superposition operators and T-uniform continuity

In order to rigorously verify assumptions ii)-iii) in Theorem 4.3, we employ the concept of superposition operators. We will only consider continuity and differentiability of these operators in L_p-spaces and the reader is referred to [138, Section 4.3.3] for a short introduction and [8, 56] for an in-depth treatment of these topics in Sobolev and Lebesgue spaces of abstract functions. Intuitively, a superposition operator is a nonlinear map between function spaces defined via an, e.g., scalar nonlinear function by superposition. The following definition of a superposition operator is adapted from [138, Section 4.3.1] and [56, Section 2]. In this work we only consider the case of a nonlinearity depending on one argument. A generalization of the presented results to nonlinearities that additionally depend on space and time is straightforward. We consider a measurable subset $S \subset \mathbb{R}^n$ with $n \in \mathbb{N}$, which serves as a placeholder for the spatial domain Ω or the temporal domain $[0,T]$. The following definition of a superposition operator is adapted from [138, Section 4.3.1] and [56, Section 2].

Definition 4.5 (Superposition operator)**.** *Let W_1 and W_2 be real valued Banach spaces. Consider a mapping $\varphi\colon W_1 \to W_2$. Then the mapping Φ defined by*

$$\Phi(x)(s) = \varphi(x(s)) \qquad \text{for } s \in S$$

assigns to an (abstract) function $x\colon S \to W_1$ a new (abstract) function $z\colon S \to W_2$ via the relation $z(s) = \varphi(x(s))$ for $s \in S$ and is called an (abstract) Nemytskij operator or (abstract) superposition operator.

We will briefly illustrate this definition by an example.

Example 4.6. *Consider $W_1 = W_2 = \mathbb{R}$ and $S = \Omega \subset \mathbb{R}^n$ bounded with $n \in \mathbb{N}$. Then the nonlinear function $\varphi\colon \mathbb{R} \to \mathbb{R}$, $\varphi(w) = w^3$ defines a superposition operator $\underline{\Phi}$ via the relation*

$$\underline{\Phi}(x)(\omega) = x(\omega)^3 \quad \text{for } \omega \in \Omega.$$

for $x\colon \Omega \to \mathbb{R}$. An immediate question that arises is the following: Given a function $x \in L_p(\Omega)$, which integrability does the image $\underline{\Phi}(x)$ have? We will provide an answer to this question in Proposition 4.7. In this example, it is intuitively clear that, e.g., $\underline{\Phi}\colon L_6(\Omega) \to L_2(\Omega)$.

Consider now $T > 0$ and the nonlinear function $\underline{\Phi}\colon L_6(\Omega) \to L_2(\Omega)$ defined above. Setting $W_1 = L_6(\Omega)$ and $W_2 = L_2(\Omega)$ and $S = [0,T]$ in Definition 4.5, we define a second superposition operator Φ for $x\colon [0,T] \to L_6(\Omega)$ via the relation

$$\Phi(x)(t) = \underline{\Phi}(x(t)) \quad \text{for } t \in [0,T].$$

As we will see later, this map is well defined and continuous as a mapping $\Phi\colon L_6(0,T;L_6(\Omega)) \to L_2(0,T;L_2(\Omega))$. We thus obtained from a scalar nonlinear function a nonlinear mapping from one space of abstract functions into another one by applying Definition 4.5 twice.

We will now discuss continuity and differentiability of nonlinear superposition operators. As to be expected and if $p < \infty$, growth conditions on the underlying function φ play a key role in establishing these properties in L_p-spaces. In the following, we give sufficient and necessary conditions for continuity and differentiability of superposition operators. However, we first characterize the image of a superposition operator under growth and boundedness conditions.

Proposition 4.7. *Let W_1 and W_2 be real valued Banach spaces. Let $\varphi\colon W_1 \to W_2$ be continuous. For $1 \le p, q < \infty$ let*

$$\|\varphi(w)\|_{W_2} \le c_1 + c_2 \|w\|_{W_1}^{\frac{p}{q}} \qquad \forall\, w \in W_1 \tag{4.18}$$

for constants $c_1 \in \mathbb{R}$ and $c_2 \geq 0$. Then the corresponding superposition operator maps $L_p(S; W_1)$ into $L_q(S; W_2)$.

If for all $c > 0$ there is a constant $\beta = \beta(c) \geq 0$ such that

$$\|\varphi(w)\|_{W_2} \leq \beta \qquad \forall\, w \in W_1 : \|w\|_{W_1} \leq c,$$

then the corresponding superposition operator maps $L_\infty(S; W_1)$ into $L_q(S; W_2)$ for all $1 \leq q < \infty$.

Proof. See [56, Theorem 1]. □

The following proposition shows that if a superposition operator maps one L_p-space into another, continuity can be derived immediately.

Proposition 4.8 (Continuity of superposition operators)**.** *Let W_1 and W_2 be real valued Banach spaces. Let $\varphi\colon W_1 \to W_2$ be continuous and $1 \leq p \leq \infty$, $1 \leq q < \infty$. If the induced superposition operator Φ maps $L_p(S; W_1)$ into $L_q(S, W_2)$, then it is continuous. If φ is locally Lipschitz continuous from W_1 to W_2, then the induced superposition operator Φ is continuous from $L_\infty(S; W_1)$ to $L_\infty(S, W_2)$.*

Proof. For the first part, see [56, Theorem 4]. For the case $p = q = \infty$ we refer to [138, Lemma 4.11]. □

We note that continuity in case of $p = q = \infty$ can also be deduced under a uniform continuity assumption on bounded sets, cf. [56, Theorem 5].

We briefly discuss Proposition 4.8 for the example of $\underline{\Phi}$ and Φ defined in Example 4.6 via the nonlinear function $\varphi(w) = w^3$. Using Proposition 4.7 and Proposition 4.8, the operator $\underline{\Phi}$ defined in Example 4.6 is continuous as a mapping from $L_{3q}(\Omega)$ to $L_q(\Omega)$ and $L_\infty(\Omega)$ to $L_q(\Omega)$ for $1 \leq q < \infty$, respectively. Additionally, Φ is continuous as a mapping from $L_{3q_1}(0, T; L_{3q_2}(\Omega))$ to $L_{q_1}(0, T; L_{q_2}(\Omega))$ for $1 \leq q_1, q_2 < \infty$, from $L_\infty(0, T; L_{3q_2}(\Omega))$ to $L_{q_1}(0, T; L_{q_2}(\Omega))$ for $1 \leq q_1, q_2 < \infty$, and from $L_\infty(0, T; L_\infty(\Omega))$ to $L_{q_1}(0, T; L_{q_2}(\Omega))$ for $1 \leq q_1, q_2 < \infty$.

As it turns out, the conditions stated in Proposition 4.7 are not only sufficient but also necessary for continuity of the induced superposition operator, cf. [56, Theorem 3].

Next, we focus on the topic of differentiability, which plays a key role in applying the implicit function theorem. The following result obtained in [56, Theorem 7] gives sufficient conditions for Fréchet differentiability.

Proposition 4.9 (Differentiability of superposition operators)**.** *Let $1 \leq q < p < \infty$. Assume that $\varphi\colon W_1 \to W_2$ is continuously Fréchet differentiable. Moreover, let the superposition operator defined by*

$$\Psi(x)(s) = \varphi'(x(s)) \quad \textit{for } s \in S$$

be continuous from $L_p(S; W_1)$ to $L_r(S; L(W_1, W_2))$ with $r = \frac{pq}{p-q}$. Then the superposition operator Φ induced by φ is continuously Fréchet differentiable and the Fréchet derivative

$$\Phi'\colon L_p(S; W_1) \to L(L_p(S; W_1), L_q(S; W_2))$$

is given by Ψ*, i.e.,*

$$(\Phi'(x)\delta x)(s) = \Psi(x)(s)\delta x(s) \quad \text{for } s \in S,\ \delta x \in L_p(S; W_1).$$

The conditions given in Proposition 4.9 are also necessary in the following sense: If a superposition operator is differentiable from $L_p(\Omega)$ to $L_q(\Omega)$ with $1 \leq p = q < \infty$, then it is affine linear. If it is differentiable from $L_p(\Omega)$ to $L_q(\Omega)$ with $1 \leq p < q \leq \infty$, then it has to be constant, cf. the discussion in [56, Section 3.1] and [8, Theorem 3.12].

Remark 4.10. *One can easily check that if* $\varphi(w)$ *is a polynomial of the form* $\varphi(w) = w^d$*,* $d \in \mathbb{N}^+$*, then it induces a continuous superposition operator* $\underline{\Phi}\colon L_{dq}(\Omega) \to L_q(\Omega)$ *via Proposition 4.8. Additionally,* $\underline{\Phi}$ *is differentiable from* $L_{dq}(\Omega)$ *to* $L_q(\Omega)$ *for all* $1 \leq q < \infty$ *by applying Proposition 4.9 with* $r = \frac{d}{(d-1)}$*. The same obviously carries over to the vector valued setting, e.g.,* $\Phi\colon L_{dq_1}(0,T; L_{dq_2}(\Omega)) \to L_{q_1}(0,T; L_{q_2}(\Omega))$ *induced by* φ *is continuous and differentiable for all* $1 \leq q_1, q_2 < \infty$*.*

In order to render the radii r_Z and r_E and the estimate (4.14) independent of T, we have to discuss the T-dependence of continuity moduli of superposition operators in unscaled and scaled L_p-spaces as introduced at the end of Section 4.1 with norms defined in (4.10).

Definition 4.11 (T-uniform continuity)**.** *Let* W_1, W_2 *be real-valued Banach spaces. We say that an operator* $\Psi\colon L_p(0,T; W_1) \to L_q(0,T; W_2)$ *is* T*-uniformly continuous if for all* $x^0 \in L_p(0,T; W_1)$ *and for all* $\varepsilon > 0$ *there is* $\delta > 0$ *independent of* T *such that if* $\|\delta x\|_{L_p(0,T;W_1)} < \delta$ *then*

$$\|\Psi(x^0 + \delta x) - \Psi(x^0)\|_{L_q(0,T;W_2)} < \varepsilon.$$

Lemma 4.12. *If the constants* c_1 *and* c_2 *in the growth condition (4.18) can be chosen independent of* T*, then the continuity of the induced superposition operator is* T*-uniform.*

Proof. The proof follows directly by the fact that the references establishing continuity under growth conditions do not assume the domain S to be bounded, cf. [8, Chapter 3] and [56]. □

Example 4.13 (Remark 4.10 revisited)**.** *We briefly illustrate the previous lemma by means of the example* $\varphi(w) = w^d$*,* $d \in \mathbb{N}^+$*. In that case it is clear that the growth condition (4.18), i.e.,*

$$|\varphi(w)| \leq c_1 + c_2|w|^{\frac{p}{q}}$$

holds with $p = dq$*,* $c_1 = 0$ *and* $c_2 = 1$*, i.e.,* φ *induces a* T*-uniformly continuous superposition operator from* $L_{dq}(0,T; L_{dp}(\Omega))$ *to* $L_q(0,T; L_p(\Omega))$ *for all* $1 \leq q, p < \infty$*. As* $[0,T]$ *and* Ω *are bounded, one can show that continuity also holds from* $L_{\hat{d}q}(0,T; L_{\hat{d}p}(\Omega))$ *to* $L_q(0,T; L_p(\Omega))$ *for* $\hat{d} > d$*, however, with constants that depend on* T *and* $|\Omega|$*. This means, that the functional analytic framework has to be chosen particularly suited to the nonlinearity to render the constants and hence the continuity uniform in* T*.*

The following lemma shows that if a superposition operator has a T-uniformly continuous Fréchet derivative, the convergence in ii) and iii) of Theorem 4.3 can be shown to be T-uniform.

Lemma 4.14. *Let W_1 and W_2 be Banach spaces, $1 \leq p \leq \infty$ and let $\Phi\colon L_p(0,T;W_1) \to L_q(0,T;W_2)$ have a T-uniformly continuous Fréchet derivative Φ'. Then,*

$$\frac{\|\Phi(x^1) - \Phi(x^2) - \Phi'(x^0)(x^1 - x^2)\|_{L_q(0,T;W_2)}}{\|x^1 - x^2\|_{L_p(0,T;W_1)}} \to 0$$

uniformly in T if $x^1, x^2 \to x^0$ in $L_p(0,T;W_1)$. Moreover,

$$\frac{\|\Phi(x^0 + \delta x) - \Phi(x^0) - \Phi'(x^0)\delta x\|_{L_q^s(0,T;W_2)}}{\|\delta x\|_{L_p^s(0,T;W_1)}} \to 0$$

uniformly in T if $\delta x \to 0$ in $L_p(0,T;W_1)$.

Proof. We compute with the fundamental theorem of calculus, cf. [82, p.51], that

$$\begin{aligned}
&\|\Phi(x^1) - \Phi(x^2) - \Phi'(x^0)(x^1 - x^2)\|_{L_q(0,T;W_2)} \\
&\quad = \|\int_0^1 \Phi'(x^2 + \theta(x^1 - x^2)) - \Phi'(x^0)\, d\theta (x^1 - x^2)\|_{L_q(0,T;W_2)} \\
&\quad \leq \Big(\sup_{\theta \in [0,1]} \|\Phi'(x^2 + \theta(x^1 - x^2)) - \Phi'(x^2)\|_{L(L_p(0,T;W_1), L_q(0,T;W_2))} \\
&\qquad\qquad + \|(\Phi'(x^0) - \Phi'(x^2))\|_{L(L_p(0,T;W_1), L_q(0,T;W_2))} \Big) \|x^1 - x^2\|_{L_p(0,T;W_1)}.
\end{aligned}$$

The first claim follows by T-uniform continuity of Φ'. For the second claim in scaled norms with scaling function s, we compute analogously

$$\begin{aligned}
&\|\Phi(x^0 + \delta x) - \Phi(x^0) - \Phi'(x^0)\delta x\|_{L_q^s(0,T;W_2)} \\
&\qquad = \|s\left(\Phi(x^0 + \delta x) - \Phi(x^0) - \Phi'(x^0)\delta x\right)\|_{L_q(0,T;W_2)} \\
&\qquad \leq \sup_{\theta \in [0,1]} \|\Phi'(x^0) - \Phi'(x^0 + \theta \delta x)\|_{L(L_p(0,T;W_1), L_q(0,T;W_2))} \|s \delta x\|_{L_p(0,T;W_1)} \\
&\qquad = \sup_{\theta \in [0,1]} \|\Phi'(x^0) - \Phi'(x^0 + \theta \delta x)\|_{L(L_p(0,T;W_1), L_q(0,T;W_2))} \|\delta x\|_{L_p^s(0,T;W_1)},
\end{aligned}$$

which concludes the proof. □

Hence, it turns out that whenever the superposition operators are differentiable with T-uniformly continuous derivative, the uniform convergence needed in Theorem 4.3 ii) and iii) to obtain T-uniform neighborhoods holds true. The remaining task in order to apply the implicit function theorem is to verify the T-uniform estimate on the solution operator of the linearized first-order optimality system, i.e., $G_z(x^0, \varepsilon^0)^{-1}$, in unscaled and scaled spaces. In the following we will derive such a bound for a wide class of nonlinear finite dimensional and semilinear infinite dimensional parabolic problems. In that context, our aim is to provide maximal flexibility in the norms of this estimate in order match the functional analytic framework where one established T-uniform continuity, cf. Example 4.13 on the importance of this issue.

4.3 Nonlinear finite dimensional problems

We briefly discuss the case of a finite dimensional system as a particular case of a parabolic problem. To this end, we consider $V = H = \mathbb{R}^n$, $\bar{A}\colon \mathbb{R}^n \to \mathbb{R}^n$ twice continuously Fréchet differentiable , $U = \mathbb{R}^m$, and $B \in L(\mathbb{R}^m, \mathbb{R}^n)$ for $n, m \in \mathbb{N}$. In this case, we have the solution space $W([0,T]) = \{v \in L_2(0,T;\mathbb{R}^n) \,|\, v' \in L_2(0,T;\mathbb{R}^n)\} =: H^1(0,T;\mathbb{R}^n)$. Further we have $H^1(0,T;\mathbb{R}^n) \hookrightarrow C(0,T;\mathbb{R}^n)$ with an embedding constant independent of T, cf. Lemma 3.4. Very similarly to Theorems 3.14 and 3.16 we will deduce a T-independent bound on the linearized extremal equations solution operator under a stabilizability and detectability assumption. To this end, we consider the linearized extremal equations at two different linearization points. First, to deduce a turnpike result, we linearize the extremal equations at the optimal steady state $(\bar{x}, \bar{\lambda}) \in (\mathbb{R}^n)^2$, leading to an autonomous linearization $L_r''(\bar{x}, \bar{\lambda})$. Second we analyze the linearization of the extremal equations at the dynamic solution $(x, \lambda) \in H^1(0,T;\mathbb{R}^n)^2$ of (4.4) to derive a sensitivity estimate. As in that case the linearized system governed by $L_r''(x, \lambda)$ is non-autonomous, we consider the stabilizability notion introduced for non-autonomous systems in Chapter 3, namely V-exponential stabilizability.

4.3.1 A T-independent bound for the solution operator

First, we introduce an important square root property for the second derivative of the reduced Lagrangian.

Lemma 4.15. *Let (x^0, λ^0) solve either the steady state problem (4.8) or the dynamic problem (4.4). Assume that $\bar{A}''$ and $\bar{J}''$ are continuous. Suppose that*

$$(L_r)_{xx}(x^0, \lambda^0)(t) = \bar{J}_{xx}(x^0(t)) - \bar{A}''(x^0(t))^T \lambda^0(t) \geq 0$$

for $t \in [0,T]$. Then, there is a self-adjoint $C \in L(L_2(0,T;\mathbb{R}^n))$ for all $1 \leq p \leq \infty$ such that

$$(L_r)_{xx}(x^0, \lambda^0) = C^2. \tag{4.19}$$

Proof. As x^0 and λ^0 are continuous in time, and by continuity of $\bar{A}''$ and $\bar{J}_{xx}$, $(L_r)_{xx}(x^0, \lambda^0) \in C(0,T;\mathbb{R}^n)$. By this continuity and the symmetry of $(L_r)_{xx}(x^0, \lambda^0)(t)$ for each $t \in [0,T]$ the result follows by concatenating the pointwise matrix square roots. For details on square roots of matrices, see [92, Chapter 10]. □

The assumption of $(L_r)_{xx}(x^0, \lambda^0)$ being positive semidefinite is rather unusual, compared to, e.g., second-order sufficient conditions, where one has only the positive definiteness for all directions satisfying the dynamics. However, it is crucial to analyze the stability of the linearized system, cf. also [136, Remark 6].

Theorem 4.16. *Let the assumptions of Lemma 4.15 hold. Let $(\bar{x}, \bar{\lambda}) \in (\mathbb{R}^n)^2$ solve the steady state problem (4.6). Consider C to be defined as in (4.19). Assume $(\bar{A}'(\bar{x}), C)$ is exponentially* *detectable and $(\bar{A}'(\bar{x}), \bar{B})$ is exponentially stabilizable in the classical sense of Assumption 2.32. Then there is a constant $c \geq 0$ independent of T such that*

$$\|L_r''(\bar{x}, \bar{\lambda})^{-1}\|_{L((L_2(0,T;\mathbb{R}^n)\times(\mathbb{R}^n))^2, H^1(0,T;\mathbb{R}^n)^2)} \leq c.$$

Additionally, for any $0 < \mu < 1$ satisfying

$$\mu < \frac{1}{\|L_r''(\bar{x}, \bar{\lambda})^{-1}\|_{L((L_2(0,T;\mathbb{R}^n)\times\mathbb{R}^n)^2, H^1(0,T;\mathbb{R}^n)^2)}},$$

there is a constant $c \geq 0$ independent of T such that setting $s(t) = \frac{1}{e^{-\mu T} + e^{-\mu(T-t)}}$ it holds that

$$\|L_r''(\bar{x}, \bar{\lambda})^{-1}\|_{L((L_2^s(0,T;\mathbb{R}^n)\times(\mathbb{R}^n)^{s(T)}\times L_2^s(0,T;\mathbb{R}^n)\times(\mathbb{R}^n)^{s(0)}),(H^1(0,T;\mathbb{R}^n),\|\cdot\|_{W^s([0,T])})^2)} \leq c.$$

Let $(x, \lambda) \in H^1(0,T;\mathbb{R}^n)^2$ solve the dynamic problem (4.4). Assume $(A'(x), C)$ and $(A'(x), B)$ are $\mathbb{R}^n$-exponentially stabilizable in the sense of Definition 3.20. Then there is a constant $c \geq 0$ independent of T such that

$$\|L_r''(x, \lambda)^{-1}\|_{L((L_2(0,T;\mathbb{R}^n)\times\mathbb{R}^n)^2, H^1(0,T;\mathbb{R}^n)^2)} \leq c.$$

Additionally, for any $0 < \mu < 1$ satisfying

$$\mu < \frac{1}{\|L_r''(x, \lambda)^{-1}\|_{L((L_2(0,T;\mathbb{R}^n)\times\mathbb{R}^n)^2, H^1(0,T;\mathbb{R}^n)^2)}},$$

there is a constant $c \geq 0$ independent of T such that setting $s(t) = e^{-\mu t}$ it holds that

$$\|L_r''(x, \lambda)^{-1}\|_{L((L_2^s(0,T;\mathbb{R}^n)\times(\mathbb{R}^n)^{s(T)}\times L_2^s(0,T;\mathbb{R}^n)\times(\mathbb{R}^n)^{s(0)}),(H^1(0,T;\mathbb{R}^n),\|\cdot\|_{W^s([0,T])})^2)} \leq c.$$

Proof. The proof of the T-independent bound on $L_r''(\bar{x}, \bar{\lambda})^{-1}$ follows completely analogously to Theorem 2.38. In that context, after having bounded the $L_2(0,T;\mathbb{R}^n)$-norm of state and adjoint, the bound on the derivative of state and adjoint can be derived analogously to the proof of Theorem 3.29. The bound in the scaled spaces follows completely analogously to Theorem 2.30.

Similarly, the proof of the T-independent bound on $L_r''(x, \lambda)^{-1}$ can be derived by as a particular case of Corollary 3.30 and Theorem 3.14. □

4.3.2 Exponential sensitivity and turnpike results

Having established T-independent invertibility of $L_r''(\bar{x}, \bar{\lambda})$ and $L_r''(x, \lambda)$ in scaled and unscaled spaces, it is crucial to ensure that L_r'' is the Fréchet derivative of L_r' in the corresponding functional analytic setting such that Theorem 4.3 ii) and iii) are satisfied.

Proposition 4.17. *Let $2 \leq p_1, p_2 < \infty$ and the assumptions of Theorem 4.16 hold. Assume that $J_x(x, u)$ induces a superposition operator from $L_{p_1}(0,T;\mathbb{R}^n)$ to $L_2(0,T;\mathbb{R}^n)$ with T-uniformly continuous Fréchet derivative. Further let $A(x)$ induce a superposition operator from $L_{p_2}(0,T;\mathbb{R}^n)$ to $L_2(0,T;\mathbb{R}^n)$ with two T-uniformly continuous Fréchet derivatives. Then $G(z, \varepsilon) := L_r'(z) - \varepsilon$ with L_r' defined in (4.4) satisfies the assumptions of Theorem 4.3 uniform in T with either* *i)* $s(t) = \frac{1}{e^{-\mu t}+e^{-\mu(T-t)}}$ *and* $(z^0, \varepsilon^0) = ((\bar{x}, \bar{\lambda}), (0, \bar{\lambda}, 0, \bar{x} - x_0))$ *for any* $0 < \mu < 1$ *satisfying*

$$\mu < \frac{1}{\|L_r''(\bar{x}, \bar{\lambda})^{-1}\|_{L((L_2(0,T;\mathbb{R}^n)\times\mathbb{R}^n)^2, H^1(0,T;\mathbb{R}^n)^2)}},$$

or

ii) $s(t) = e^{-\mu t}$ *and* $(z^0, \varepsilon^0) = ((x, \lambda), 0)$ *for any* $0 < \mu < 1$ *satisfying*

$$\mu < \frac{1}{\|L_r''(x, \lambda)^{-1}\|_{L((L_2(0,T;\mathbb{R}^n)\times\mathbb{R}^n)^2, H^1(0,T;\mathbb{R}^n)^2)}},$$

with the spaces $Z = H^1(0, T; \mathbb{R}^n)$ *and* $E = ((L_2(0, T; \mathbb{R}^n) \times \mathbb{R}^n)^2$ *and with the scaled norms* $\|\cdot\|_{Z_s} = \|\cdot\|_{W^s([0,T])^2}$ *and* $\|\cdot\|_{E_s} = \|\cdot\|_{L_2^s(0,T;\mathbb{R}^n)\times(\mathbb{R}^n)^{s(T)}\times L_2^s(0,T;\mathbb{R}^n)\times(\mathbb{R}^n)^{s(0)}}$.

Proof. First, we observe that the assumptions of Lemma 4.14 with $W_1 = W_2 = \mathbb{R}^n$ are satisfied for $\Phi = J_x(x, u)$ with $p = p_1$ and $q = 2$ and for $\Phi = A(x)$ resp. $\Phi = A'(x)^T\lambda$ for $p = p_2$ and $q = 2$, respectively. Further, with the T-independent embedding $H^1(0, T; \mathbb{R}^n) \hookrightarrow L_p(0, T; \mathbb{R}^n)$ for all $2 \le p < \infty$ independently of T, cf. [5, Theorem 3], the assumptions ii) and iii) of Theorem 4.3 follow. The T-independent bounds on the solution operators L_r'' at either linearization point follow directly from Theorem 4.16. □

This result now directly implies a turnpike and sensitivity result for the nonlinear system via the implicit function theorem of Theorem 4.3. Note that the implicit function theorem provides the estimates for state and adjoint, whereas the estimates for the control follow straightforwardly by the relation $\delta u(t) = Q^{-1}B^*\delta\lambda(t)$ for a.e. $t \in [0, T]$ for either $\delta u(t) = u(t) - \bar{u}$ and $\delta\lambda(t) = \lambda(t) - \bar{\lambda}$ (distance to the turnpike) or $\delta u(t) = \tilde{u}(t) - u(t)$ and $\delta\lambda(t) = \tilde{\lambda}(t) - \lambda(t)$ (absolute error for perturbed system), cf. the elimination of the control after (4.5).

Corollary 4.18. *Suppose the assumptions of Proposition 4.17 hold. Let* (x, u, λ) *solve the nonlinear dynamic problem* (4.3) *and* $(\bar{x}, \bar{u}, \bar{\lambda})$ *the nonlinear static problem* (4.7). *Define* $(\delta x, \delta u, \delta\lambda) := (x - \bar{x}, u - \bar{u}, \lambda - \bar{\lambda})$. *Then there is a radius* $r_E > 0$ *independent of* T *such that, if*

$$\|x_0 - \bar{x}\| + \|\bar{\lambda}\| \le r_E,$$

then for any $0 < \mu < 1$ *satisfying*

$$\mu < \frac{1}{\|L_r''(\bar{x}, \bar{\lambda})\|_{L((L_2(0,T;\mathbb{R}^n)\times\mathbb{R}^n)^2, H^1(0,T;\mathbb{R}^n)^2)}}$$

there is a constant $c \ge 0$ *independent of* T *such that we have*

$$\left\|\frac{1}{e^{-\mu t} + e^{-\mu(T-t)}}\delta x(t)\right\|_{H^1(0,T;\mathbb{R}^n)} + \left\|\frac{1}{e^{-\mu t} + e^{-\mu(T-t)}}\delta u(t)\right\|_{L_\infty(0,T;\mathbb{R}^m)} + \left\|\frac{1}{e^{-\mu t} + e^{-\mu(T-t)}}\delta\lambda(t)\right\|_{H^1(0,T;\mathbb{R}^n)} \le cr_E.$$

Corollary 4.19. *Let the assumptions of Proposition 4.17 hold. Let (x, λ) solve the extremal equations (4.4) and $(\tilde{x}, \tilde{\lambda})$ the perturbed extremal equations (4.5). Define $(\delta x, \delta\lambda) := (\tilde{x}-x, \tilde{\lambda}-\lambda)$ and $\delta u := Q^{-1}B^*\delta\lambda$. Then there is a radius $r_E > 0$ independent of T such that, if*

$$\|\varepsilon_1\|_{L_2(0,T;\mathbb{R}^n)} + \|\varepsilon_T\| + \|\varepsilon_2\|_{L_2(0,T;\mathbb{R}^n)} + \|\varepsilon_0\| \le r_E,$$

then for any $0 < \mu < 1$ satisfying

$$\mu < \frac{1}{\|L_r''(x,\lambda)\|_{L((L_2(0,T;\mathbb{R}^n)\times\mathbb{R}^n)^2, H^1(0,T;\mathbb{R}^n)^2)}}$$

there is a constant $c \ge 0$ independent of T such that, setting

$$\rho := \left\|e^{-\mu t}\varepsilon_1\right\|_{L_2(0,T;\mathbb{R}^n)} + \left\|e^{-\mu T}\varepsilon_T\right\| + \left\|e^{-\mu t}\varepsilon_2\right\|_{L_2(0,T;\mathbb{R}^n)} + \|\varepsilon_0\|,$$

it holds that

$$\left\|e^{-\mu t}\delta x(t)\right\|_{H^1(0,T;\mathbb{R}^n)} + \left\|e^{-\mu t}\delta u(t)\right\|_{L_\infty(0,T;\mathbb{R}^m)} + \left\|e^{-\mu t}\delta\lambda(t)\right\|_{H^1(0,T;\mathbb{R}^n)} \le c\rho.$$

4.4 Semilinear heat equations

In this part we will verify the assumptions of the abstract implicit function theorem, i.e., Theorem 4.3, for a class of semilinear heat equations. To this end, we assume that $\Omega \subset \mathbb{R}^n$, $n \ge 2$ is a bounded domain with smooth boundary. The analysis in this part is heavily motivated by the approach taken in [119], where the authors derive a Maximum Principle for optimal control problems governed by semilinear parabolic PDEs. In that work it is shown that for sufficiently smooth data, the solution x of a semilinear parabolic PDE with monotone nonlinearity indeed satisfies $x \in L_\infty(0,T;L_\infty(\Omega))$. This allows for existence results globally in time without global Lipschitz conditions on the nonlinearity. For convenience of the reader, we briefly introduce the setting considered in [119]. To this end, we assume that the PDE of interest is semilinear parabolic, i.e., the nonlinear operator of the state equation of (4.1) is given by

$$\bar{A}(x) = \mathcal{A}x - \varphi(x),$$

where $\varphi : \mathbb{R} \to \mathbb{R}$ is a smooth nonlinearity satisfying $\varphi'(x) \ge c_0$ for $c_0 \in \mathbb{R}$. The operator $-\mathcal{A}$ is considered to be an elliptic differential operator of second order, i.e.,

$$\mathcal{A}x := \sum_{i,j=0}^{n} D_i(a_{ij}D_jx), \tag{4.20}$$

where $a_{ij} = a_{ji} \in C(\bar{\Omega}, \mathbb{R})$ and $a_{i,j}(\omega)v \cdot v > 0$ for all $\omega \in \Omega$ and $v \in \mathbb{R}^n$. By $\frac{\partial x}{\partial \nu_\mathcal{A}}(t,s) = \sum_{i,j=0}^n a_{ij}(s)\partial_j x(t,s)\nu_i(s)$ we denote the conormal derivative of x, where $\nu = (\nu_1, \dots, \nu_n)$ is the outward unit normal to $\partial\Omega$. We consider the domain

$$D(\mathcal{A}) = \{v \in C^2(\Omega) \mid v = 0 \text{ on } \partial\Omega\} \qquad \text{or} \qquad D(\mathcal{A}) = \{v \in C^2(\Omega) \mid \frac{\partial v}{\partial \nu_\mathcal{A}} = 0 \text{ on } \partial\Omega\} \tag{4.21}$$

for either homogeneous Dirichlet or homogeneous Neumann boundary conditions. We assume w.l.o.g. that there is $\alpha > 0$ such that

$$-\int_\Omega \mathcal{A}vv\, d\omega \geq \frac{\alpha}{2}\|v\|^2_{H^1(\Omega)} \tag{4.22}$$

for $v \in D(\mathcal{A})$. In case of Dirichlet boundary conditions this immediately follows with integration by parts and the Poincaré inequality. For Neumann boundary conditions, we can replace $\mathcal{A}x$ by $(\mathcal{A} - kI)$ for any $k > 0$ by $\bar{A}(x) = \mathcal{A}x - \varphi(x) = (\mathcal{A} - kI)x + kx - \varphi(x)$ and redefine $\varphi(x) := \varphi(x) - kx$ accordingly.

It can be shown that for all $1 \leq l < \infty$ the closure $\mathcal{A}_l$ of $\mathcal{A}$ in $L_l(\Omega)$ generates an analytic semigroup $(\mathcal{T}_l(t))_{t\geq 0}$ in $L_l(\Omega)$. For $1 < l < \infty$, the domain is given by $D(\mathcal{A}_l) = \{v \in W^{2,l}(\Omega) \,|\, v = 0 \text{ on } \partial\Omega\}$ or $D(A_l) = \{v \in W^{2,l}(\Omega) \,|\, \frac{\partial v}{\partial \nu_{\mathcal{A}}} = 0 \text{ on } \partial\Omega\}$, depending on the choice in (4.21). Additionally, the spectrum of $\mathcal{A}_l$ does not depend on $1 \leq l < \infty$. For details we refer to [122] and [119, Section 3].

For the semigroup $(\mathcal{T}_l(t))_{t\geq 0}$, we have the following stability estimate, which is the main tool of this part.

Proposition 4.20. *For any $\delta > 0$ there is $\mu_0 > 0$ and a constant $c > 0$ independent of t, such that*

$$\|\mathcal{T}_l(t)\psi_0\|_{L_q(\Omega)} \leq c\frac{e^{-\mu_0 t}}{t^{\frac{n}{2}(\frac{1}{l}-\frac{1}{q}+\delta)}}\|\psi_0\|_{L_l(\Omega)} \qquad \forall t > 0 \tag{4.23}$$

for all $\psi_0 \in L_l(\Omega)$ and $1 \leq l \leq q \leq \infty$ with $l < \infty$. In the case of homogeneous Dirichlet boundary conditions $\delta = 0$ can be chosen.

Proof. See [122, Lemma 1] or [4, Proposition 12.5]. □

This stability result for analytic semigroups turns out to be crucial to derive estimates in L_p-spaces for large p for, e.g., right-hand sides in $L_2(0,T; L_2(\Omega))$ as performed in the following. As a consequence of those L_p-estimates, we can allow for a wide range of different functional analytic settings, i.e., different choices of integrability parameters. This flexibility can then be leveraged when verifying T-uniformity in the context of the superposition operator, i.e., rendering the constants in Proposition 4.7 independent of T, cf. Example 4.13. We will again come back to this issue in Remark 4.34.

Depending on the choice of boundary conditions above, we will set $V = H^1(\Omega)$ in the case of homogeneous Neumann boundary conditions or $V = H^1_0(\Omega)$ in the case of homogeneous Dirichlet boundary conditions. Further suppose that the control is distributed, i.e., $B \in L(L_2(\Omega_c)), L_2(\Omega))$ for a control domain $\Omega_c \subset \Omega$..

An important assumption in the remainder of this part is boundedness on the linearization point. This allows us to analyze the linearized PDE system by classical methods as the involved coefficients are bounded. Another crucial point is positivity of the second derivative of the Lagrangian with respect to the state. We will make the following assumptions.

Assumption 4.21. *Let the following hold.*

i) $(x, \lambda) \in L_\infty(0, T; L_\infty(\Omega))$ *for any solution* (x, u, λ) *of* (4.3) *and* $(\bar{x}, \bar{\lambda}) \in L_\infty(\Omega)$ *for any solution* $(\bar{x}, \bar{u}, \bar{\lambda})$ *of* (4.7).

ii) Set $(x^0, \lambda^0) = (x, \lambda)$ *solving* (4.3) *or* $(x^0, \lambda^0) = (\bar{x}, \bar{\lambda})$ *solving* (4.7). *We assume that* $(L_r)_{xx}(x, \lambda) = J_{xx}(x^0) + \varphi''(x^0)\lambda^0 \in L_\infty(0, T; L_\infty(\Omega))$ *induces a nonnegative multiplication operator, i.e., for* $v\colon [0, T] \times \Omega \to \mathbb{R}$

$$((L_r)_{xx}(x^0, \lambda^0)v)(t, \omega) := (L_r)_{xx}(x^0, \lambda^0)(t, \omega) \cdot v(t, \omega)$$

and $(L_r)_{xx}(x^0, \lambda^0)(t, \omega) \geq 0$ *for a.e.* $t \in [0, T]$ *and* $\omega \in \Omega$.

We briefly remark on these assumptions

Remark 4.22. *In order to render Assumption 4.21 i) satisfied, one usually assumes that the data of* (4.1) *and* (4.6) *is sufficiently smooth. Boundedness of solutions in time and space for parabolic problems was proven in [119]. Similarly, for semilinear elliptic equations, a proof for continuity of solutions can be found in [30]. The interested reader is also referred to the respective parts in the monograph [138].*

Regarding Assumption 4.21 ii), $(L_r)_{xx}(x^0, \lambda^0)$ *induces a multiplication operator if, e.g., the cost functional is of the form*

$$J(x, u) = \frac{1}{2} \int_0^T \|x - x_d\|^2_{L_2(\Omega_o)} + \frac{\alpha}{2} \|u(t)\|^2_U$$

for $\Omega_o \subset \Omega$ *and if the nonlinearity is given by* $\varphi(x) = x^3$. *In that case,*

$$(L_r)_{xx}(x^0, \lambda^0) = \chi_{\Omega_o} + 6x^0\lambda^0,$$

where χ_{Ω_o} *is the characteristic function of the observation region* Ω_o. *The positivity assumption is fulfilled if, e.g.,* $\Omega_o = \Omega$ *and if* λ^0 *and* x^0 *are small in* $L_\infty(0, T; L_\infty(\Omega))$. *For* $(x^0, \lambda^0) = (\bar{x}, \bar{\lambda})$ *or* $(x^0, \lambda^0) = (\tilde{x}, \tilde{\lambda})$, *the latter can be verified by imposing smallness conditions on the data of the underlying steady-state or dynamic OCP, cf. Example 4.33. Again, we note that, as seen in this example and as stated in [136, Remark 6], this assumption is not standard in optimal control. In particular, it is not clear how to verify it by, e.g., second-order sufficient conditions. However, this assumption is crucial to define a square root as we will do in the following, which itself is necessary to obtain stability results for the linearized system, cf. the proof of Theorem 4.27. The assumption that* L_{xx} *is positive semidefinite, was also made in [136, Theorem 1] and [135, Theorem 1].*

We now introduce a square root property for the second derivative of the reduced Lagrange function with respect to the state.

Lemma 4.23. *Let Assumption 4.21 hold and set $(x^0,\lambda^0) = (x,\lambda)$ or $(x^0,\lambda^0) = (\bar{x},\bar{\lambda})$. Then, there is a multiplication operator $C \in L(L_{p_1}(0,T;L_{p_2}(\Omega)))$ for all $1 \le p_1, p_2 \le \infty$ defined by*

$$(Cv)(t,\omega) := \sqrt{(L_r)_{xx}(x^0,\lambda^0)(t,\omega)} \cdot v(t,\omega) \tag{4.24}$$

such that

$$(L_r)_{xx}(x^0,\lambda^0) = C^2.$$

Proof. The claim follows directly from Assumption 4.21 by positivity of $(L_r)_{xx}(x^0,\lambda^0)(t,\omega)$ and regularity of x^0 and λ^0. □

4.4.1 A T-independent bound for the solution operator

In order to apply Theorem 4.3, we will show a bound on the inverse of

$$L_r''(x^0,\lambda^0)\colon (L_{p_1}(0,T;L_{p_2}(\Omega)) \cap W^{1,2}(0,T,D(\mathcal{A}_2),L_2(\Omega))^2 \to (L_2(0,T;L_2(\Omega)) \times V)^2,$$

where $2 \le p_1, p_2 \le \infty$, $\frac{n}{2}(\frac{1}{2} - \frac{1}{p_2}) < \frac{1}{p_1} + \frac{1}{2}$, $p_2 < \frac{2n}{n-2}$ and (x^0,λ^0) either solves the static system (4.8) or the dynamic system (4.4).

To derive an operator norm we consider the linear system

$$\underbrace{\begin{pmatrix} J_{xx}(x^0) + \varphi''(x^0)\lambda^0 & -\frac{d}{dt} - \mathcal{A}_2^* + \varphi'(x^0) \\ 0 & E_T \\ \frac{d}{dt} - \mathcal{A}_2 + \varphi'(x^0) & -BQ^{-1}B^* \\ E_0 & 0 \end{pmatrix}}_{L_r''(x^0,\lambda^0)} \begin{pmatrix} \delta x \\ \delta\lambda \end{pmatrix} = \begin{pmatrix} l_1 \\ \delta\lambda_T \\ l_2 \\ \delta x_0 \end{pmatrix} \tag{4.25}$$

for $(l_1, \delta\lambda_T, l_2, \delta x_0) \in (L_2(0,T;L_2(\Omega) \times V)$. Note that due to $x^0 \in L_\infty(0,T;L_\infty(\Omega))$ and due to the smoothness of φ, the terms $J_x(x^0)$, $\varphi'(x^0)$, and $\varphi''(x^0)$ are in $L_\infty(0,T;L_\infty(\Omega))$ because of Proposition 4.8 and hence can be interpreted as pointwise multiplications. With slight abuse of notation, we denote by the same symbol the corresponding superposition operator.

Moreover, under Assumption 4.21 and with Lemma 4.23, we obtain a self-adjoint operator $C \in L(L_{p_1}(0,T;L_{p_2}(\Omega)))$ as defined in (4.24) for all $1 \le p_1, p_2 \le \infty$ such that we may write

$$J_{xx}(x^0) + \varphi''(x^0)\lambda^0 = C^2.$$

We now aim to estimate $(\delta x, \delta\lambda)$ by means of the right-hand side of (4.25) in appropriate norms. To this end, we make the following stabilizability assumptions.

Assumption 4.24. *Let Assumption 4.21 hold. Further, set $(x^0,\lambda^0) = (x,\lambda)$ or $(x^0,\lambda^0) = (\bar{x},\bar{\lambda})$ and let $c_0 \in \mathbb{R}$ such that $c_0 \le \varphi'(w)$ for all $w \in \mathbb{R}$. Consider $C \in L(L_{p_1}(0,T;L_{p_2}(\Omega)))$ for all $1 \le p_1, p_2 \le \infty$ defined by (4.24). Let $\bar{C} \in L(L_p(\Omega), L_p(\Omega))$ for all $2 \le p \le \infty$ be such that $\|\bar{C}v\|_{L_2(0,T;L_2(\Omega))} \le \|Cv\|_{L_2(0,T;L_2(\Omega))}$ for all $v \in L_2(0,T;L_2(\Omega))$, where C is defined in (4.24).*

Additionally assume:

i) $(\mathcal{A}_l - c_0 I, \bar{B})$ *is exponentially stabilizable in the sense that for all* $1 \le l < \infty$ *there is* $\bar{K}_{\bar{B}} \in L(L_l(\Omega), L_l(\Omega_c))$ *satisfying* $\bar{B}\bar{K}_{\bar{B}} \in L(L_l(\Omega))$ *such that* $\mathcal{A}_l - c_0 I + \bar{B}\bar{K}_{\bar{B}}$ *generates an exponentially stable analytic semigroup in* $L_l(\Omega)$ *satisfying* (4.23).

ii) $(\mathcal{A}_l - c_0 I, \bar{C})$ *is exponentially stabilizable in the sense that for all* $1 \le l < \infty$ *there is* $\bar{K}_{\bar{C}} \in L(L_l(\Omega))$ *such that* $\mathcal{A}_l - c_0 I + \bar{C}^* \bar{K}_{\bar{C}}$ *generates an exponentially stable analytic semigroup in* $L_l(\Omega)$ *satisfying* (4.23).

iii) $(\mathcal{A}_2 - \varphi'(x^0), C)$ *and* $(\mathcal{A}_2 - \varphi'(x^0), B)$ *are exponentially stabilizable in the following sense: There are operators* $K_B \in L(L_2(0,T;L_2(\Omega)), L_2(0,T;L_2(\Omega_c)))$ *and* $K_C \in L(L_2(0,T;L_2(\Omega)))$ *such that*

$$-\int_0^T \int_\Omega \big(\mathcal{A}_2 - \varphi'(x^0) + BK_B\big)\, vv \, d\omega\, dt \ge \alpha \|v\|^2_{L_2(0,T;V)}$$
$$-\int_0^T \int_\Omega \big(\mathcal{A}_2 - \varphi'(x^0) + CK_C\big)\, vv \, d\omega\, dt \ge \alpha \|v\|^2_{L_2(0,T;V)}$$

for all $v \in L_2(0,T;D(\mathcal{A}_2))$.

iv) $\|x^0\|_{L_\infty(0,T;L_\infty(\Omega))}$ *and* $\|C\|_{L(L_p(0,T;L_p(\Omega)),L_p(0,T;L_p(\Omega)))}$ *are bounded independently of* T *for all* $1 \le p \le \infty$.

We briefly comment on these assumptions.

Remark 4.25. *The assertions Assumption 4.24 i) and ii) ensure that the linearized system is stabilizable and detectable and that the closed-loop operators generate a strongly continuous exponentially stable analytic semigroup in* $L_l(\Omega)$ *for all* $1 \le l < \infty$ *satisfying the particular stability estimate* (4.23)*. The third assumption, i.e., iii) allows us to deduce the* $W([0,T])$*-bound analogously to Theorem 3.29 and was introduced for non-autonomous parabolic problems in Definition 3.20. The last assumption ensures that the coefficients in the linearized system are bounded independently of* T*. This is trivially fulfilled for a steady state linearization point* (x^0, λ^0)*. In case that the linearization point is the time-dependent optimal solution, this estimate is satisfied if, e.g., a turnpike property in this uniform norm holds. The latter was proven in cf. [110, Theorem 0.2] under a smallness assumption on the reference state in case of a tracking type cost functional. If* $\varphi'(x) \ge 0$*, i.e., the nonlinearity is monotone, then conditions i)-iii) are trivially satisfied by choosing zero for all feedback operators.*

A central tool in the following will be a convolution estimate, similar to the proof of Lemma 2.36 in the linear case. This proof is further motivated by the approach of [119, Proposition 3.1].

Theorem 4.26. *Let Assumption 4.24 hold and let* $(\delta x, \delta\lambda) \in W([0,T])^2$ *solve* (4.25)*. Then, for all* $p_1, p_2 \ge 2$ *satisfying* $\frac{n}{2}(\frac{1}{2} - \frac{1}{p_2}) < \frac{1}{p_1} + \frac{1}{2}$ *with* $p_1 < \infty$ *and* $p_2 < \frac{2n}{n-2}$*, there is a constant* $c > 0$ *independent of* T*, such that*

$$\|(\delta x, \delta\lambda)\|_{W^{1,2}(0,T,D(\mathcal{A}_2),L_2(\Omega))^2} + \|(\delta x, \delta\lambda)\|_{L_{p_1}(0,T;L_{p_2}(\Omega))^2}$$
$$\le c\big(\|C\delta x\|_{L_2(0,T;L_2(\Omega))} + \|B^*\delta\lambda\|_{L_2(0,T;L_2(\Omega_c))} + \|r\|_{(L_2(0,T;L_2(\Omega))\times V)^2}\big),$$

where $r := (l_1, \delta\lambda_T, l_2, \delta x_0)$*. In the case of homogeneous Dirichlet boundary conditions,* $p_1 = \infty$ *can be chosen.*

Proof. We will first show all estimates for the state. To this end, for the $W^{1,2}(0, T, D(\mathcal{A}_2), L_2(\Omega))$-estimate, we consider the state equation of (4.25), i.e.,

$$\delta x' + (-\mathcal{A}_2 + \varphi'(x^0))\delta x - BQ^{-1}B^*\delta\lambda = l_2$$

with initial condition $\delta x(0) = \delta x_0$. Adding the stabilizing feedback CK_C from Assumption 4.24 iii), we obtain

$$\delta x' + (-\mathcal{A}_2 + \varphi'(x^0) - C^*K_C)\delta x = BQ^{-1}B^*\delta\lambda + l_2 - C^*K_C\delta x.$$

Testing the equation with δx, using the coercivity of Assumption 4.24 iii) we get

$$\|\delta x\|_{L_2(0,T;V)} \leq c\left(\|\delta x_0\|_V + \|B^*\delta\lambda\|_{L_2(0,T;L_2(\Omega))} + \|l_2\|_{L_2(0,T;L_2(\Omega))} + \|C\delta x\|_{L_2(0,T;L_2(\Omega))}\right) \tag{4.26}$$

As $\mathcal{A}_2$ generates an exponentially stable analytic semigroup in $L_2(\Omega)$ by applying the maximal regularity result [19, Part II-1, Theorem 3.1] to

$$\delta x' - \mathcal{A}_2\delta x = -\varphi'(x^0)\delta x + BQ^{-1}B^*\delta\lambda + l_2$$

we obtain (similarly to Lemma 2.64)

$$\begin{aligned}\|\delta x'\|_{L_2(0,T;L_2(\Omega))} &+ \|\mathcal{A}_2\delta x\|_{L_2(0,T;L_2(\Omega))}\\ &\leq c\left(\|\delta x\|_{L_2(0,T;L_2(\Omega))} + \|B^*\delta\lambda\|_{L_2(0,T;L_2(\Omega))} + \|l_2\|_{L_2(0,T;L_2(\Omega))}\right).\end{aligned}$$

Together with (4.26) we conclude

$$\begin{aligned}&||\delta x\|_{W^{1,2}(0,T,D(\mathcal{A}_2),L_2(\Omega))}\\ &\quad\leq c\left(\|l_2\|_{L_2(0,T;L_2(\Omega))} + \|B^*\delta\lambda\|_{L_2(0,T;L_2(\Omega))} + \|C\delta x\|_{L_2(0,T;L_2(\Omega))} + \|\delta x_0\|_V\right).\end{aligned}$$

To obtain the estimate in $L_{p_1}(0, T; L_{p_2}(\Omega))$, we proceed similar to [119, Proof of Proposition 3.1]. Let $\psi_0 \in L_l(\Omega)$, $1 \leq l < \infty$ and $\psi \in C(0, T; L_l(\Omega))$ solve the auxiliary problem

$$\psi' = (\mathcal{A}_l - c_0 I + \bar{C}^*\bar{K}_{\bar{C}})\psi, \qquad \psi(0) = \psi_0,$$

where $\bar{K}_{\bar{C}}$ is a stabilizing feedback for $(\mathcal{A}_l - c_0 I, \bar{C})$ in the sense of Assumption 4.24 ii). Thus, by Proposition 4.20 for all $\delta > 0$ and $1 \leq l \leq q \leq \infty$ with $l < \infty$ and $t > \tau \geq 0$ we have the estimate

$$\|\psi(t-\tau)\|_{L_q(\Omega)} \leq c\frac{e^{-\mu_0(t-\tau)}}{(t-\tau)^{\frac{n}{2}(\frac{1}{l}-\frac{1}{q}+\delta)}}\|\psi_0\|_{L_l(\Omega)}. \tag{4.27}$$

We will now assume that ψ_0 is smooth, and the result for $\psi_0 \in L_l(\Omega)$ can be verified via a density argument. We compute

$$\int_\Omega \psi_0(\omega)\delta x(t,\omega)\,d\omega = \underbrace{\int_0^t \left(\frac{d}{d\tau}\int_\Omega \psi(t-\tau,\omega)\delta x(\tau,\omega)\,d\omega\right)d\tau}_{\textcircled{I}} + \underbrace{\int_\Omega \psi(t,\omega)\delta x_0(\omega)\,d\omega}_{\textcircled{II}}. \tag{4.28}$$

For the first part of (4.28) we obtain with $c_0 - \varphi'(x^0) \le 0$ and by self-adjointness that

$$\begin{aligned}
\textcircled{I} &= \int_0^t \left(\int_\Omega -\psi'(t-\tau,\omega)\delta x(\tau,\omega) + \psi(t-\tau,\omega)\delta x'(\tau,\omega)\,d\omega\right)d\tau\\
&= \int_0^t \Bigg(\int_\Omega -\mathcal{A}_l\psi(t-\tau,\omega)\delta x(\tau,\omega) - \bar{C}^*K\psi(t-\tau,\omega)\delta x(\tau,\omega) + c_0\psi(t-\tau,\omega)\delta x(\tau,\omega)\\
&\qquad + \psi(t-\tau,\omega)\mathcal{A}_2\delta x(\tau,\omega) - \varphi'(x^0)\psi(t-\tau,\omega)\delta x(\tau,\omega) + l_2(t,\omega)\psi(t-\tau,\omega)\\
&\qquad + \bar{B}Q^{-1}\bar{B}^*\delta\lambda(\tau,\omega)\psi(t-\tau,\omega)\,d\omega\Bigg)d\tau\\
&\le \int_0^t \left(\int_\Omega -\bar{C}^*K\psi(t-\tau,\omega)\delta x(\tau,\omega) + \psi(t-\tau,\omega)l_2(t,\omega) + \psi(t-\tau,\omega)\bar{B}Q^{-1}\bar{B}^*\delta\lambda(\tau,\omega)\,d\omega\right)d\tau.
\end{aligned}$$

In the following we denote by p_2' the dual exponent to p_2, i.e., $\frac{1}{p_2} + \frac{1}{p_2'} = 1$. Using the exponential stability estimate of (4.27) and setting $l = p_2'$ and $q = 2$, we obtain for the first summand of (4.28) that

$$\textcircled{I} \le c\|\psi_0\|_{L_{p_2'}(\Omega)} \int_0^t \frac{e^{-\mu_0(t-\tau)}}{(t-\tau)^{\frac{n}{2}(\frac{1}{p_2'}-\frac{1}{2}+\delta)}} \left(\|\bar{C}\delta x(\tau)\|_{L_2(\Omega)} + \|\bar{B}^*\delta\lambda(\tau)\|_{L_2(\Omega_c)} + \|l_2(\tau)\|_{L_2(\Omega)}\right)d\tau.$$

For the second part of (4.28) we use Hölder's inequality and (4.27) with $q = l = p_2'$ and obtain for any $\delta > 0$ that

$$\textcircled{II} \le \|\psi(t)\|_{L_{p_2'}(\Omega)}\|\delta x_0\|_{L_{p_2}(\Omega)} \le c\frac{e^{-\mu_0 t}}{t^\delta}\|\psi_0\|_{L_{p_2'}(\Omega)}\|\delta x_0\|_{L_{p_2}(\Omega)}. \tag{4.29}$$

Taking the supremum over all $\psi_0 \in L_{p_2'}(\Omega)$ yields for any $t \in [0,T]$ that

$$\begin{aligned}
\|\delta x(t)\|_{L_{p_2}(\Omega)} \le c\int_0^t \frac{e^{-\mu_0(t-\tau)}}{(t-\tau)^{\frac{n}{2}(\frac{1}{p_2'}-\frac{1}{2}+\delta)}} \left(\|\bar{C}\delta x(\tau)\|_{L_2(\Omega)} + \|\bar{B}^*\delta\lambda(\tau)\|_{L_2(\Omega_c)} + \|l_2(\tau)\|_{L_2(\Omega)}\right)d\tau\\
+ c\frac{e^{-\mu_0 t}}{t^\delta}\|x_0\|_{L_{p_2}(\Omega)}.
\end{aligned}$$

We now integrate this inequality over time. To this end, we recall Young's convolution inequality, cf. [147, Theorem II.4.4], which states for $\frac{1}{p_1} + 1 = \frac{1}{2} + \frac{1}{h}$ that

$$\|w * g\|_{L_{p_1}(\mathbb{R})} \le \|w\|_{L_h(\mathbb{R})}\|g\|_{L_2(\mathbb{R})}.$$

We apply this convolution inequality to the functions defined by

$$g(\tau) := \|\bar{C}\delta x(\tau)\|_{L_2(\Omega)} + \|\bar{B}^*\delta\lambda(\tau)\|_{L_2(\Omega_c)} + \|l_2(\tau)\|_{L_2(\Omega)},$$
$$w(\tau) := \frac{e^{-\mu_0\tau}}{\tau^{\frac{n}{2}(\frac{1}{p_2'}-\frac{1}{2}+\delta)}}$$

for any $\tau \in [0,T]$ and extended by zero otherwise. Additionally, we require that $\frac{n}{2}(\frac{1}{p_2}-\frac{1}{2}+\delta) < \frac{1}{h}$ and $p_1 < \infty$ to ensure $\|w\|_{L_h(\mathbb{R})} < \infty$. This yields

$$\begin{aligned}
&\|\delta x\|_{L_{p_1}(0,T;L_{p_2}(\Omega))}\\
&\le c\left(\|\bar{C}\delta x\|_{L_2(0,T;L_2(\Omega))} + \|B^*\delta\lambda\|_{L_2(0,T;L_2(\Omega_c))} + \|l_2\|_{L_2(0;T;L_2(\Omega))} + \|\frac{e^{-\mu t}}{t^\delta}\|_{L_{p_1}(\mathbb{R})}\|\delta x_0\|_{L_{p_2}(\Omega)}\right)\\
&\le c\left(\|C\delta x\|_{L_2(0,T;L_2(\Omega))} + \|B^*\delta\lambda\|_{L_2(0,T;L_2(\Omega_c))} + \|l_2\|_{L_2(0;T;L_2(\Omega))} + \|\delta x_0\|_V\right),
\end{aligned}$$

where the last estimate follows from Assumption 4.24, i.e., $\|\bar{C}v\|_{L_2(0,T;L_2(\Omega))} \le \|Cv\|_{L_2(0,T;L_2(\Omega))}$ for all $v \in L_2(0,T;L_2(\Omega))$, by $0 \le \delta < 1$ and by the classical Sobolev embedding theorem $V \hookrightarrow L_{p_2}(\Omega)$ for $p_2 < \frac{2n}{n-2}$, cf. [1, Theorem 5.4]. For Dirichlet boundary conditions $\delta = 0$ can be chosen and thus we can take the supremum over all t in (4.29), i.e., choose $p_1 = \infty$.

For the adjoint state λ, one proceeds analogously: First, adding the stabilizing feedback BK_B from Assumption 4.24 iii) allows to conclude the $W^{1,2}(0,T,D(\mathcal{A}_2),L_2(\Omega))$-estimate. The remaining $L_{p_1}(0,T;L_{p_2}(\Omega))$-estimate follows by the same argumentation by replacing the time argument $t-\tau$ for $0 \le \tau < t$ by $t+\tau$ for $0 \le \tau \le T-t$ in (4.27) and integrating from t to T in (4.28). □

This stability estimate can be used to derive a T-uniform estimate for the solution operators norm. This in turn can then be used to also bound the solution operator in exponentially scaled spaces. Both these bounds play a central role in the assumptions of the implicit function theorem Theorem 4.3. The following theorem states the main result of this section.

Theorem 4.27. *Let Assumption 4.24 hold. Further, set $z^0 = (x,\lambda)$ solving (4.3) or $z^0 = (\bar{x},\bar{\lambda})$ solving (4.7). Then, for all $2 \le p_1, p_2$ satisfying $\frac{n}{2}(\frac{1}{2}-\frac{1}{p_2}) < \frac{1}{p_1}+\frac{1}{2}$, $p_1 < \infty$ and $p_2 < \frac{2n}{n-2}$ there is a constant $c \ge 0$ independent of T such that*

$$\|L_r''(z^0)^{-1}\|_{L((L_2(0,T;L_2(\Omega))\times V)^2,(L_{p_1}(0,T;L_{p_2}(\Omega))\cap W^{1,2}(0,T,D(\mathcal{A}_2),L_2(\Omega))^2)} \le c \tag{4.30}$$

Moreover for all $0 < \mu < 1$ satisfying

$$\mu < \frac{1}{\|L_r''(z^0)^{-1}\|_{L((L_2(0,T;L_2(\Omega))\times V)^2,(L_{p_1}(0,T;L_{p_2}(\Omega))\cap W^{1,2}(0,T,D(\mathcal{A}_2),L_2(\Omega))^2)}} \tag{4.31}$$

there is a constant $c \ge 0$ independent of T such that

$$\|L_r''(z^0)^{-1}\|_{L((L_2^s(0,T;L_2(\Omega))\times V^{s(T)}\times L_2^s(0,T;L_2(\Omega))\times V^{s(0)}),(L_{p_1}^s(0,T;L_{p_2}(\Omega))\cap W_s^{1,2}(0,T,D(\mathcal{A}_2),L_2(\Omega)))^2)} \le c \tag{4.32}$$

for the scaling functions $s(t) = e^{-\mu t}$ or $s(t) = \frac{1}{e^{-\mu t}+e^{-\mu(T-t)}}$. In the case of homogeneous Dirichlet boundary conditions, $p_1 = \infty$ can be chosen.

Proof. Using the bound derived in Theorem 4.26, it only remains to estimate $\|C\delta x\|^2_{L_2(0,T;L_2(\Omega))}+\|B^*\delta\lambda\|^2_{L_2(0,T;L_2(\Omega_c))}$. This follows by testing in (4.25) the adjoint equation with the state, the state equation with the adjoint, integrating by parts and subtracting, which yields

$$\|C\delta x\|^2_{L_2(0,T;L_2(\Omega))}+\|B^*\delta\lambda\|^2_{L_2(0,T;L_2(\Omega_c))}\leq \\ |\langle l_1,\delta x\rangle_{L_2(0,T;L_2(\Omega))}|+|\langle l_2,\delta\lambda\rangle_{L_2(0,T;L_2(\Omega))}|+|\langle \delta x_0,\delta\lambda(0)\rangle_{L_2(\Omega)}|+|\langle \delta\lambda_T,\delta x(T)\rangle_{L_2(\Omega)}|.$$

The bound (4.30) then follows. To prove the bound in the scaled spaces we proceed analogously to the proofs of Theorems 3.14 and 3.16. Hence we define $M := L_r''(z^0)$ and set $Z := (L_{p_1}(0,T;L_{p_2}(\Omega))\cap W^{1,2}(0,T;D(\mathcal{A}_2),L_2(\Omega)))^2$ and $E := (L_2(0,T;L_2(\Omega))\times V)^2$. First, setting $s(t)=\frac{1}{e^{-\mu t}+e^{-\mu(T-t)}}$ a straightforward computation shows that for $\varepsilon\in E$

$$\begin{aligned} M\delta z &= \varepsilon\\ (M-\mu P)(s\delta z) &= s\varepsilon\\ (I-\mu M^{-1}P)(s\delta z) &= M^{-1}s\varepsilon \end{aligned}$$

where $P := \begin{pmatrix} 0 & F\\ 0 & 0\\ -F & 0\\ 0 & 0\end{pmatrix}$ and $F := \frac{(e^{-\mu(T-t)}-e^{-\mu t})}{(e^{-\mu t}+e^{-\mu(T-t)})}<1$. Thus, choosing $\mu<\frac{1}{\|M^{-1}\|_{L(E,Z)}}$ and setting $\beta=\mu\|M^{-1}\|_{L(E,Z)}<1$, a standard Neumann argument, cf. [85, Theorem 2.14] yields,

$$\|s\delta z\|_Z\leq \frac{\|M^{-1}\|_{L(E,Z)}}{1-\beta}\|s\varepsilon\|_E.$$

Thus, by definition of the scaled norms, the bound (4.32) for $s(t)=\frac{1}{e^{-\mu t}+e^{-\mu(T-t)}}$ follows. Completely analogously we conclude (4.32) for $s(t)=e^{-\mu t}$ with the same argumentation and $P := \begin{pmatrix} 0 & -I\\ 0 & 0\\ I & 0\\ 0 & 0\end{pmatrix}$. ☐

We briefly comment on the estimates of Theorem 4.26 and Theorem 4.27 in the case of homogeneous Dirichlet boundary conditions.

Remark 4.28. *In the case of $n=2$, the restrictions for p_1,p_2, i.e., $\frac{n}{2}(\frac{1}{2}-\frac{1}{p_2})<\frac{1}{p_1}+\frac{1}{2}$ with $p_1<\infty$ and $p_2<\frac{2n}{n-2}$, allow for all $2\leq p_1,p_2\leq\infty$ except $p_1=p_2=\infty$. If $n=3$, e.g., the choice $2\leq p_1=p_2<6$ is allowed. The pointwise in time estimates, i.e., choosing $p_1=\infty$ and thus requiring $p_2<\infty$ for $n=2$ and $p_2<6$ for $n=3$ are consistent with maximal parabolic regularity theory. In that case, for initial values in $H_0^1(\Omega)$ and right-hand sides in $L_2(0,T;L_2(\Omega))$ the maximal parabolic regularity theory leads to solutions continuous in time with values in $H_0^1(\Omega)$, even in case of time-dependent generators, cf. [6]. By classical embedding theorems we get $C(0,T;H_0^1(\Omega))\hookrightarrow C(0,T;L_p(\Omega))$ with $1\leq p<\infty$ for $n=2$ and $1\leq p<6$ for $n=3$ which coincides with the choice of p_2 specified above.*

4.4.2 Exponential sensitivity and turnpike results

We can now combine the results of Section 4.2.2 regarding superposition operators and the bound on the solution operator to the linearized problem of Section 4.4.1 to apply the implicit function theorem Theorem 4.3 to semilinear parabolic problems. In that case, we will choose $p_1 = p_2$ as the image space of the superposition operators is $L_2(0,T;L_2(\Omega))$, i.e., spatial and temporal integrability coincide. In that case, the assumptions of Theorem 4.27 on $p_1 = p_2 = p$ simplify to $p < \frac{2n}{n-2}$. This choice of p represents the largest exponent such that $H^1(\Omega) \hookrightarrow L_p(\Omega)$.

Theorem 4.29. *Let Assumption 4.24 hold and consider $2 \le p < \frac{2n}{n-2}$ arbitrary. Further, let $\varphi(x)$ and $J_x(x)$ induce twice continuously Fréchet differentiable superposition operators from $L_p(0,T;L_p(\Omega))$ to $L_2(0;T;L_2(\Omega))$ with T-uniformly continuous derivatives. Then, the nonlinear operator $G(z,\varepsilon) := L_r'(z) - \varepsilon$ with L_r' given in (4.4) satisfies the assumptions of Theorem 4.3 uniformly in T for any $0 < \mu < 1$ satisfying (4.31) with*

- *$z^0 = (\bar{x}, \bar{\lambda})$ solving the steady-state problem (4.8) and $\varepsilon^0 = (0, \bar{\lambda}, 0, \bar{x} - x_0)$ and the scaling $s(t) = \frac{1}{e^{-\mu t} + e^{-\mu(T-t)}}$, or*
- *$z^0 = (x, \lambda)$ solving the dynamic problem (4.4) and $\varepsilon^0 = 0$ and the scaling $s(t) = e^{-\mu t}$*

with the spaces $Z = \left(L_p(0,T;L_p(\Omega)) \cap W^{1,2}(0,T,D(\mathcal{A}_2),L_2(\Omega))\right)^2$ and $E = (L_2(0,T;L_2(\Omega)) \times V)^2$ and the scaled norms $\|\cdot\|_{Z_s} = \|\cdot\|_{\left(L_p^s(0,T;L_p(\Omega)) \cap W_s^{1,2}(0,T,D(\mathcal{A}_2),L_2(\Omega))\right)^2}$ and $\|\cdot\|_{E_s} = \|\cdot\|_{L_2^s(0,T;L_2(\Omega)) \times V^{s(T)} \times L_2^s(0,T;L_2(\Omega)) \times V^{s(0)}}$.

Proof. Assumption ii) and iii) of Theorem 4.3 follow by T-uniform continuity of the derivative of the superposition operators via Lemma 4.14. T-uniform continuous invertibility of $L_r''(\bar{x}, \bar{\lambda})^{-1}$ resp. $L_r''(\bar{x}, \bar{\lambda})^{-1}$ in scaled and unscaled spaces follows from Theorem 4.27 setting $p_1 = p_2 = p$. □

This result can now be used to deduce a local turnpike result, stating that solutions of the dynamic problem (4.3) are close to solutions of the static problem (4.8) for the majority of the time under the assumption that initial resp. terminal values are close enough to the turnpike.

Again note that the implicit function theorem Theorem 4.3 provides the estimates for state and adjoint. The estimates for the control can be concluded via $\delta u(t) = Q^{-1} B^* \delta\lambda(t)$, i.e., inserting the eliminating relation for the control, cf. (4.5).

Corollary 4.30. *Let the assumptions of Theorem 4.29 hold with $2 \le p < \frac{2n}{n-2}$. Consider (x, u, λ) solving the nonlinear dynamic problem (4.1) and $(\bar{x}, \bar{u}, \bar{\lambda})$ solving the nonlinear static problem (4.6). Define $(\delta x, \delta u, \delta\lambda) := (x - \bar{x}, u - \bar{u}, \lambda - \bar{\lambda})$. Then there are $r_E > 0$, $0 < \mu < 1$ (satisfying (4.31) with $z^0 = (\bar{x}, \bar{\lambda})$) and $c \ge 0$, all independent of T, such that if*

$$\|x_0 - \bar{x}\|_V + \left\|\bar{\lambda}\right\|_V \le r_E$$

it holds that

$$\left\|\frac{1}{e^{-\mu t}+e^{-\mu(T-t)}}\delta x(t)\right\|_{L_p(0,T;L_p(\Omega))\cap W^{1,2}(0,T,D(\mathcal{A}_2),L_2(\Omega))}+\left\|\frac{1}{e^{-\mu t}+e^{-\mu(T-t)}}\delta u(t)\right\|_{L_\infty(0,T;L_2(\Omega_c))}$$
$$+\left\|\frac{1}{e^{-\mu t}+e^{-\mu(T-t)}}\delta\lambda(t)\right\|_{L_p(0,T;L_p(\Omega))\cap W^{1,2}(0,T,D(\mathcal{A}_2),L_2(\Omega))}\leq cr_E.$$

Further, we can conclude a sensitivity result, which states that perturbations of the extremal equations' dynamics that are small on an initial part lead to disturbances in the variables that are small at an initial part. More specifically we obtain that solutions to the perturbed dynamic problem (4.5) are close to the solutions of the unperturbed dynamic problem (4.3) on an initial part, even if the perturbations increase exponentially. In that context, we have to assume that the perturbations in unscaled norms are sufficiently small.

Corollary 4.31. *Let the assumptions of Theorem 4.29 hold with $2\leq p<\frac{2n}{n-2}$. Let (x,λ) solve the nonlinear extremal equations (4.4) and $(\tilde{x},\tilde{\lambda})$ solve the perturbed extremal equations (4.5). Define $(\delta x,\delta\lambda):=(\tilde{x}-x,\tilde{\lambda}-\lambda)$ and $\delta u=Q^{-1}B^*\delta\lambda$. Then there are $r_E>0$, $0<\mu<1$ (satisfying (4.31) with $z^0=(x,\lambda)$) and $c\geq 0$, all independent of T, such that if*

$$\|\varepsilon_1\|_{L_p(0,T;L_p(\Omega))}+\|\varepsilon_T\|_V+\|\varepsilon_2\|_{L_2(0,T;L_2(\Omega))}+\|\varepsilon_0\|_V\leq r_E$$

and setting

$$\rho:=\left\|e^{-\mu t}\varepsilon_1\right\|_{L_2(0,T;L_2(\Omega))}+\left\|e^{-\mu T}\varepsilon_T\right\|_V+\left\|e^{-\mu t}\varepsilon_2\right\|_{L_2(0,T;L_2(\Omega))}+\|\varepsilon_0\|_V,$$

it holds that

$$\left\|e^{-\mu t}\delta x(t)\right\|_{L_p(0,T;L_p(\Omega))\cap W^{1,2}(0,T,D(\mathcal{A}_2),L_2(\Omega))}+\left\|e^{-\mu t}\delta u(t)\right\|_{L_\infty(0,T;L_2(\Omega_c))}$$
$$+\left\|e^{-\mu t}\delta\lambda(t)\right\|_{L_p(0,T;L_p(\Omega))\cap W^{1,2}(0,T,D(\mathcal{A}_2),L_2(\Omega))}\leq c\rho.$$

We conclude this section by some remarks for possible extensions and an example.

Remark 4.32. *We assumed in this part that the control operator is bounded as linear operator to $L_2(\Omega)$, ruling out the case of boundary control. The case of boundary control could be included if one can ensure that the closed-loop semigroup is analytic and satisfies the stability estimate Proposition 4.20. Perturbations of analytic semigroups can be analyzed with the notion of A-boundedness or A-compactness, cf. [44, Chapter III].*

Example 4.33. *We present an example with distributed control of a heat equation with Dirichlet boundary conditions. To this end we set $V=H_0^1$, $\bar{B}=\chi_{\Omega_c}$, where $\Omega_c\subset\Omega$ non-empty, $\mathcal{A}=\Delta$ and the nonlinearity $\varphi(x)=x^3-c_0x$ for $c_0\in\mathbb{R}$. If c_0 is larger than the smallest eigenvalue of $-\Delta$ in $H_0^1(\Omega)$, the uncontrolled PDE is unstable. Let the cost functional be given by $J(x,u)=\frac{1}{2}\int_0^T\|x-x_d\|^2_{L_2(\Omega)}+\|u\|^2_{L_2(\Omega_c)}$, where $x_d=\bar{x}_d\in L_2(\Omega)$ is a static reference. Using maximal elliptic regularity, cf. [30], we get for the solution of the static system*

$$\|\bar{x}\|_{L_\infty(\Omega)}+\|\bar{\lambda}\|_{L_\infty(\Omega)}\leq c(\Omega)\|x_d\|_{L_2(\Omega)},$$

i.e., $(\bar{x}, \bar{\lambda})$ satisfy Assumption 4.21. Thus, choosing x_d sufficiently small such that

$$\|\bar{\lambda}\bar{x}\|_{L_\infty(0,T;L_\infty(\Omega))} = \|\bar{\lambda}\bar{x}\|_{L_\infty(\Omega)} = \underline{m} < 1,$$

the operator

$$(L_r)_{xx}(\bar{x}, \bar{\lambda}) = I - \bar{\lambda}\bar{x}.$$

is nonnegative and satisfies the assumptions of Lemma 4.23. The square root $\bar{C}$ of this operator in the sense of (4.24) can thus be defined pointwise for $v\colon \Omega \to \mathbb{R}$ and a.e. $\omega \in \Omega$ by

$$(Cv)(\omega) = \sqrt{(1 - \bar{\lambda}(\omega)\bar{x}(\omega))}v(\omega).$$

As the linearization point is the turnpike, i.e., a steady-state, C as defined in (4.24) is time-independent and we can set $\bar{C} = C$ in Assumption 4.24. We will now verify the stabilizability assumptions, i.e., Assumption 4.24 i)-iii) for two particular cases. First, assume $c_0 \geq 0$. Then we can choose the feedback operators $\bar{K}_{\bar{B}} = 0$ and $\bar{K}_{\bar{C}} = 0$ in Assumption 4.24 i) and ii) and $K_B = 0$ and $K_C = 0$ in Assumption 4.24 iii). If $c_0 < 0$, the uncontrolled system can be unstable. If $\Omega_c = \Omega$ and $\bar{B} = I$, i.e., the control is active on the whole domain, one can choose, e.g., the feedback $\bar{K}_{\bar{B}}$ in Assumption 4.24 i) such that $\bar{B}\bar{K}_{\bar{B}} = c_0 I$ and hence $\mathcal{A}_l - c_0 I + \bar{B}\bar{K}_{\bar{B}} = \mathcal{A}_l$. One can proceed analogously in Assumption 4.24 ii) and iii). Alternatively, if neither $c_0 \geq 0$ nor $\Omega_c = \Omega$, Assumption 4.24 i) and ii) can be verified by null controllability results for heat equations in Banach spaces, which imply stabilizability [154, Theorem 3.3]. Further, Assumption 4.24 iii) can be shown analogously as in Example 3.21 with the generalized Poincaré inequality, cf. [138, Lemma 2.5]. Hence, if Assumption 4.24 is satisfied and we can apply the turnpike result of Corollary 4.30.

In order to apply the sensitivity result of Corollary 4.31 we need to analyze

$$(L_r)_{xx}(x, \lambda) = I - x\lambda$$

where (x, λ) solves (4.4). In [119], the authors deduce a T-dependent bound

$$\begin{aligned}\|x\|_{L_\infty(0,T;L_\infty(\Omega))} + &\|\lambda\|_{L_\infty(0,T;L_\infty(\Omega))} \\ &\leq c(T)\left(\|x_d\|_{L_{2+\delta}(0,T;L_{2+\delta}(\Omega))} + \|x_0\|_{L_\infty(\Omega)}\right).\end{aligned}$$

for any $\delta > 0$. If the nonlinear and linearized uncontrolled equation is stable, e.g., if $c_0 \geq 0$, it is possible to show the above bound for the state independently of T, cf. [110, Lemma 1.1] where such an estimate was shown under the assumption that $x_d \in L_\infty(0, T; L_\infty(\Omega))$. Having bounded the state the corresponding bound on the adjoint can be obtained, cf. [110, Lemma A.1] by parabolic regularity. Hence, choosing the data x_d, u_d and x_0 small enough, similar to the elliptic case, we have $\|\lambda x\|_{L_\infty(0,T;L_\infty(\Omega))} = \underline{m} < 1$ and thus the operator $(L_r)_{xx}(x, \lambda)$ satisfies the assumptions of Lemma 4.23 and and we can define the square root C via (4.24) with $\|C\|_{L(L_p(0,T;L_q(\Omega))} = \|\sqrt{(1 - \lambda x)}\|_{L_\infty(0,T;L_\infty(\Omega))}$ for all $1 \leq p, q \leq \infty$. Choosing $\bar{C} = \underline{c}I$ in

Assumption 4.24, where $\underline{c} \le \sqrt{1-\underline{m}^2}$ yields for $v\colon (0,T)\times\Omega \to \mathbb{R}$ and a.e. $(t,\omega)\in(0,T)\times\Omega$ the estimate

$$\underline{c}^2|v(t,\omega)| \le |1-\underline{m}||v(t,\omega)| = \left(1-\|\lambda x\|_{L_\infty(0,T;L_\infty(\Omega))}\right)|v(t,\omega)| \le \|1-\lambda x\|_{L_\infty(0,T;L_\infty(\Omega))}|v(t,\omega)|.$$

Taking the square root yields $\|\bar{C}v\|_{L_2(0,T;L_2(\Omega))} < \|Cv\|_{L_2(0,T;L_2(\Omega))}$ for all $v \in L_2(\Omega)$. Thus, Assumption 4.24 is satisfied and we can apply the sensitivity result of Corollary 4.31.

Remark 4.34. *We briefly discuss the case of nonlinearities that are sums of monotone polynomials, e.g., $\varphi(x) = x^3 + x^5$. In standard applications of superposition operators where the estimates do not need to be uniform in the size of the domain, only the behavior of the nonlinearity towards infinity is important. Thus, in case of $\varphi(x) = x^3 + x^5$ one would estimate the cubic term on the set where $x > 1$ by the higher order term x^5 and bound the remainder by the measure of the domain, as there $x \le 1$. This is not possible if one is particularly interested in estimates independent of the size of the domain, i.e., in our case, independent of T, cf. also Example 4.13. As a remedy, one has to invoke Theorem 4.3 with the space $Z = \left(L_{10}(0,T;L_{10}(\Omega)) \cap L_6(0,T;L_6(\Omega)) \cap W^{1,2}(0,T;D(\mathcal{A}_2),L_2(\Omega))\right)^2$, where the bound on the solution operator follows by Theorem 4.27 if $n=2$.*

4.5 Numerical results

In this part, we will showcase the theoretical results of this chapter by means of numerical examples of nonlinear parabolic equations. First, we will illustrate the turnpike property and second we evaluate the a priori time refinement strategies presented in Section 3.3. Even though the theoretical results in this chapter only consider distributed control of semilinear equations, we will also investigate the boundary control of a quasilinear equation and find that, at least numerically, the same results hold.

4.5.1 Distributed control of a semilinear equation

We briefly recall the numerical example of Section 3.3.1 and add a semilinearity to the state equation. Again, we consider $T = 10$, $\Omega = [0,3]\times[0,1]$, and the cost functional

$$J(x,u) := \frac{1}{2}\|(x-x_{\mathrm{d}})\|^2_{L_2(0,T;L_2(\Omega))} + \frac{\alpha}{2}\|u\|^2_{L_2(0,T;L_2(\Omega))},$$

where $\alpha > 0$ and x_{d} is either the static reference defined in (3.17) or the dynamic reference defined in (3.18). We consider dynamics governed by the semilinear heat equation

$$\begin{aligned} x' - 0.1\Delta x + ex^3 &= u && \text{in } \Omega\times(0,T),\\ x &= 0 && \text{in } \partial\Omega\times(0,T),\\ x(0) &= 0 && \text{in } \Omega, \end{aligned}$$

where $e \ge 0$ is a nonlinearity parameter. In Figure 4.1, the norm of the optimal state and control for different nonlinearity parameters e are depicted. The turnpike property emerges in all four cases, even for very large choices of the nonlinearity parameter. Additionally, we observe that the norm of the turnpike decreases for increasing nonlinearity. This is due to the fact that the nonlinearity increases the stability of the state equation (towards zero), which can be seen by testing the state equation with the state, integrating by parts in time and space and using the Poincaré inequality, which leads to

$$\|x(t)\|^2 \leq -c(\Omega) \int_0^t \|x(s)\|^2 + e\|x(s)\|^4 \, ds + \int_0^t u(s)x(s) \, ds.$$

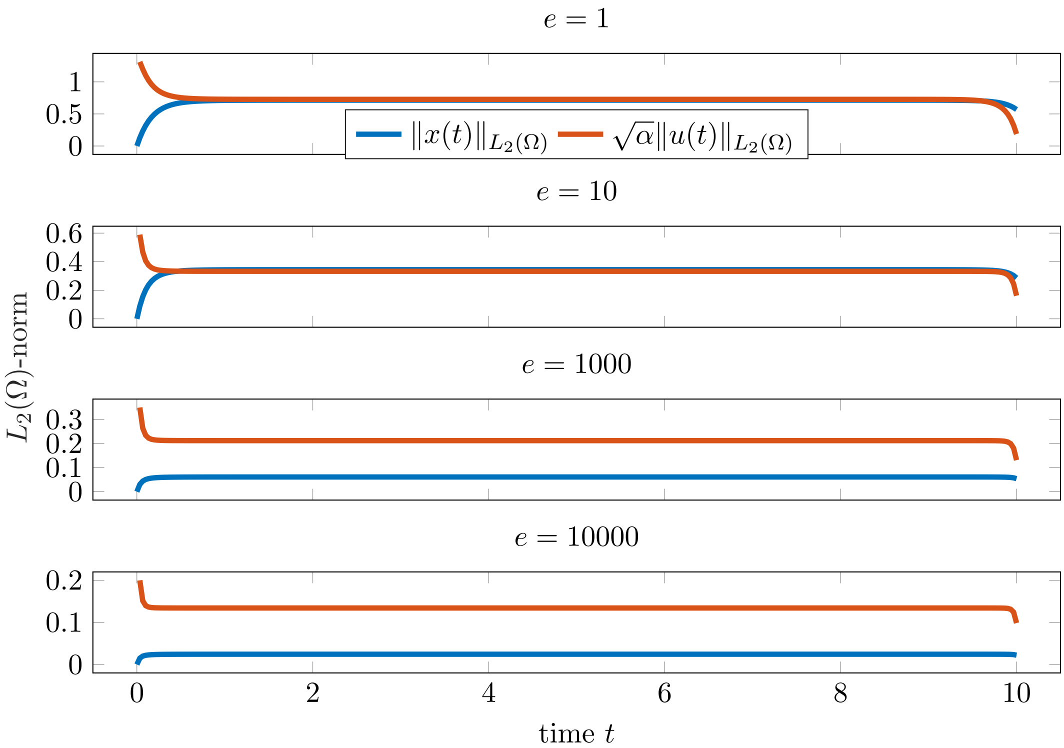


Figure 4.1: Spatial norm of open-loop state and control over time with the static reference $x_{\mathrm{d}}^{\mathrm{stat}}$ and $\alpha = 10^{-1}$ for different nonlinearity parameters.

Second, we apply four steps of the MPC Algorithm 1 to the optimal control problem above. We set the implementation horizon $\tau = 1$ and choose the dynamic reference $x_{\mathrm{d}}^{\mathrm{dyn}}$ defined in (3.18). The simulation of the closed-loop trajectory emerging from the MPC feedback is again computed on three uniform refinements of the initial grid. In Figure 4.2, the closed-loop cost for different a priori time discretization schemes as introduced in Section 3.3 is depicted. It can be seen that the exponential and piecewise uniform time grids achieve lower closed-loop cost than a conventional uniform grid. As all grids are constructed a priori, we note again that the numerical effort is the same for all three techniques.

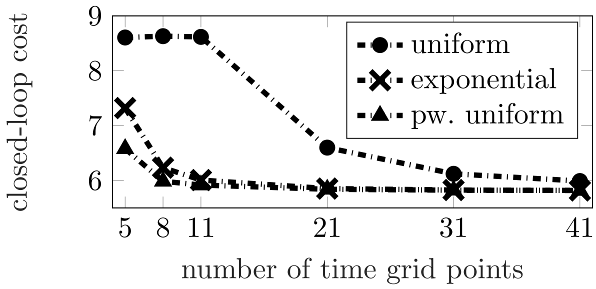


Figure 4.2: Comparison of MPC closed-loop cost for different a priori time discretization schemes with dynamic reference $x_{\mathrm{d}}^{\mathrm{dyn}}$ and parameters $e = 1$ and $\alpha = 10^{-2}$.

4.5.2 Boundary control of a quasilinear equation

As a second numerical example, we consider a heat equation with heat conductivity depending on the temperature. To this end, we introduce the heat conduction tensor

$$\kappa(x)(t,\omega) := \left(c|x(t,\omega)|^2 + 0.1\right),$$

where $c \geq 0$ is a nonlinearity parameter and consider the quasilinear dynamics

$$\begin{aligned} x' - \nabla \cdot (\kappa(x)\nabla x) &= 0 && \text{in } \Omega \times (0,T), \\ \kappa(x)\frac{\partial x}{\partial \nu} &= u && \text{in } \partial\Omega \times (0,T), \\ x(0) &= 0 && \text{in } \Omega. \end{aligned}$$

We use the same tracking-type cost functional as in Section 4.5.1. For an in-depth analysis of optimal control problems governed by quasilinear parabolic equations, the interested reader is referred to [21, 31, 87, 104, 108]. Our theoretical results in Section 4.4 do not cover the case of a quasilinear equation. However, the turnpike property can be observed in Figure 4.3 even for very large choices of the nonlinearity parameter c. Moreover, we observe the same behavior of the norm of the turnpike as in the semilinear example: for increased nonlinearity, the norm of the turnpike decreases. This again reflects the stabilizing effect of the nonlinearity (towards zero). We depict the turnpike property in a norm that is motivated by the second derivative of the Lagrange function, i.e., the scaled $H^1(\Omega)$-norm $\|v\|_{\alpha d,H^1(\Omega)} := \|v\|_{L_2(\Omega)} + \sqrt{d\alpha}\|\nabla v\|_{L_2(\Omega)}$ with $d = 0.1$.

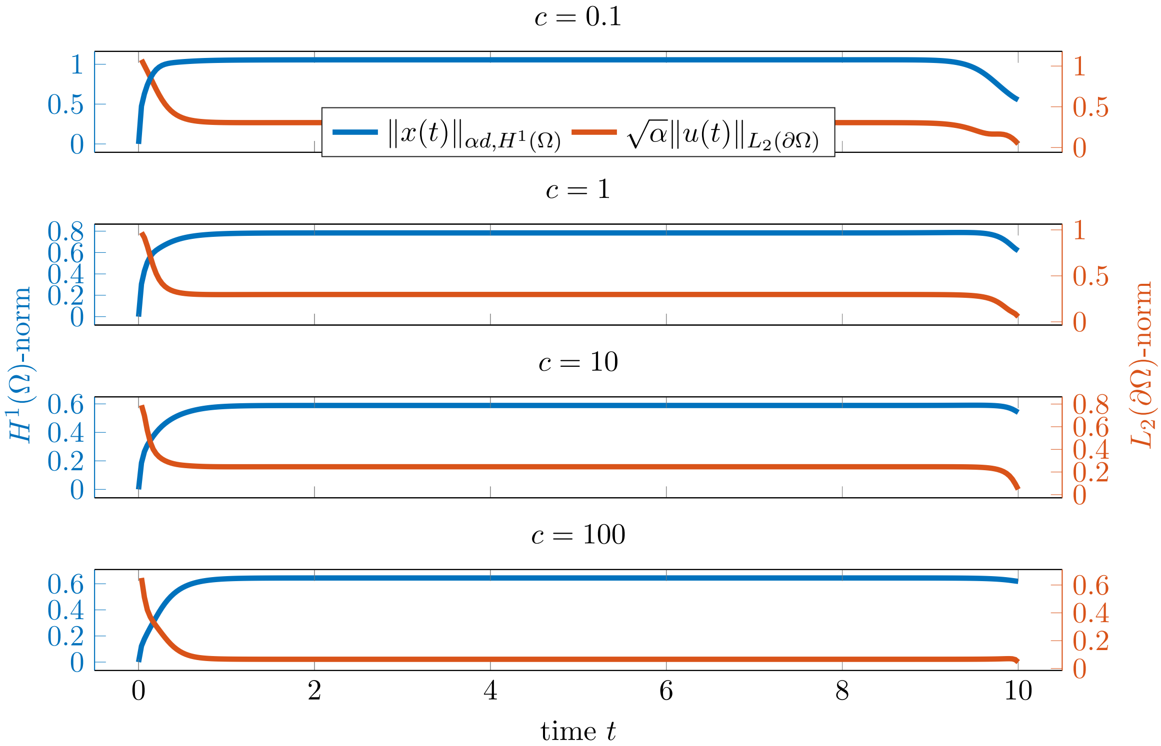


Figure 4.3: Spatial norm of open-loop state and control over time with the static reference $x_{\mathrm{d}}^{\mathrm{stat}}$ for different nonlinearity parameters and $\alpha = 10^{-1}$.

In Figure 4.4 we compare the closed-loop cost of different a priori time discretization schemes. Similar to the semilinear example investigated before, we observe that exponential and piecewise uniform a priori time grids outperform the conventional uniform grid.

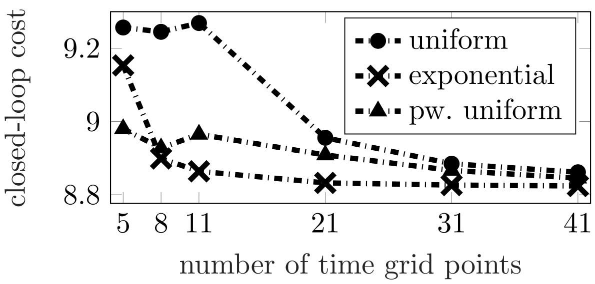


Figure 4.4: Comparison of MPC closed-loop cost for different a priori time discretization schemes with dynamic reference $x_{\mathrm{d}}^{\mathrm{dyn}}$ for different priori time discretization schemes with parameters $c = 0.1$ and $\alpha = 10^{-2}$.

4.6 Outlook

We will briefly discuss several extensions of the analysis performed in this chapter.

- A first extension would be to replace the parabolic equation by a general Cauchy problem. A result on T-independent invertibility was established for the autonomous linear case in Chapter 2. This can be applied to the linearization of the nonlinear first-order necessary optimality conditions, if the linearization point is a steady state and the linearized generator gives rise to a strongly continuous semigroup. Thus, turnpike results can be established in that case when choosing a functional analytic setting where the superposition operators are continuous and differentiable. For a time-dependent linearization point however, the non-autonomous case has to be considered. Under appropriate assumptions, it should be possible to derive estimates in the same spaces as for the autonomous case. The main difficulty, however, is that there is no smoothing effect in general evolution equations. Thus, we have to establish continuity and differentiability of the superposition operators mapping from X to X, which is only possible for $X = L_p(\Omega)$ if the nonlinearity is indeed affine linear, $X = \mathbb{R}^n$ or $p = \infty$, cf. [56, Section 3.1] and [8, Theorem 3.12]. For the latter however, one has to ensure that the underlying dynamics give rise to a continuous semigroup on $L_\infty(\Omega)$, which is, e.g., not the case for the Laplacian, cf. [119, Section 3.1]. A remedy is to not allow for general nonlinearities in the Cauchy problems but rather particular cases, e.g., a semilinear wave equation of the form
$$x'' - \Delta x + \varphi(x) = u,$$
where the nonlinearity only depends on x. In that case, utilizing the solution theory for wave equations, there is indeed a smoothing effect of the solution operator in the following sense. Let $V \hookrightarrow H \hookrightarrow V^*$ form a Gelfand triple. Then one obtains solutions in $x \in L_2(0,T;V)$, $x' \in L_2(0,T;H)$, $x'' \in L_2(0,T;V^*)$ for right hand sides in $L_2(0,T;H)$ and further $x \in C(0,T;V)$, $x' \in C(0,T;H)$, cf. e.g., [96]. After writing the equation as a first-order system and deriving the estimate on the linearized equations' solution operator with range $C(0,T;X)^2$ and $L_2(0,T;X)^2$ for $X = V \times H$ with the results of Chapter 2, a T-independent bound in $L_p(0,T;V)$ for all $p \in [2,\infty]$ follows by the generalized Hölder inequality. Moreover, if, e.g., $V = H^1(\Omega)$, then $V \hookrightarrow L_p(\Omega)$ for all $p < \infty$ in space dimension two and differentiability of the superposition operator corresponding to a polynomial nonlinearity $\varphi(x)$ in these spaces can be deduced straightforwardly.

- The local analysis presented in this chapter fails for equations where the nonlinearity is not continuously differentiable. This is, e.g., the case for problems with control constraints, where the constrains can be eliminated via a max-operator in the optimality conditions. A second example are parabolic non-smooth dynamics of the form
$$x' - \Delta x + \max(x, 0) = u.$$
with, e.g., homogeneous Dirichlet boundary conditions. In these cases, the implicit function theorem fails due to non-smoothness.

- Finally, we discuss an extension to quasilinear equations. In Section 4.5.2, we observed that the solutions to quasilinear problem indeed enjoy turnpike behavior and that localized grids on $[0, \tau]$ yield an increased MPC performance. The abstract framework presented in this chapter, in particular the implicit function theorem Theorem 4.3 can, in principle, be applied to quasilinear equations. In that context, after choosing an appropriate functional analytic framework, a T-independent bound on the solution operator to the linearized equation and T-uniform continuity has to be derived.

Chapter 5

Goal oriented error estimation for Model Predictive Control

In this chapter, we illustrate how a posteriori goal oriented grid adaptivity can be used to efficiently solve the subproblems arising in a Model Predictive Control (MPC) algorithm. This is motivated by the theoretical findings of Chapters 2 to 4, cf. in particular Theorems 2.27, 2.48 and 2.55 for general linear evolution equations, Section 2.6 for linear autonomous parabolic equations, Theorem 3.14 for linear non-autonomous parabolic equations, and Corollary 4.31 for semilinear parabolic equations, where we showed that in order to obtain a low absolute error of the state and control on an initial part, the perturbations of the extremal equations only have to be small on this initial part. This directly implies that in order to have an MPC feedback of high quality, any adaptive space-time discretization scheme should predominantly refine the grid on $[0, \tau]$.

We briefly touched the subject of grid adaptivity in Sections 3.3 and 4.5, where we presented different a priori discretization techniques for MPC. The question that remained was how to determine a suitable discretization that is specialized for a Model Predictive Controller, automatically. In this chapter, we will employ a posteriori goal oriented error estimation techniques to adaptively refine the grids in every loop of the MPC algorithm to obtain highly efficient discretizations in time and space. We will illustrate the performance of this approach and show that adaptive space and time mesh refinement aiming for a small discretization error on $[0, \tau]$ leads to grids that are fine on $[0, \tau]$ and coarse on the remainder.

The aim of goal oriented error estimation techniques is to refine the time and/or space grid to reduce the error in an arbitrary functional $I(x, u)$, the so called *quantity of interest (QOI)*, with, e.g., the goal to guarantee that

$$I(x, u) - I(\tilde{x}, \tilde{u}) < \text{tol},$$

where (x, u) is the optimal solution and $(\tilde{x}, \tilde{u})$ a numerical approximation on a time and/or space grid. In the particular case of MPC, this methodology can be used to minimize the error of the MPC feedback and its influence on the state, meaning that $I(x, u)$ is a functional incorporating only $x\big|_{[0,\tau]}$ and $u\big|_{[0,\tau]}$. To this end, we present a truncated version of the cost functional as an objective for refinement that is specialized for MPC. The main objective of this chapter will be to illustrate the efficiency gain from using a goal oriented error estimation technique in a Model Predictive Controller in the sense that for a fixed number of total degrees of freedom for the solution of the OCP we will significantly reduce the closed-loop cost when using the truncated cost functional for refinement compared to using the full cost function. We will further show that the error indicators computed by the goal oriented error estimator for this truncated QOI decay exponentially outside of $[0, \tau]$.

Structure. After defining the abstract problem setting, the time and space discretization scheme and recalling basic properties of goal oriented error estimation in Section 5.1, we will present a specialized QOI for MPC in Section 5.2. We further provide an extension of the sensitivity result of, e.g., Theorem 3.14, proving that goal oriented error indicators decay exponentially in time on $[\tau, T]$ if the QOI is localized at $[0, \tau]$. In Section 5.3 we provide various numerical examples to illustrate the behavior of goal oriented error estimation specialized for MPC in time and space. We consider a linear quadratic setting in Section 5.3.1, semilinear dynamics in Section 5.3.2 and boundary controlled quasilinear dynamics in Section 5.3.3. We will compare the resulting grids, solutions and MPC closed-loop performance of goal oriented adaptivity with the cost functional as QOI to results of goal oriented adaptivity, where we choose a truncated cost functional as QOI in the context of time, space and space-time adaptivity. Finally, in Section 5.3.4, we provide implementation details and present various aspects that can be taken into account for fast adaptive MPC methods.

5.1 Setting and preliminaries

In this section, we briefly recall the parabolic optimal control problem of the previous chapter and the corresponding optimality conditions. We further present the spatial and temporal discretization scheme and recall the basics of goal oriented error estimation for parabolic optimization problems.

5.1.1 Optimal control problem and optimality conditions

Analogously to Section 4.1, suppose that $(V, \|\cdot\|_V)$ is a separable Banach space, $(H, \langle\cdot,\cdot\rangle)$ is a separable and real Hilbert space, and $V \hookrightarrow H \cong H^* \hookrightarrow V^*$ forms a Gelfand triple. We consider the optimal control problem

$$\begin{aligned} \min_{(x,u)}\ & J(x,u) := \int_0^T \bar{J}(t, x(t), u(t))\, dt \\ \text{s.t. }\ & x'(t) = \bar{A}(x(t)) + \bar{B}u(t) + f(t), \\ & x(0) = x_0, \end{aligned} \tag{5.1}$$

where $x_0 \in H$, $f \in L_2(0,T;V^*)$, and $J(x,u)$ is a twice continuously differentiable functional from $L_2(0,T;V) \times L_2(0,T;U)$ to $\mathbb{R}$. We assume that the operator $\bar{B} \in L(U,V^*)$ and $\bar{A}\colon V \to V^*$ is twice continuously Fréchet differentiable. For clarity of presentation, we consider only the case where $J_{xu} = J_{ux} = 0$, which is, e.g., the case for standard tracking type cost functionals. Again, we will assume that the optimal control problem has a solution in $W([0,T])$ and the reader is referred to the discussion in Section 4.1 for that matter. We briefly recall the definition of the Lagrange function with Lagrange multiplier $(\lambda, \lambda_0) \in L_2(0,T;V) \times H$, that is,

$$L(x,u,(\lambda,\lambda_0)) := J(x,u) + \langle x' - A(x) - Bu - f, \lambda\rangle_{L_2(0,T;V^*)\times L_2(0,T;V)} + \langle x(0) - x_0, \lambda_0\rangle, \quad (5.2)$$

where

$$\langle A(x), \lambda\rangle_{L_2(0,T;V^*)\times L_2(0,T;V)} := \int_0^T \langle \bar{A}(x), \lambda\rangle_{V^*\times V}\, dt,$$
$$\langle Bu, \lambda\rangle_{L_2(0,T;V^*)\times L_2(0,T;V)} := \int_0^T \langle \bar{B}u, \lambda\rangle_{V^*\times V}\, dt.$$

The corresponding optimality conditions read

$$L'(x,u,\lambda) = \begin{pmatrix} J_x(x,u) - \lambda' - A'(x)^*\lambda \\ \lambda(T) \\ J_u(x,u) - B^*\lambda \\ x' - A(x) - Bu - f \\ x(0) - x_0 \end{pmatrix} = 0. \quad (5.3)$$

We note that one could straightforwardly incorporate a terminal state penalization $J_T(x(T))$ into the cost functional, which would result in a nonzero terminal condition for the adjoint. We will omit this terminal cost for ease of presentation. Correspondingly, the second derivative of the Lagrange function is given by

$$L''(x,u,\lambda) = \begin{pmatrix} J_{xx}(x,u) - A''(x)^*\lambda & 0 & -\frac{d}{dt} - A'(x)^* \\ 0 & 0 & E_T \\ 0 & J_{uu}(x,u) & -B^* \\ \frac{d}{dt} - A'(x) & -B & 0 \\ E_0 & 0 & 0 \end{pmatrix}, \quad (5.4)$$

where again $E_t x = x(t)$ for $t \in [0,T]$ as defined in Definition 2.26.

5.1.2 Discretization and goal oriented error estimation

For the discretization of the infinite dimensional problem, we use a discontinuous Galerkin approach of order zero in time (denoted by dG(0)), and a continuous Galerkin approach of order one in space (denoted by cG(1)) as presented in [100, 101]. In the literature, this combined approach is often referred to as dG(0)cG(1)-discretization. We will briefly recall some of the work considering this discretization technique.

Discretization and adaptivity for parabolic equations with discontinuous Galerkin methods was first established in the seminal papers [45, 46]. For the particular case of $\bar{A}(x) = \Delta x$, a priori time and space discretization error estimates for optimal control of parabolic PDEs of order $k+1$ and $s+1$, respectively, are given in [102, Section 5.1], where k and s are the orders of the polynomials in the ansatz space in time and space, respectively. Control constraints were included in [131]. For semilinear parabolic PDEs, a priori bounds were obtained in [106] under growth conditions, whereas the case of semilinear parabolic PDEs without growth conditions was treated recently in [103]. Considering efficient numerical realization, the reader is referred to [120] for a PDE context and to [16] for the case of optimal control. Lastly, there are recent discrete maximal parabolic regularity results for the discrete-time equations, cf. [93, 94].

For the reader's convenience, we will briefly recall the definition of this discretization scheme and the corresponding a posteriori error goal oriented estimation. In the following, we will abbreviate

$$\mathcal{W} := W([0,T]), \qquad \mathcal{U} = L_2(0,T;U), \qquad \langle v, w\rangle_I := \int_I \langle v(t), w(t)\rangle_{V^*\times V}\, dt.$$

Time discretization

We split up the interval $[0,T] = \{0\} \cup I_1 \cup I_2 \cup \cdots \cup I_M$ into subintervals $I_m = (t_{m-1}, t_m]$ of corresponding size $k_m := t_m - t_{m-1}$ for $m \in \{1, \ldots, M\}$ and set $I_0 := \{0\}$, where $0 = t_0 < t_1 \cdots < t_M = T$. We define the discrete-time spaces of piecewise constant in time ansatz functions by

$$\begin{aligned}\mathcal{W}_k &:= \{v_k \in L_2(0,T;H) \,|\, v_k\big|_{I_m} \in V,\ m = 1, \ldots, M,\ v_k(0) \in H\},\\ \mathcal{U}_k &:= \{u_k \in L_2(0,T;U) \,|\, u_k\big|_{I_m} \in U,\ m = 1, \ldots, M\}.\end{aligned}$$

By continuity of elements in $\mathcal{W} = W([0,T]) \hookrightarrow C(0,T;H)$, cf. Lemma 3.4, this forms a nonconforming ansatz space, as elements of $\mathcal{W}_k$ are not necessarily continuous. However, despite the nonconformity, the important feature of Galerkin orthogonality of the difference of continuous and discrete solution to the test space is preserved, cf. [100, Remark 5.2]. To capture the possible discontinuities, we denote the right and left sided limits and the jump at time grid point t_m for $v_k \in \mathcal{W}_k$ via

$$v_{k,m}^+ := \lim_{t\to 0^+} v_k(t_m + t), \quad v_{k,m}^- := \lim_{t\to 0^+} v_k(t_m - t), \quad [v]_{k,m} := v_{k,m}^+ - v_{k,m}^-,$$

and illustrate this definition in Figure 5.1.

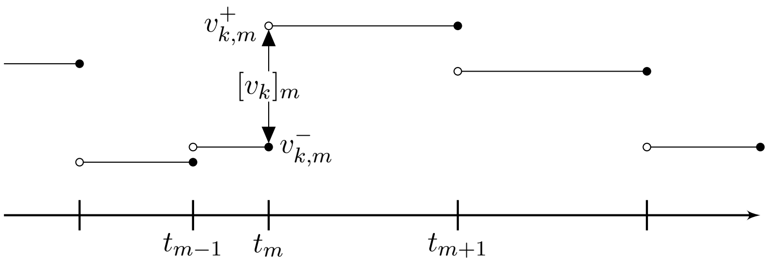


Figure 5.1: One sided limits and jumps of discrete-time variables.

Due to the nonconformity of the ansatz space, the Lagrange function defined in (5.2) is not defined on $\mathcal{W}_k$. Thus, we define the discrete-time Lagrange function $L^k\colon \mathcal{W}_k \times \mathcal{U}_k \times \mathcal{W}_k \to \mathbb{R}$ by

$$L^k(x_k, u_k, \lambda_k) := \sum_{m=1}^{M} \int_{t_{m-1}}^{t_m} \bar{J}(s, x_k, u_k)\, ds + \sum_{m=1}^{M} \big(\langle x_k', \lambda_k \rangle_{I_m} - \langle \bar{A}(x_k) - \bar{B}u_k - f, \lambda_k \rangle_{I_m} \big) \\ + \sum_{m=1}^{M} \langle [x_k]_{m-1}, \lambda_{k,m-1}^+ \rangle + \langle x_{k,0}^- - x_0, \lambda_{k,0}^- \rangle, \tag{5.5}$$

where the jump terms $[x_k]_{m-1}$ capture the discontinuities of the state. This Lagrange function is also well defined for state and adjoint state belonging to the continuous function space $\mathcal{W}$ and on this space it coincides with the continuous-time Lagrangian defined in (5.2). For piece-wise constant functions of the space $\mathcal{W}_k$, the time derivative vanishes, whereas for functions continuous in time belonging to $\mathcal{W}$, the jump terms vanish.

The discrete-time version for the state equation of (5.3) reads

$$\langle L_\lambda^k(x_k, u_k, \lambda_k), \varphi_k \rangle_{\mathcal{W}_k^* \times \mathcal{W}_k} = \sum_{m=1}^{M} \big(\langle x_k', \varphi_k \rangle_{I_m} - \langle \bar{A}(x_k) - \bar{B}u_k - f, \varphi_k \rangle_{I_m} \big) \\ + \sum_{m=1}^{M} \langle [x_k]_{m-1}, \varphi_{k,m-1}^+ \rangle + \langle x_{k,0}^- - x_0, \varphi_{k,0}^- \rangle = 0 \tag{5.6}$$

for $\varphi_k \in \mathcal{W}_k$. Analogously, the discrete-time counterpart to the third equation of (5.3), is given by

$$\langle L_u^k(x_k, u_k, \lambda_k), \varphi_k \rangle_{\mathcal{U}_k^* \times \mathcal{U}_k} = \sum_{m=1}^{M} \langle \bar{J}_u(\cdot, x_k, u_k) - \bar{B}^* \lambda_k, \varphi_k \rangle_{I_m} = 0 \tag{5.7}$$

for $\varphi_k \in \mathcal{U}_k$. Using integration by parts on each subinterval in the state equation (5.6), one can derive the adjoint equation as discrete-time counterpart to the first equation of (5.3), that is,

$$\langle L_x^k(x_k, u_k, \lambda_k), \varphi_k \rangle_{\mathcal{W}_k^* \times \mathcal{W}_k} = \sum_{m=1}^{M} \langle \bar{J}_x(\cdot, x_k, u_k), \varphi_k \rangle_{I_m} + \sum_{m=1}^{M} \big(\langle -\lambda_k' - \bar{A}'(x_k)^* \lambda_k, \varphi_k \rangle_{I_m} \\ - ([\lambda_k]_{m-1}, \varphi_{k,m-1}^-) \big) + \langle \lambda_{k,M}^-, \varphi_{k,M}^- \rangle = 0 \tag{5.8}$$

for all $\varphi_k \in \mathcal{W}_k$. The resulting time-stepping scheme is equivalent to an implicit Euler method if the temporal integrals are approximated via the box rule, cf. [100, Section 3.4.1], and thus inherits its A-stability.

Space discretization and time-stepping on dynamic meshes

For spatial discretization we use linear continuous finite elements as treated in the standard literature [23, 32, 77]. To this end, we assign a regular triangulation $\mathcal{K}_h^m$ and corresponding conforming finite element spaces $V_h^m \subset V$ and $U_h^m \subset U$ to each interval I_m and obtain the fully discrete spaces

$$\begin{aligned} \mathcal{W}_{kh} &:= \{v_{kh} \in L_2(0,T,H) \,|\, v_{kh}\big|_{I_m} \in V_h^m,\, m = 1,\ldots,M,\, v_{kh}(0) \in V_h^0\}, \\ \mathcal{U}_{kh} &:= \{u_{kh} \in L_2(0,T,U) \,|\, u_{kh}\big|_{I_m} \in U_h^m,\, m = 1,\ldots,M\}. \end{aligned} \tag{5.9}$$

Due to conformity of these spaces with respect to the discrete-time spaces, i.e., $\mathcal{W}_{kh} \subset \mathcal{W}_k$ and $\mathcal{U}_{kh} \subset \mathcal{U}_k$, the discrete-time Lagrangian (5.5) is well defined on $\mathcal{W}_{kh} \times \mathcal{U}_{kh} \times \mathcal{W}_{kh}$.

In order to allow full flexibility for the spatial adaptivity, it is possible that the triangulation $\mathcal{K}_h^m$ on the interval I_m is different from the triangulation $\mathcal{K}_h^{m-1}$ on the interval I_{m-1}. In terms of numerical realization, this leads to difficulties in efficiently evaluating the scalar product of basis elements of different time steps as needed for the assembly of the Euler step equations (5.6) and (5.8). A remedy is presented in [128], where the authors suggest the evaluation of scalar products on a common triangulation of $\mathcal{K}_h^m$ and $\mathcal{K}_h^{m-1}$, which we denote by $\mathcal{K}_h^{m-1/2}$. This common triangulation is depicted in Figure 5.2, where the original meshes have been independently *red-green refined*, cf. [40, Section 6.2.2] and [11]. If both meshes stem from the same original mesh by refinement, then the common refinement leads to a regular triangulation and to a finite element space $V_h^{m-1/2}$ such that $V_h^{m-1}, V_h^m \subset V_h^{m-1/2}$.

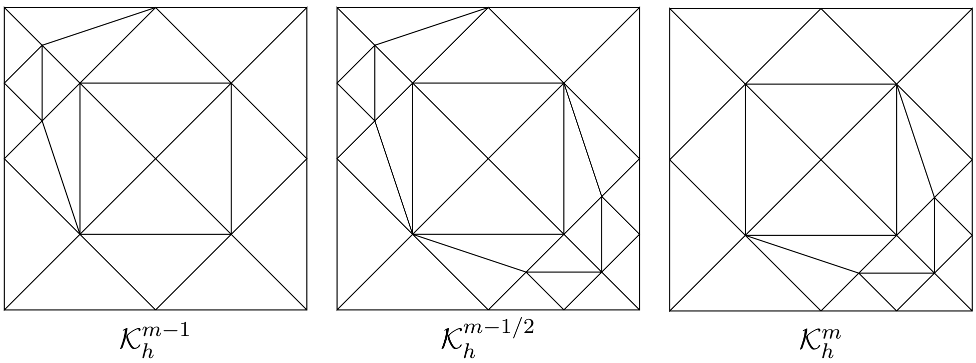


Figure 5.2: Sketch of common refinement $\mathcal{K}_h^{m-1/2}$ of two triangulations $\mathcal{K}_h^{m-1}$ and $\mathcal{K}_h^m$.

In our case, this common refinement is computed by the module *dune-gridglue* [14] of the *DUNE* C++-library [20] and allows us to compute scalar products of basis elements $\psi_m \in V_h^m$

and $\psi_{m-1} \in V_h^{m-1}$ via

$$\int_\Omega \psi_m \psi_{m-1} = \sum_{K \in \mathcal{K}_h^{m-1/2}} \int_K \psi_m \psi_{m-1}. \tag{5.10}$$

By construction of the grids, for each cell $K \in \mathcal{K}_h^{m-1/2}$, there are corresponding parent cells $K^m \in \mathcal{K}_h^m$ and $K^{m-1} \in \mathcal{K}_h^{m-1}$ such that $K \subset K^m$ and $K \subset K^{m-1}$. The *dune-gridglue* module provides the index of the associated parent cells K^m and K^{m-1} in each original mesh. Thus, the integral over cells of the commonly refined triangulation in (5.10) can be evaluated efficiently with local evaluation in K^m and K^{m-1} and a suitable quadrature rule. Hence, the price to pay for dynamic space grids is the computation of the common triangulations and the assembly of $M-1$ mass matrices, assigning to finite element functions defined on one space grid a linear functional on a neighboring space grid. The algorithms completing these tasks can be implemented in parallel using all available CPU-cores. Further, after refinement of space grid $\mathcal{K}_h^m$, only the common refinements $\mathcal{K}_h^{m-1/2}$ and $\mathcal{K}_h^{m+1/2}$ and the corresponding mass matrices need to be updated. We will discuss this topic in detail in Section 5.3.4.

Goal oriented error estimation

We will now introduce the concept of goal oriented error estimation for optimal control of parabolic PDEs. There are a lot of works considering goal oriented error estimation starting with the seminal papers [15, 17, 18], which were extended to systems with state or control constraints [116], optimal control of hyperbolic equations [86] and optimal control of parabolic equations [100, 101, 102]. A comprehensive introduction to adaptive finite element methods for ODEs and PDEs with applications is given in the monograph [10]. The main idea of goal oriented error estimation is to estimate and reduce the discretization error with respect to an arbitrary functional $I(x, u)$, called the quantity of interest (QOI). Motivations for the definition of QOIs range from allowing error estimation outside of the usual energy norm for, e.g., flow simulation in the PDE case [18, 78] to the case of optimal control, where applications include parameter estimation and optimal choice of regularization parameters [102, 141] to the standard case of choosing the cost functional as the QOI.

We follow the literature [100, 101] and denote by $(x, u, \lambda) \in (\mathcal{W} \times \mathcal{U} \times \mathcal{W})$ a continuous-time solution of the extremal equations (5.3), by $(x_k, u_k, \lambda_k) \in (\mathcal{W}_k \times \mathcal{U}_k \times \mathcal{W}_k)$ and by $(x_{kh}, u_{kh}, \lambda_{kh}) \in (\mathcal{W}_{kh} \times \mathcal{U}_{kh} \times \mathcal{W}_{kh})$ time and fully discrete solutions of the system described by (5.6), (5.7), and (5.8). One intermediate aim of goal oriented a posteriori error estimation is to derive error estimators η_k and η_h such that

$$I(x, u) - I(x_{kh}, u_{kh}) \approx \eta_k + \eta_h,$$

where η_k approximates the time discretization error and η_h approximates the space discretization error. A detailed derivation of the estimators is performed in [100, Chapter 6] and [101]. We briefly recall the main steps for the convenience of the reader and for later use. For more details, the interested reader is referred to the references above. Besides the solution triple $\xi := (x, u, \lambda)$ of the first-order necessary conditions, a second triple of variables $\chi := (v, q, z)$ has to be considered. These *secondary variables* solve the linear system

$$L''(\xi)\chi = (L^k)''(\xi)\chi = -\begin{pmatrix} I_x'(x,u) \\ 0 \\ I_u'(x,u) \\ 0 \\ 0 \end{pmatrix} \qquad \text{in } \mathcal{W}^* \times H \times \mathcal{U}^* \times \mathcal{W}^* \times H, \tag{5.11}$$

on the continuous-time level, the system

$$(L^k)''(\xi_k)\chi_k = -\begin{pmatrix} I_x'(x_k,u_k) \\ 0 \\ I_u'(x_k,u_k) \\ 0 \\ 0 \end{pmatrix} \qquad \text{in } \mathcal{W}_k^* \times H \times \mathcal{U}_k^* \times \mathcal{W}_k^* \times H, \tag{5.12}$$

on the discrete-time level, and the system

$$(L^k)''(\xi_{kh})\chi_{kh} = -\begin{pmatrix} I_x'(x_{kh},u_{kh}) \\ 0 \\ I_u'(x_{kh},u_{kh}) \\ 0 \\ 0 \end{pmatrix} \qquad \text{in } \mathcal{W}_{kh}^* \times V_h^0 \times \mathcal{U}_{kh}^* \times \mathcal{W}_{kh}^* \times V_h^M. \tag{5.13}$$

on the fully discrete level. These equations are similar to the defining equation of a Lagrange-Newton step, where the derivative of the Lagrangian on the right hand side is replaced by the derivative of the QOI.

With the continuous triples $\xi = (x, u, \lambda)$ and $\chi = (v, q, z)$ and the corresponding discrete counterparts, we define the residual of the first-order optimality condition via

$$\begin{aligned} \rho^\lambda(x,u,\lambda)\varphi &:= \langle L_x^k(x,u,\lambda), \varphi\rangle_{\mathcal{W}_k^*\times\mathcal{W}_k}, \\ \rho^u(x,u,\lambda)\varphi &:= \langle L_u^k(x,u,\lambda), \varphi\rangle_{\mathcal{U}_k^*\times\mathcal{U}_k}, \\ \rho^x(x,u,\lambda)\varphi &:= \langle L_\lambda^k(x,u,\lambda), \varphi\rangle_{\mathcal{W}_k^*\times\mathcal{W}_k}, \end{aligned}$$

and a residual involving the secondary variables $\chi = (v, q, z)$ via

$$\begin{aligned} \rho^z(\xi,v,q,z)\varphi &:= L_{\lambda x}^k(\xi)(z,\varphi) + L_{ux}^k(\xi)(q,\varphi) + L_{xx}^k(\xi)(v,\varphi) + I_x'(x,u)\varphi, \\ \rho^q(\xi,v,q,z)\varphi &:= L_{uu}^k(\xi)(q,\varphi) + L_{xu}^k(\xi)(v,\varphi) + L_{\lambda u}^k(\xi)(z,\varphi) + I_u'(x,u)\varphi, \\ \rho^v(\xi,v,q)\varphi &:= L_{x\lambda}^k(\xi)(v,\varphi) + L_{u\lambda}^k(\xi)(q,\varphi). \end{aligned}$$

With these residuals, the time discretization error can be estimated via

$$\begin{aligned}&I(x,u)-I(x_k,u_k)\approx\\&\frac{1}{2}\big(\rho^\lambda(x_k,u_k,\lambda_k)(v-\underline{v}_k)+\rho^u(x_k,u_k,\lambda_k)(q-\underline{q}_k)+\rho^x(x_k,u_k)(z-\underline{z}_k)\\&+\rho^z(\xi_k,v_k,q_k,z_k)(x-\underline{x}_k)+\rho^q(\xi_k,v_k,q_k,z_k)(u-\underline{u}_k)+\rho^v(\xi_k,v_k,q_k)(\lambda-\underline{\lambda}_k)\big)\end{aligned}$$

for $(\underline{v}_k,\underline{q}_k,\underline{z}_k),(\underline{x}_k,\underline{u}_k,\underline{\lambda}_k)\in\mathcal{W}_k\times\mathcal{U}_k\times\mathcal{W}_k$ arbitrary. Similarly, the space discretization error estimator can be approximated via

$$\begin{aligned}I(x_k,u_k)-I(x_{kh},u_{kh})\approx&\\&\frac{1}{2}\big(\rho^\lambda(x_{kh},u_{kh},\lambda_{kh})(v_k-\underline{v}_{kh})+\rho^u(x_{kh},u_{kh},\lambda_{kh})(q_k-\underline{q}_{kh})\\&+\rho^x(x_{kh},u_{kh})(z_k-\underline{z}_{kh})+\rho^z(\xi_{kh},v_{kh},q_{kh},z_{kh})(x_k-\underline{x}_{kh})\\&+\rho^q(\xi_{kh},v_{kh},q_{kh},z_{kh})(u_k-\underline{u}_{kh})+\rho^v(\xi_{kh},v_{kh},q_{kh})(\lambda_k-\underline{\lambda}_{kh})\big)\end{aligned}$$

for $(\underline{v}_{kh},\underline{q}_{kh},\underline{z}_{kh}),(\underline{x}_{kh},\underline{u}_{kh},\underline{\lambda}_{kh})\in\mathcal{W}_{kh}\times\mathcal{U}_{kh}\times\mathcal{W}_{kh}$. The arbitrary choice of the test functions originates in Galerkin orthogonality, cf. [101, Proposition 4.1, Theorem 4.3]. The terms $v-\underline{v}_k$, $q-\underline{q}_k$, $z-\underline{z}_k$, $x-\underline{x}_k$, $u-\underline{u}_k$, $\lambda-\underline{\lambda}_k$ resp. $v_k-\underline{v}_{kh}$, $q_k-\underline{q}_{kh}$, $z_k-\underline{z}_{kh}$, $x_k-\underline{x}_{kh}$, $u_k-\underline{u}_{kh}$, $\lambda_k-\underline{\lambda}_{kh}$ are often called weights and need to be approximated to obtain computable error estimates as the solutions in the infinite dimensional spaces, i.e., expressions with no subscript or subscript k, are not at hand. Approximating the weights by elements of $\mathcal{W}_k$ and $\mathcal{W}_{kh}$, respectively, causes the estimators to vanish due to Galerkin orthogonality. Hence, will discuss options to efficiently approximate the weights in Section 5.3.4. Having approximated the weights for the time discretization error by w_v^k, w_q^k, w_z^k, w_x^k, w_u^k and w_λ^k and the weights for the space discretization error by w_v^h, w_q^h, w_z^h, w_x^h, w_u^h and w_λ^h we define the error indicators by

$$\begin{aligned}\eta_k:=&\frac{1}{2}\big(\rho^\lambda(x_{kh},u_{kh},\lambda_{kh})(w_v^k)+\rho^u(x_{kh},u_{kh},\lambda_{kh})(w_q^k)+\rho^x(x_{kh},u_{kh})(w_z^k)\\&+\rho^z(\xi_{kh},v_{kh},q_{kh},z_{kh})(w_x^k)+\rho^q(\xi_{kh},v_{kh},q_{kh},z_{kh})(w_u^k)+\rho^v(\xi_{kh},v_{kh},q_{kh})(w_\lambda^k)\big)\end{aligned}\tag{5.14}$$

and

$$\begin{aligned}\eta_h:=&\frac{1}{2}\big(\rho^\lambda(x_{kh},u_{kh},\lambda_{kh})(w_v^h)+\rho^u(x_{kh},u_{kh},\lambda_{kh})(w_q^h)+\rho^x(x_{kh},u_{kh})(w_z^h)\\&+\rho^z(\xi_{kh},v_{kh},q_{kh},z_{kh})(w_x^h)\ +\rho^q(\xi_{kh},v_{kh},q_{kh},z_{kh})(w_u^h)+\rho^v(\xi_{kh},v_{kh},q_{kh})(w_\lambda^h)\big).\end{aligned}\tag{5.15}$$

5.2 Exponential decay of error indicators

Having introduced the concept of goal oriented error estimation, we will present a quantity of interest particularly well suited for the adaptive solution of the optimal control problems in a Model Predictive Controller as described in Algorithm 1. In every iteration of the MPC loop, the control on $[0,\tau]$ is used as feedback. Hence, we suggest using a truncation of the cost functional as a quantity of interest, namely

$$I^\tau(x,u) := \int_0^\tau \bar{J}(t,x,u)\,dt. \tag{5.16}$$

This specialized quantity of interest in goal oriented error estimation yields time and space grids such that the error of the MPC feedback is small. The stability results in Chapters 2 to 4 suggest that the propagation of discretization errors over time is exponentially damped for optimal control problems satisfying stabilizability and detectability conditions. In other words, to obtain a low error in $I^\tau(x,u)$, it is expected that it suffices to use a fine grid in space and time on $[0,\tau]$ that becomes coarser towards T. In this part, we will prove that the error indicators η_k and η_h defined in (5.14) and (5.15) for the QOI defined in (5.16) decay exponentially outside the interval $[0,\tau]$. First, we observe that by the linear dependence, the error indicators inherit the behavior of the secondary variables. Thus, it suffices to analyze the behavior of the continuous version of these variables, i.e., $\chi = (v,q,z)$ defined in (5.11) or the discrete-time version $\chi_k = (v_k,q_k,z_k)$ defined in (5.12). In these defining equations, we observe that the right hand side depends on the derivatives of the QOI. In case of a QOI as defined in (5.16), these functionals only integrate over a small part of the time horizon if $\tau \ll T$. In the following we will show that the continuous-time secondary variables $\chi = (v,q,z)$ defined in (5.11) or the discrete-time secondary variables $\xi_k = (v_k,q_k,z_k)$ defined in (5.12) inherit the locality of the QOI in the sense that they are large on $[0,\tau]$ and small on $[\tau,T]$. As the involved operator in the defining equation for the secondary variables is the second derivative of the Lagrange function, the assumptions made in this part will be of the same nature as in the nonlinear case considered in Chapter 4.

Assumption 5.1. *Let (x,u,λ) be a solution of the optimality system (5.3). Assume the following:*

- *There is a Hilbert space $(Y,\langle\cdot,\cdot\rangle_Y)$ and an operator $C \in L(L_2(0,T;V),L_2(0,T;Y))$ such that $L_{xx}(x,u) = C^*C$.*
- *There is an operator $R \in L(L_2(0,T;U),L_2(0,T;U))$ satisfying $\|Ru\|_{L_2(0,T;U)} \ge \alpha\|u\|_U$ for $\alpha > 0$ such that $J_{uu}(x,u) = R^*R$.*
- *$A'(x) \in L(L_2(0,T;V),L_2(0,T;V^*))$.*
- *$(A'(x),C)$ and $(A'(x),B)$ are V-exponentially stabilizable in the sense of Definition 3.20, i.e., there are feedback operators $K_C \in L(L_2(0,T;Y),L_2(0,T;V^*))$ and $K_B \in L(L_2(0,T;V),L_2(0,T;U))$ and a constant $\alpha > 0$ such that*

$$-\langle A'(x) + K_C C v, v\rangle_{L_2(0,T;V^*)\times L_2(0,T;V)} \ge \alpha\|v\|^2_{L_2(0,T;V)},$$
$$-\langle A'(x) + BK_B v, v\rangle_{L_2(0,T;V^*)\times L_2(0,T;V)} \ge \alpha\|v\|^2_{L_2(0,T;V)}$$

  for all $v \in L_2(0,T;V)$.

Theorem 5.2. *Let Assumption 5.1 hold. Consider the QOI $I^\tau(x,u)$ defined in (5.16). Let $(v,q,z) \in W([0,T]) \times L_2(0,T;U) \times W([0,T])$ solve (5.11), i.e.,*

$$L''(x,u,\lambda)\begin{pmatrix} v\\ q\\ z\end{pmatrix} = \begin{pmatrix} L_{xx}(x,u) & 0 & -\frac{d}{dt} - A'(x)^* \\ 0 & 0 & E_T \\ 0 & J_{uu}(x,u) & -B^* \\ \frac{d}{dt} - A'(x) & -B & 0 \\ E_0 & 0 & 0 \end{pmatrix}\begin{pmatrix} v\\ q\\ z\end{pmatrix} = -\begin{pmatrix} I^\tau_x(x,u) \\ 0 \\ I^\tau_u(x,u) \\ 0 \\ 0\end{pmatrix}.$$

Then, setting

$$M := \begin{pmatrix} C^*C & -\frac{d}{dt} - A'(x)^* \\ 0 & E_T \\ \frac{d}{dt} - A'(x) & -BJ_{uu}(x,u)^{-1}B^* \\ E_0 & 0 \end{pmatrix},$$

the solution operator norm $\|M^{-1}\|_{L((L_2(0,T;V^*)\times H)^2, W([0,T])^2)}$ can be bounded independently of T. Further, for all $\mu > 0$ satisfying

$$\mu < \frac{1}{\|M^{-1}\|_{L((L_2(0,T;V^*)\times H)^2, W([0,T])^2)}}$$

there is a constant $c(\tau) > 0$ independent of T such that

$$\left\| e^{\mu t}\begin{pmatrix} v\\ q\\ z\end{pmatrix}\right\|_{W([0,T])\times L_2(0,T;U)\times W([0,T])} \leq c(\tau)\big(\|J_x(x,u)\|_{L_2(0,\tau;V^*)} + \|J_u(x,u)\|_{L_2(0,\tau;U)}\big). \quad (5.17)$$

Proof. We first rewrite the system by eliminating the control via $q = J_{uu}^{-1}(x,u)\left(B^*z - I^\tau_u(x,u)\right)$ as

$$\begin{pmatrix} C^*C & -\frac{d}{dt} - A'(x)^* \\ 0 & E_T \\ \frac{d}{dt} - A'(x) & -BJ_{uu}(x,u)^{-1}B^* \\ E_0 & 0 \end{pmatrix}\begin{pmatrix} v\\ z\end{pmatrix} = -\begin{pmatrix} I^\tau_x(x,u) \\ 0 \\ BJ_{uu}(x,u)^{-1}I^\tau_u(x,u) \\ 0. \end{pmatrix}.$$

With the assumptions of V-exponential stabilizability the bound on M^{-1} follows analogously to Corollary 3.30. A bound on the variables scaled by $e^{\mu t}$ follows completely analogously to the proof of Theorem 3.14 by replacing the scaling $e^{-\mu t}$ with $e^{\mu t}$. Finally, we estimate the scaled right hand side via

$$\begin{aligned}&\left\|\int_0^\tau e^{\mu t}\bar{J}_x(t,x,u)\,dt\right\|_{L_2(0,T;V^*)} + \left\|BJ_{uu}(x,u)^{-1}\int_0^\tau e^{\mu t}\bar{J}_u(t,x,u)\,dt\right\|_{L_2(0,T;U)} \\ &\leq c(\tau)\left(\|J_x(x,u)\|_{L_2(0,\tau,V^*)} + \|J_u(x,u)\|_{L_2(0,\tau;U)}\right),\end{aligned}$$

which concludes the proof. ∎

We will derive a similar estimate in Theorem 5.6 for the discrete-time secondary variables v_k, q_k and z_k and the fully discrete secondary variables v_{kh}, q_{kh} and z_{kh}. Further, we will show in Remark 5.3 that the term on the right hand side, i.e., $\|J_x(x,u)\|_{L_2(0,\tau;V^*)} + \|J_u(x,u)\|_{L_2(0,\tau;U)}$ is bounded independently of T if a turnpike property holds. Before that, however, we give a short interpretation of the estimate (5.17): As the scaling $e^{\mu t}$ grows exponentially in time, the variables (v,q,z) have to decay exponentially in time such that the product is bounded independently of the end time T. Thus, the secondary variables (v,q,z) inherit the behavior of the QOI $I^\tau(x,u)$ being localized on $[0,\tau]$. Due to the linear dependence, this also carries over to the error indicators in (5.14) and (5.15).

Remark 5.3. *We will briefly give sufficient conditions under which the upper bound in* (5.17) *can be shown to be bounded independently of T in the case of a linear quadratic problem. It turns out that when a turnpike property holds, the initial part of the optimal solution is only affected by the horizon negligibly, if the horizon is large. Consider a time horizon $T > 0$ and the linear quadratic optimal control problem*

$$\min_{(x,u)} J(x,u) := \frac{1}{2}\int_0^T \|C(x(t)-x_d)\|_Y^2 + \|R(u(t)-u_d)\|_U^2 \, dt \qquad s.t.\ x' = Ax + Bu, \quad x(0) = x_0,$$

where $x_d \in V$ and $u_d \in U$ and all operators are time-independent. Suppose that the involved operators satisfy the stabilizability assumptions of Assumption 5.1. Then it follows by Theorem 3.16 that the state and control satisfy the turnpike estimate

$$\left\|\left(x(t)-\bar{x}, u(t)-\bar{u}\right)\right\|_{H\times U} \le c(e^{-\mu t} + e^{-\mu(T-t)})\left(\|\bar{\lambda}\| + \|\bar{x}-x_0\|\right) \tag{5.18}$$

for a.e. $t \in [0,T]$, where $(\bar{x},\bar{u})$ denotes the optimal solution of the corresponding steady state problem, $\bar{\lambda}$ is the corresponding adjoint state and $c \ge 0$ is independent of T. Hence, in particular we have

$$\left\|\left(x(t), u(t)\right)\right\|_{H\times U} \le c_1 \tag{5.19}$$

for a.e. $t \in [0,T]$ with $c_1 \ge 0$ independent of T. Thus,

$$\int_0^\tau \|x(t)-x_d\| + \|u(t)-u_d\|_U \, dt \le \tau c_1 + c_2$$

with $c_2 \ge 0$ independent of T. Hence, together with (5.19) *we get*

$$\|J_x(x,u)\|_{L_2(0,\tau;V^*)} + \|J_u(x,u)\|_{L_2(0,\tau;U)} \le c$$

with $c \ge 0$ independent of T. Finally we note that the steady state turnpike assumed in (5.18) *can be replaced by a dynamic turnpike concept and the proof remains valid, if the dynamic turnpike is bounded independently of T. In particular, for time-varying problems in discrete time, a similar property was proven in [67, Theorem 3].*

The result of Theorem 5.2 does not immediately carry over to the discrete-time secondary variables as defined in (5.12) due to the nonconformity of the discrete-time ansatz space. Thus, we will give a separate proof of this matter in the following. To this end, we introduce a suitable function space for scaled functions of $\mathcal{W}_k$ which are not necessarily piecewise constant in time.

Definition 5.4. *We define the space of functions that are weakly differentiable on every subinterval via*

$$W^M([0,T]) := \{v \in L_2(0,T;V) \,|\, v_{|I_m} \in W([t_{m-1},t_m]),\, m = 1,\dots,M,\, v(0) \in H\}$$

and endow it with the natural norm

$$\|v\|_{W^M([0,T])} = \sum_{m=1}^{M} \left(\|v\|_{W([t_{m-1},t_m])} + \|v_{m-1}^- - v_{m-1}^+\|\right) + \|v(0)\|.$$

Additionally, we define linear operators $\Lambda^k, \Lambda^{k,-} \colon W^M([0,T]) \to W^M([0,T])^*$ via the relations

$$\begin{aligned}
&\langle \Lambda^k v, \varphi\rangle_{W^M([0,T])^*\times W^M([0,T])} := \\
&\qquad \sum_{m=1}^{M} \left(\langle v', \varphi\rangle_{I_m} + \langle [v]_{m-1}, \varphi_{m-1}^+\rangle\right) - \langle A'(x)v, \varphi\rangle_{L_2(0,T;V^*)\times L_2(0,T;V)} + \langle v_0^-, \varphi_0^-\rangle, \\
&\langle \Lambda^{k,-} v, \varphi\rangle_{W^M([0,T])^*\times W^M([0,T])} := \\
&\qquad -\sum_{m=1}^{M} \left(\langle v', \varphi\rangle_{I_m} + \langle [v]_{m-1}, \varphi_{m-1}^-\rangle\right) - \langle A'(x)^* v, \varphi\rangle_{L_2(0,T;V^*)\times L_2(0,T;V)} + \langle v_M^-, \varphi_M^-\rangle.
\end{aligned}$$

It is clear that $\mathcal{W}_k \hookrightarrow W^M([0,T])$ and that $W^M([0,T])$ with the norm defined above is a Banach space due to $W([t_k,t_{k+1}]) \hookrightarrow C(t_k,t_{k+1};H)$ for all $0 \le k \le M-1$. Testing of the initial resp. terminal condition is included in the operators Λ^k resp. $\Lambda^{k,-}$ due to the terms $\langle v_0^-, \varphi_0^-\rangle$ and $\langle v_M^-, \varphi_M^-\rangle$. We first employ a T-independent invertibility result for the discrete-time operator occurring in (5.12). To this end, we note that $L_{xx} = L_{xx}^k$ and $L_{uu} = L_{uu}^k$, i.e., the second derivatives with respect to the state and control of the continuous and discrete-time Lagrange function coincide. This is because the time derivative and the jump terms enters the Lagrange function in a linear way, i.e., they vanish in the second derivative.

Theorem 5.5. *If Assumption 5.1 holds, the inverse of the operator*

$$M^k := \begin{pmatrix} L_{xx}(x,u) & \Lambda^{k,-} \\ \Lambda^k & -BJ_{uu}(x,u)^{-1}B^* \end{pmatrix}$$

can be bounded by

$$\|(M^k)^{-1}\|_{L((L_2(0,T;V^*)\times H)^2, W^M([0,T])^2)} \le c,$$

where $c \ge 0$ is T-independent constant.

Proof. We extend the proof of Corollary 3.30 to corresponding time-discretized differential operators Λ^k and $\Lambda^{k,-}$ that allow for discontinuous states. Consider the system

$$\begin{pmatrix} C^*C & \Lambda^{k,-} \\ \Lambda^k & -BJ_{uu}(x,u)^{-1}B^* \end{pmatrix} \begin{pmatrix} v \\ z \end{pmatrix} = \begin{pmatrix} (l_1, z_T) \\ (l_2, v_0) \end{pmatrix} \tag{5.20}$$

for $l_1, l_2 \in L_2(0,T;V^*)$ and $z_T, v_0 \in H$. First, we test the state equation, i.e., the second equation of (5.20) with v and obtain

$$\sum_{m=1}^{M} \left(\langle v', v \rangle_{I_m} + \langle [v]_{m-1}, v_{m-1}^+ \rangle \right) + \|v_0^-\|^2 - \langle A'(x)v + BJ_{uu}(x,u)^{-1}B^*z, v \rangle_{L_2(0,T;V^*)\times L_2(0,T;V)}$$
$$= \langle l_2, v \rangle_{L_2(0,T;V^*)\times L_2(0,T;V)} + \|v_0\|^2$$

and use the formula from Lemma 3.4 iv) and the definition of the jump terms $[v]_m := v_m^+ - v_m^-$ applied on every subinterval to compute

$$\begin{aligned}
&\sum_{m=1}^{M} \left(\langle v', v \rangle_{I_m} + \langle [v]_{m-1}, v_{m-1}^+ \rangle \right) + \|v_0^-\|^2 \\
&\quad = \sum_{m=1}^{M} \left(\frac{1}{2}\|v_m^-\|^2 - \frac{1}{2}\|v_{m-1}^+\|^2 + \|v_{m-1}^+\|^2 - \langle v_{m-1}^-, v_{m-1}^+ \rangle \right) + \|v_0^-\|^2 \\
&\quad = \sum_{m=1}^{M} \left(\frac{1}{2}\|v_m^-\|^2 - \langle v_{m-1}^-, v_{m-1}^+ \rangle + \frac{1}{2}\|v_{m-1}^+\|^2 \right) + \|v_0^-\|^2 \\
&\quad = \sum_{m=1}^{M} \left(\frac{1}{2}\|v_{m-1}^-\|^2 - \langle v_{m-1}^-, v_{m-1}^+ \rangle + \frac{1}{2}\|v_{m-1}^+\|^2 \right) + \frac{1}{2}\left(\|v_M^-\|^2 + \|v_0^-\|^2 \right) \\
&\quad = \sum_{m=1}^{M} \frac{1}{2}\|v_{m-1}^- - v_{m-1}^+\|^2 + \frac{1}{2}\left(\|v_M^-\|^2 + \|v_0^-\|^2 \right)
\end{aligned}$$

for the first three terms. Thus, adding the stabilizing feedback K_C from Definition 3.20, we obtain

$$\sum_{m=1}^{M} \frac{1}{2}\|v_{m-1}^- - v_{m-1}^+\|^2 + \frac{1}{2}\left(\|v_M^-\|^2 + \|v_0^-\|^2 \right) - \langle (A'(x) + K_C C)v, v \rangle_{L_2(0,T;V^*)\times L_2(0,T;V)}$$
$$\leq c \left(\|Cv\|_{L_2(0,T;Y)} + \|B^*z\|_{L_2(0,T;U)} + \|l_2\|_{L_2(0,T;V^*)} \right) \|v\|_{L_2(0,T;V)} + \|v_0\|^2.$$

Hence, by $L_2(0,T;V)$-ellipticity of $-(A'(x) + K_C C)$, we get

$$\begin{aligned}
&\sum_{m=1}^{M} \frac{1}{2}\|v_{m-1}^- - v_{m-1}^+\|^2 + \frac{1}{2}\left(\|v_M^-\|^2 + \|v_0^-\|^2 \right) + \|v\|^2_{L_2(0,T;V)} \\
&\qquad \leq c \left(\|Cv\|^2_{L_2(0,T;Y)} + \|B^*z\|^2_{L_2(0,T;U)} + \|l_2\|^2_{L_2(0,T;V^*)} + \|v_0\|^2 \right).
\end{aligned} \tag{5.21}$$

Analogously, we test the adjoint equation with z and compute

$$
\begin{aligned}
-\sum_{m=1}^{M} &\left(\langle z', z\rangle_{I_m} + \langle [z]_{m-1}, z_{m-1}^-\rangle\right) + \|z_M^-\|^2 \\
&= -\sum_{m=1}^{M} \left(\frac{1}{2}\|z_m^-\|^2 - \frac{1}{2}\|z_{m-1}^+\|^2 + \langle z_{m-1}^+, z_{m-1}^-\rangle - \|z_{m-1}^-\|^2\right) + \|z_M^-\|^2 \\
&= \|z_0^-\|^2 + \sum_{m=1}^{M} \left(\frac{1}{2}\|z_m^-\|^2 - \langle z_{m-1}^+, z_{m-1}^-\rangle + \frac{1}{2}\|z_{m-1}^+\|^2\right) \\
&= \frac{1}{2}\left(\|z_0^-\|^2 + \|z_M^-\|^2\right) + \sum_{m=1}^{M} \frac{1}{2}\|z_{m-1}^- - z_{m-1}^+\|^2
\end{aligned}
$$

and thus, analogously to the state, by using V-exponential stabilizability of $(A'(x), B)$, we get for the adjoint that

$$
\begin{aligned}
\sum_{m=1}^{M} \frac{1}{2}\|z_{m-1}^- - z_{m-1}^+\|^2 + \frac{1}{2}\left(\|z_M^-\|^2 + \|z_0^-\|^2\right) + \|z\|_{L_2(0,T;V)}^2 \\
\leq c(\|Cv\|_{L_2(0,T;Y)}^2 + \|B^*z\|_{L_2(0,T;U)}^2 + \|l_1\|_{L_2(0,T;V^*)}^2 + \|z_T\|^2).
\end{aligned}
\tag{5.22}
$$

It remains to estimate the term $\|Cv\|_{L_2(0,T;Y)}^2 + \|B^*z\|_{L_2(0,T;U)}^2$. To this end, we test the first equation of (5.20) with v, the second equation of (5.20) with z, subtract the latter from the former and obtain

$$
\begin{aligned}
&\|Cv\|_{L_2(0,T;Y)}^2 + \|B^*z\|_{L_2(0,T;U)}^2 \\
&\leq \left|\langle \Lambda^{k,-} z, v\rangle_{W^M([0,T])^* \times W^M([0,T])} - \langle \Lambda^k v, z\rangle_{W^M([0,T])^* \times W^M([0,T])}\right| \\
&+ (\|(l_1, z_T)\|_{L_2(0,T;V^*)\times H} + \|(l_2, v_0)\|_{L_2(0,T;V^*)\times H})\left(\|v\|_{L_2(0,T;V)} + \|v_0^-\| + \|z\|_{L_2(0,T;V)} + \|z_M^-\|\right)
\end{aligned}
\tag{5.23}
$$

We proceed to show that $\langle \Lambda^{k,-} z, v\rangle_{W^M([0,T])^* \times W^M([0,T])} = \langle \Lambda^k v, z\rangle_{W^M([0,T])^* \times W^M([0,T])}$ and compute

$$
\begin{aligned}
&\langle \Lambda^{k,-} z, v\rangle_{W^M([0,T])^* \times W^M([0,T])} \\
&= -\sum_{m=1}^{M} \left(\langle z', v\rangle_{I_m} + \langle [z]_{m-1}, v_{m-1}^-\rangle\right) + \langle A'(x)^* z, v\rangle_{L_2(0,T;V^*)\times L_2(0,T;V)} + \langle z_M^-, v_M^-\rangle \\
&= \sum_{m=1}^{M} \left(\langle z, v'\rangle_{I_m} - \langle z_m^-, v_m^-\rangle + \langle z_{m-1}^+, v_{m-1}^+\rangle - \langle z_{m-1}^+ - z_{m-1}^-, v_{m-1}^-\rangle\right) \\
&\qquad + \langle A'(x) v, z\rangle_{L_2(0,T;V^*)\times L_2(0,T;V)} + \langle z_M^-, v_M^-\rangle \\
&= \sum_{m=1}^{M} \left(\langle z, v'\rangle_{I_m} + \langle z_{m-1}^+, v_{m-1}^+ - v_{m-1}^-\rangle\right) + \langle A'(x) v, z\rangle_{L_2(0,T;V^*)\times L_2(0,T;V)} + \langle z_0^-, v_0^-\rangle \\
&= \langle \Lambda^k v, z\rangle_{W^M([0,T])^* \times W^M([0,T])}.
\end{aligned}
$$

The interested reader is referred to a similar result in [127, Proposition 3.6] in a continuous-time setting. Thus, together with (5.21), (5.22), and (5.23) we obtain with $c \geq 0$ independent of T:

$$\begin{aligned} \sum_{m=1}^{M} \frac{1}{2}\|v_{m-1}^{-} - v_{m-1}^{+}\|^2 + \frac{1}{2}\left(\|v_M^{-}\|^2 + \|v_0^{-}\|^2\right) + \sum_{m=1}^{M} \frac{1}{2}\|z_{m-1}^{-} - z_{m-1}^{+}\|^2 \\ + \frac{1}{2}\left(\|z_M^{-}\|^2 + \|z_0^{-}\|^2\right) + \|(v,z)\|^2_{L_2(0,T;V)^2} \leq c\|(l_1, z_T, l_2, v_0)\|^2_{(L_2(0,T;V^*)\times H)^2} \end{aligned} \tag{5.24}$$

To obtain an estimate on the derivatives, we test the state equation with a test function $\varphi_m \in C^\infty([t_{m-1}, t_m]; V)$ such that $\varphi(t_{m-1}) = \varphi(t_m) = 0$ and obtain

$$\sum_{m=1}^{M} \langle v', \varphi \rangle_{I_m} = \langle BJ_{uu}^{-1}B^*z + l_2 + A'(x)v, \varphi \rangle_{L_2(0,T;V^*)\times L_2(0,T;V)}.$$

By density of $C_0^\infty([t_{m-1}, t_m]; V)$ in $L_2(t_{m-1}, t_m; V)$, cf. [127, Lemma 2.1], we conclude the estimate

$$\begin{aligned} &\|v'\|_{L_2(t_{m-1},t_m;V^*)} \leq \big(\|A'(x)\|_{L(L_2(t_{m-1},t_m;V),L_2(t_{m-1},t_m;V^*))} \\ &+ \|BJ_{uu}^{-1}B^*\|_{L(L_2(t_{m-1},t_m;V),L_2(t_{m-1},t_m;V^*))}\big)\|(v,z)\|_{L_2(t_{m-1},t_m;V)^2} + \|l_2\|_{L_2(t_{m-1},t_m;V^*)}, \end{aligned}$$

which, together with (5.24) and proceeding analogously for the adjoint, yields the result. □

We now obtain an analogous result to Theorem 5.2 for the discrete-time system.

Theorem 5.6. *Let Assumption 5.1 hold and consider the QOI $I^\tau(x,u)$ defined in (5.16). Let $(v_k, q_k, z_k) \in W^M([0,T]) \times L_2(0,T;U) \times W^M([0,T])$ solve (5.12), i.e.,*

$$\begin{pmatrix} L_{xx}(x,u) & 0 & \Lambda^{k,-} \\ 0 & J_{uu}(x,u) & -B^* \\ \Lambda^k & -B & 0 \end{pmatrix} \begin{pmatrix} v_k \\ q_k \\ z_k \end{pmatrix} = -\begin{pmatrix} I_x^\tau(x,u) \\ I_u^\tau(x,u) \\ 0 \end{pmatrix}. \tag{5.25}$$

Then for all $\mu > 0$ satisfying

$$\mu < \frac{1}{\|(M^k)^{-1}\|_{L((L_2(0,T;V^*)\times H)^2, W^M([0,T])^2)}},$$

there is a constant $c(\tau) > 0$ independent of T such that

$$\left\| e^{\mu t} \begin{pmatrix} v_k \\ q_k \\ z_k \end{pmatrix} \right\|_{W^M([0,T])\times L_2(0,T;U)\times W^M([0,T])} \leq c(\tau)\left(\|J_x(x,u)\|_{L_2(0,\tau;V^*)} + \|J_u(x,u)\|_{L_2(0,\tau;U)}\right). \tag{5.26}$$

Proof. Again, we eliminate the control via $q_k = J_{uu}^{-1}(x,u)\left(B^* z_k + I_u(x,u)\right)$ and obtain

$$\underbrace{\begin{pmatrix} C^*C & \Lambda^{k,-} \\ \Lambda^k & -BJ_{uu}(x,u)^{-1}B^* \end{pmatrix}}_{=M^k} \begin{pmatrix} v_k \\ z_k \end{pmatrix} = -\begin{pmatrix} I_x^\tau(x,u) \\ BJ_{uu}(x,u)^{-1} I_u^\tau(x,u) \end{pmatrix}.$$

We further choose $\mu < \frac{1}{\|(M^k)^{-1}\|_{L((L_2(0,T;V^*)\times H)^2, W^M([0,T])^2)}}$ independently of T, cf. Theorem 5.5, introduce scaled variables $\tilde{v}_k = e^{\mu t} v_k$ and $\tilde{z}_k = e^{\mu t} z_k$ and compute that

$$\begin{aligned}
&\langle \Lambda^k v_k, \varphi\rangle_{W^M([0,T])^*\times W^M([0,T])} = \langle \Lambda^k (e^{-\mu t}\tilde{v}_k), \varphi\rangle_{W^M([0,T])^*\times W^M([0,T])} \\
&= \sum_{m=1}^M \langle (e^{-\mu t}\tilde{v}_k)', \varphi\rangle_{I_m} + \langle [e^{-\mu t}\tilde{v}_k]_{m-1}, \varphi_{m-1}^+\rangle - \langle A'(x) e^{-\mu t}\tilde{v}_k, \varphi\rangle_{L_2(0,T;V^*)\times L_2(0,T;V)} \\
&\qquad + \langle (e^{-\mu t}\tilde{v}_k)_0^-, \varphi_0^-\rangle \\
&= \sum_{m=1}^M \langle (\tilde{v}_k)' - \mu \tilde{v}_k, e^{-\mu t}\varphi\rangle_{I_m} + \langle [\tilde{v}_k]_{m-1}, e^{-\mu t}\varphi_{m-1}^+\rangle - \langle A'(x)\tilde{v}_k, e^{-\mu t}\varphi\rangle_{L_2(0,T;V^*)\times L_2(0,T;V)} \\
&\qquad + \langle (\tilde{v}_k)_0^-, (e^{-\mu t}\varphi)_0^-\rangle \\
&= \langle (\Lambda^k - \mu I)\tilde{v}_k, \tilde{\varphi}\rangle_{W^M([0,T])^*\times W^M([0,T])},
\end{aligned}$$

where $\tilde{\varphi} = e^{-\mu t}\varphi$. Proceeding analogously for the adjoint equation we get

$$\begin{aligned}
M^k \begin{pmatrix} v_k \\ z_k \end{pmatrix} &= -\begin{pmatrix} I_x(x,u) \\ BJ_{uu}(x,u)^{-1} I_u(x,u) \end{pmatrix} \\
(M^k + \mu P) \begin{pmatrix} \tilde{v}_k \\ \tilde{z}_k \end{pmatrix} &= -\begin{pmatrix} \int_0^\tau e^{\mu t} \bar{J}_x(t,x,u)\cdot\, dt \\ BJ_{uu}(x,u)^{-1} \int_0^\tau e^{\mu t} \bar{J}_u(t,x,u)\cdot\, dt \end{pmatrix}
\end{aligned}$$

where $P = \begin{pmatrix} 0 & I \\ -I & 0 \end{pmatrix}$ and hence $\|P\|_{L(W^M([0,T])^2, (L_2(0,T;V^*)\times H)^2)} < 1$. Completely analogously to the proof of, e.g., Theorem 3.14, we multiply the equation by $(M^k)^{-1}$ and employ a Neumann-series argument as $\mu < \frac{1}{\|(M^k)^{-1}\|_{L((L_2(0,T;V^*)\times H), W^M([0,T])^2)}}$ and obtain

$$\begin{aligned}
&\left\| \begin{pmatrix} v_k \\ z_k \end{pmatrix} \right\|_{W^M([0,T])^2} \\
&\quad \le \|(I + \mu (M^k)^{-1} P)^{-1}\|_{L(W^M([0,T])^2, W^M([0,T])^2)} \|(M^k)^{-1}\|_{L((L_2(0,T;V^*)\times H)^2, W^M([0,T])^2)} \\
&\qquad \left\| \begin{pmatrix} \int_0^\tau e^{\mu t} \bar{J}_x(t,x,u)\cdot\, dt \\ BJ_{uu}(x,u)^{-1} \int_0^\tau e^{\mu t} \bar{J}_u(t,x,u)\cdot\, dt \end{pmatrix} \right\|_{L_2(0,T;V^*)^2} \\
&\quad \le c(\tau)\left(\|J_x(x,u)\|_{L_2(0,\tau,V^*)} + \|J_u(x,u)\|_{L_2(0,\tau;U)} \right)
\end{aligned}$$

with a constant $c > 0$ independent of T. For the control, we compute

$$\begin{aligned}
\|q_k\|_{L_2(0,T;U)} &= \|J_{uu}(x,u)^{-1} B^* z_k + J_{uu}(x,u)^{-1} I_u(x,u)\|_{L_2(0,T;U)} \\
&\le c(\tau)\left(\|J_x(x,u)\|_{L_2(0,\tau,V^*)} + \|J_u(x,u)\|_{L_2(0,\tau;U)} \right),
\end{aligned}$$

which concludes the proof. □

We will briefly illustrate the exponential decay of the secondary variables proven in Theorems 5.2 and 5.6 for the linear quadratic problem of Section 3.3.1. The plots in Figure 5.3 show the exponential decay of the linearly interpolated discrete-time secondary variables for the QOI $I^\tau(x,u)$ defined in (5.16). In Figure 5.3, we observe that for all values of the Tikhonov parameter α, the state and the control decay exponentially after the time $\tau = 0.5$. The ledges in the plot are introduced by the tolerance of the linear solver used for solution of the linear system (5.13). The smaller we choose α, the faster the secondary variables decay in time. The reason for this can be found when inspecting the proof of, e.g., Lemma 3.25, showing that $\|M^{-1}\|_{L((L_2(0,T;V^*)\times H)^2, W([0,T])^2)}$ is proportional to $\|R\|_{L(L_2(0,T;U))} = \alpha$. Thus, decreasing α allows for a larger choice of the scaling parameter $\mu > 0$ in Theorem 5.2 due to the bound

$$\mu < \frac{1}{\|M^{-1}\|_{L((L_2(0,T;V^*)\times H)^2, W([0,T])^2)}}.$$

This straightforwardly carries over to the discrete-time setting considered in Theorems 5.5 and 5.6. Further, as the decay parameter μ is chosen in the same fashion in the turnpike results, i.e., e.g., Theorem 3.16, also the speed of exponential convergence to the turnpike is increased when decreasing α, which is in accordance with the findings of [3, Table 3.1, p. 41] where a lower stabilizing horizon for MPC could be chosen when decreasing the Tikhonov parameter α.

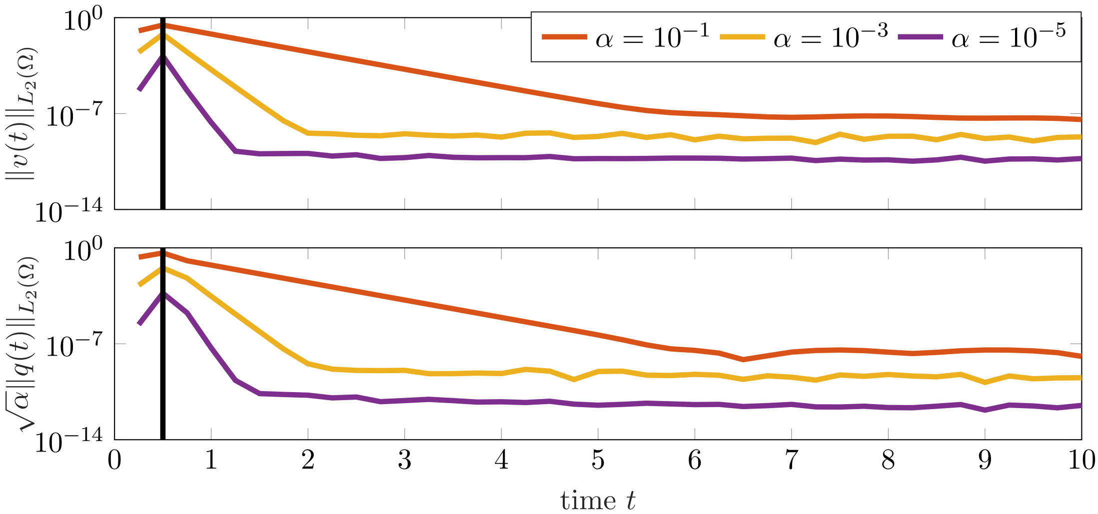


Figure 5.3: Norm of the secondary state v and control q of (5.12) over time. The vertical black line indicates the implementation horizon $\tau = 0.5$.

5.3 Numerical results

In this part we qualitatively and quantitatively examine the results of goal oriented error estimation with the specialized QOI defined in (5.16), i.e.,

$$I^\tau(x,u) := \int\limits_0^\tau \bar{J}(t,x,u)\,dt$$

and compare it with classical error estimation using the full cost functional as QOI, i.e.,

$$J(x,u) = \int\limits_0^T \bar{J}(t,x,u)\,dt.$$

We inspect the error indicators for time, space and space-time adaptivity, the resulting grids and the performance of a Model Predictive Controller evaluated via the cost functional value of the MPC trajectory, using goal oriented error estimation with both QOIs used for adaptivity in every solution of an OCP.

We briefly comment on adaptive time discretization if the solutions satisfy steady state turnpike behavior. If the optimal control problem is autonomous and satisfies a stabilizability and detectability condition, we showed in Theorems 2.30 and 3.16 for linear quadratic problems that the optimal triple exhibits turnpike behavior. This was extended to the nonlinear case in Corollary 4.30. We further observed this property numerically in Figures 3.2, 4.1 and 4.3. An important property of turnpike behavior is that the approaching and leaving arcs' lengths are independent of the time horizon. As in between these transient arcs the solution stays close to an equilibrium of the dynamics, any adaptive time discretization scheme will predominantly refine the time grid at the beginning and the end of the time interval to resolve dynamic parts. Further, the independence of leaving and approaching arc of the size of the interval suggests that in case of time adaptivity, the resulting computational cost will be almost independent of the length of the interval. This naturally suggests that in case of a steady state turnpike, an adaptive time discretization can be very efficient. Moreover, the turnpike property was also exploited in [136] to construct an efficient shooting algorithm. These considerations are however not applicable when considering non-autonomous problems that do not possess a corresponding static optimization problem and hence no optimal equilibrium. One then has to rely on classical a posteriori grid refinement techniques to adaptively refine the time and space grid.

Problem setting

In the following, we fix $\Omega = [0,3] \times [0,1]$ and the time horizon $T = 10$. We briefly recall the reference trajectories as defined in Section 3.3. That is, using the function

$$g(s) := \begin{cases} 10e^{1-\frac{1}{1-s^2}} & s < 1 \\ 0 & \text{else,} \end{cases}$$

we construct a static reference trajectory via

$$x_{\mathrm{d}}^{\mathrm{stat}}(\omega) := g\left(\frac{10}{3}\left\|\omega - \begin{pmatrix}1.5\\0.5\end{pmatrix}\right\|\right), \tag{5.27}$$

and a dynamic reference trajectory via

$$x_{\mathrm{d}}^{\mathrm{dyn}}(t,\omega) := g\left(\frac{10}{3}\left\|\omega - \begin{pmatrix}\omega_{1,\mathrm{peak}}(t)\\\omega_{2,\mathrm{peak}}(t)\end{pmatrix}\right\|\right), \tag{5.28}$$

where

$$\omega_{1,\mathrm{peak}}(t) := 1.5 - \cos\left(\pi\left(\frac{t}{10}\right)\right), \qquad w_{2,\mathrm{peak}}(t) := \left|\cos\left(\pi\left(\frac{t}{10}\right)\right)\right|.$$

The static trajectory $x_{\mathrm{d}}^{\mathrm{stat}}$ and the dynamic trajectory $x_{\mathrm{d}}^{\mathrm{dyn}}$ are depicted in Figure 3.1 and Figure 3.4, respectively. We will further consider examples with a reference concentrated towards the boundary that grows exponentially in time, i.e.,

$$x_{\mathrm{d}}^{\mathrm{exp}}(t,\omega) := e^{\frac{t}{2}} g\left(\frac{10}{3}\left\|\omega - \begin{pmatrix}1.5\\\omega_{2,\mathrm{peak}}\end{pmatrix}\right\|\right), \tag{5.29}$$

where $\omega_{2,\mathrm{peak}} \in [0,1]$ will be specified later. We consider the cost functional

$$\int_0^T \bar{J}(t,x,u)\,dt := \frac{1}{2}\int_0^T \|x(t) - x_{\mathrm{d}}(t)\|_{L_2(\Omega)}^2 + \alpha\|u(t)\|_U^2\,dt, \tag{5.30}$$

where x_{d} is one of the reference trajectories defined above and $\alpha > 0$ is a Tikhonov parameter. Depending on the governing dynamics, we will set $U = L_2(\Omega)$ for the case of distributed control and $U = L_2(\partial\Omega)$ in the case of boundary control. Whenever we use the autonomous reference trajectory defined in (5.27), we will use the implementation horizon $\tau = 0.5$. In case of the non-autonomous trajectory (5.28), we consider a larger implementation horizon $\tau = 1$.

#uniform refs.	0	1	2	3	4	5
#Triangles	12	48	192	768	3072	12288
#Vertices	11	33	113	417	1602	6273

Table 5.1: Number of elements and degrees of freedom for different hierarchies of the spatial grid.

In the following we will consider different linear and nonlinear unstable and stable dynamics with distributed and boundary control. We apply the MPC Algorithm 1 to these different model problems and perform goal oriented error estimation and grid refinement for either $I^\tau(x,u)$ or $J(x,u)$ as QOI after termination of the nonlinear OCP solver. After refinement, we use the interpolated solution on the refined grid as starting guess and solve the nonlinear OCP again on the refined mesh. This procedure is repeated until the maximal number of time or space grid points is reached.

In all MPC simulations, we will perform four steps of Algorithm 1. In case of time adaptivity, we compute all trajectories on a space grid uniformly refined three times, cf. Table 5.1, and start the adaptive algorithm with three time points. The simulation is performed on a grid with 51 time grid points on $[0, \tau]$. In case of space adaptivity, we fix the number of total time grid points to 41, perform adaptive space adaptivity starting with a grid with one uniform refinement, and perform the simulation with the same time step size. The space grids for the simulation are five times uniformly refined. In case of space-time adaptivity, we perform the simulation with 51 time grid points on $[0, \tau]$, where every space grid is five times uniformly refined and we start with five time grid points and one uniform refinement.

5.3.1 Linear quadratic optimal control problems

We will first consider linear quadratic problems and dynamics governed by a linear heat equation with distributed control, i.e.,

$$\begin{aligned} \dot{x} &= 0.1\Delta x + sx + u && \text{in } (0,T)\times\Omega, \\ x(0) &= 0 && \text{in } \Omega, \\ x &= 0 && \text{in } \Omega\times(0,T), \end{aligned} \tag{5.31}$$

where $s \in \mathbb{R}$ is a stability (if $s < 0$) or instability (if $s > 0$) parameter. Alternatively, we consider Neumann boundary control, i.e.,

$$\begin{aligned} \dot{x} &= 0.1\Delta x + sx && \text{in } (0,T)\times\Omega, \\ x(0) &= 0 && \text{in } \Omega, \\ 0.1\frac{\partial x}{\partial \nu} &= u && \text{in } \Omega\times(0,T), \end{aligned} \tag{5.32}$$

where $\frac{\partial}{\partial \nu}$ denotes the outward unit normal derivative.

We set $U = L_2(\Omega)$ or $U = L_2(\partial\Omega)$ and aim to minimize the standard tracking-type cost functional (5.1) subject to either of the dynamics defined above.

We will observe that the stability of the underlying optimal solution plays a major role in the adaptive time discretization. In that context, we investigate the case of an unstable uncontrolled equation, i.e., choosing $s > 0$ large enough. In that case, the presented dG(0) scheme can be numerically unstable. Thus, for strongly unstable open-loop dynamics, a multiple shooting approach, cf. [27, 28, 80, 81], should be considered to prevent instabilities or blow-ups of the numerical solution.

Time adaptivity

In Figure 5.4, we depict the spatial norm of state and control over time for an autonomous problem with reference trajectory $x_{\mathrm{d}}^{\mathrm{stat}}$ and Tikhonov parameter $\alpha = 10^{-1}$ governed by dynamics described by (5.31) with instability parameter $s = 4$. We observe that the refinement with respect to the truncated QOI $I^\tau(x,u)$ only takes place at the beginning of the time interval. Further, we see that the error indicators decay exponentially shortly after the implementation horizon $\tau = 0.5$ due to the exponential decay of the secondary variables proven in Theorems 5.2 and 5.6. Second, choosing the entire cost functional as a QOI, we see that the refined time grid is fine towards $t = 0$ and $t = T$. This is because the dynamics exhibit a steady state turnpike behavior, i.e., the highly dynamic parts are located at the beginning and the end of the time horizon. Hence, in order to obtain an accurate solution on the whole horizon, these parts need to be refined. Further, we observe that the solution obtained by refinement via $I^\tau(x,u)$ does not exhibit the leaving arc despite very clearly showing the approaching arc.

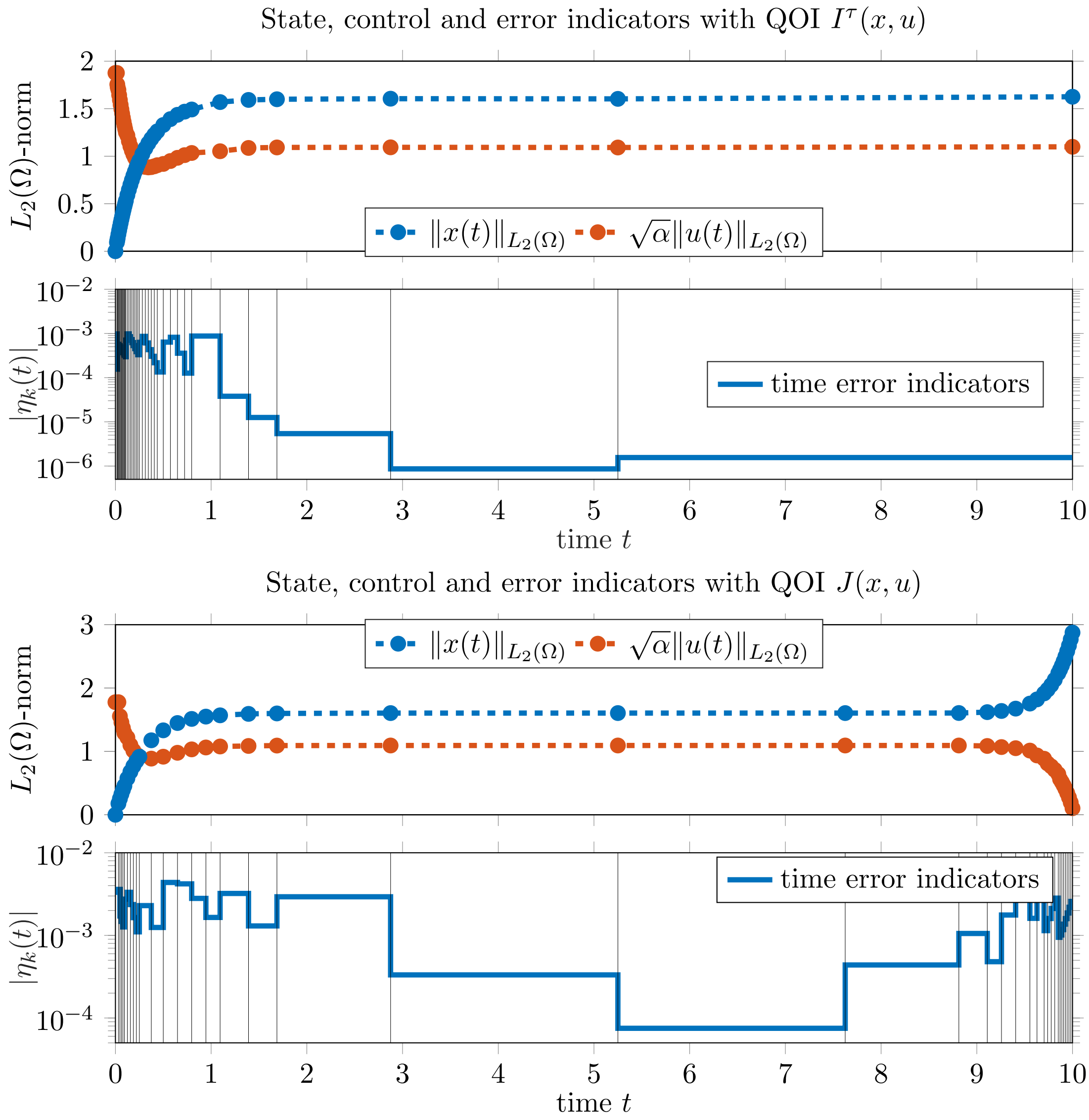


Figure 5.4: Open-loop trajectories and error indicators in the first MPC step after adaptive refinement with 41 time grid points for an unstable problem with distributed control and static reference. The vertical lines illustrate the adaptively refined time grid.

Further, in Figure 5.5, we depict the same quantities for a non-autonomous problem and choose the time-dependent reference trajectory $x_{\mathrm{d}}^{\mathrm{dyn}}$ and $\alpha = 10^{-3}$. The dynamics are governed by (5.32) with $s = 0$ and we set the implementation horizon to $\tau = 1$. Similarly to the autonomous problem we again observe that the refinement and error indicators are concentrated on the implementation horizon $[0, \tau]$ if $I^\tau(x, u)$ is chosen as QOI. If we use the cost functional as QOI, the time refinement and error indicators are distributed over the whole time horizon, as the reference and hence the solution is dynamic at all times.

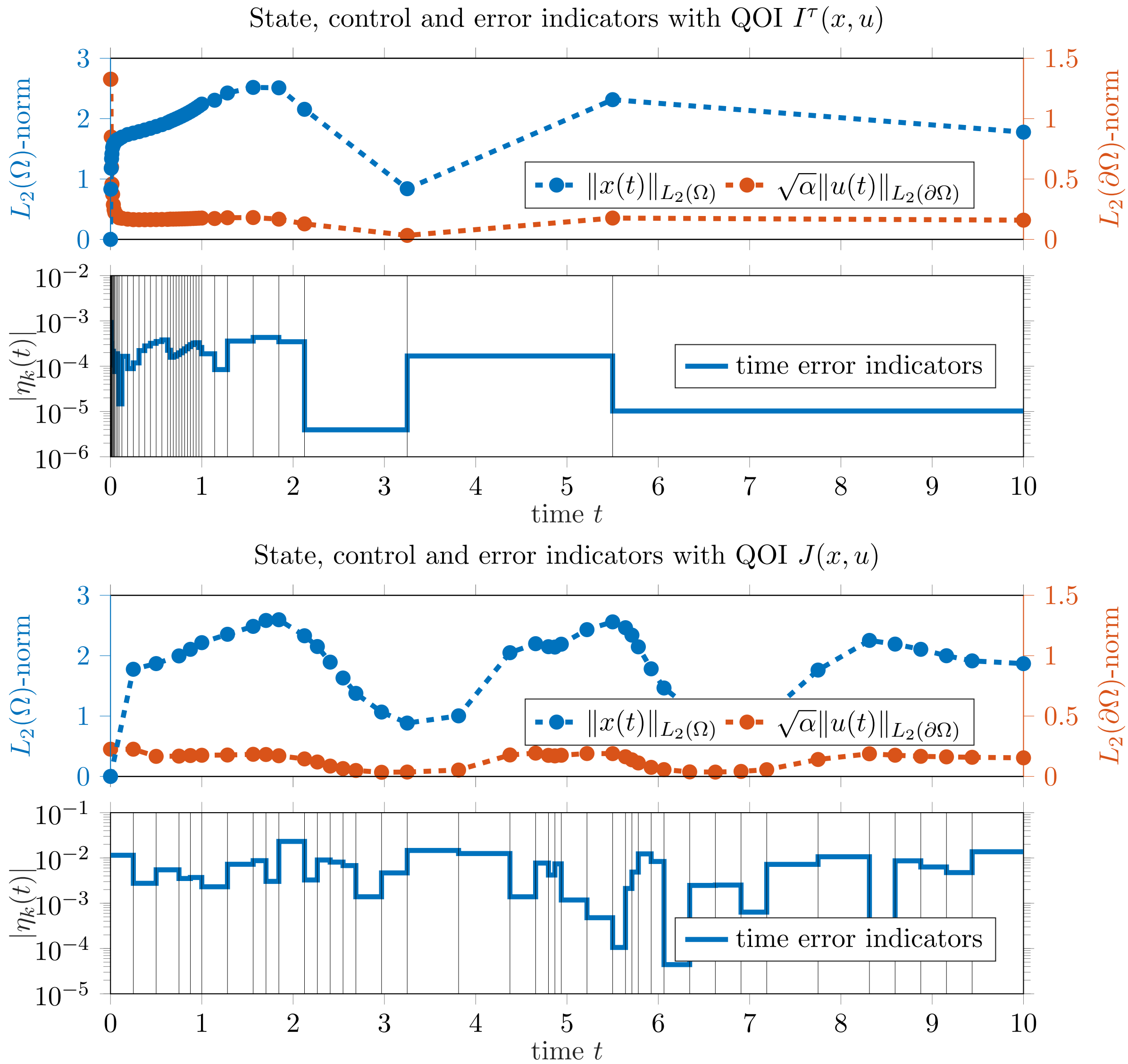


Figure 5.5: Open-loop trajectories and error indicators in the first MPC step after adaptive refinement with 41 time grid points for a boundary controlled stable non-autonomous problem with dynamic reference. The vertical lines illustrate the adaptively refined time grid.

Having investigated the error indicators and the resulting time refinement in the context of one optimal control problem, we depict the performance gain in a Model Predictive Controller with three examples when using the truncated QOI in Figure 5.6 for adaptivity in every MPC step. We show the closed-loop cost of the MPC trajectory obtained by applying four steps of the MPC algorithm Algorithm 1 to the optimal control problem. The plot on the top right and on the bottom correspond to the setting of Figure 5.4 and Figure 5.5, respectively, whereas the plot on the top left compares the closed-loop cost for a stable autonomous problem with

$s = 0$, $\alpha = 10^{-3}$, and reference x_d^{stat}. In all three cases we observe that for a given number of maximal time steps, choosing the specialized QOI $I^\tau(x, u)$ as an objective for refinement leads to a significant reduction of the closed-loop cost, i.e., a better controller performance.

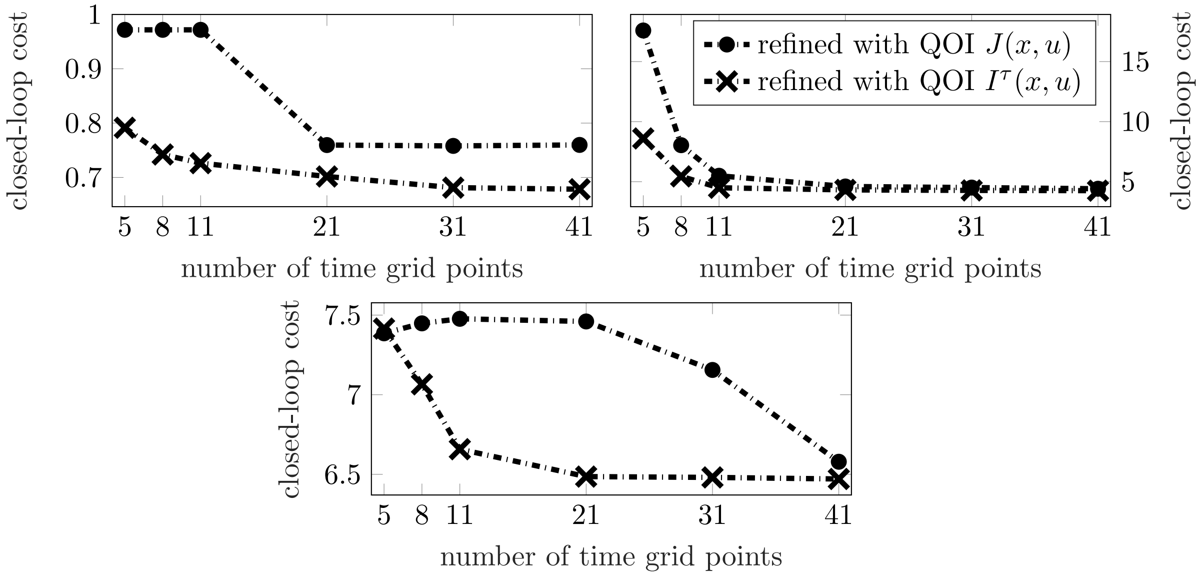


Figure 5.6: Comparison of cost functional values of the MPC closed-loop trajectory for different QOIs used for temporal refinement. Top left: Stable autonomous problem. Top right: Unstable autonomous problem. Bottom: Boundary controlled non-autonomous problem.

Space adaptivity

In this part, we investigate the case of space refinement. To this end, we compare the error indicators, the resulting space grids and the closed-loop cost for refinement with $I^\tau(x, u)$ and $J(x, u)$. In the upper row of Figure 5.7, the space error indicators for an autonomous optimal control problem governed by the linear dynamics with distributed control defined in (5.31) with reference x_d^{stat} and Tikhonov parameter $\alpha = 10^{-3}$ are depicted. Again, the error indicators for the objective $I^\tau(x, u)$ decay exponentially after the implementation horizon, whereas they stay almost constant over the whole time horizon in case of the QOI $J(x, u)$. This again is due to the turnpike property, i.e., the dynamic trajectories are close to the solution of the steady state problem for the majority of the time. The higher indicators at the beginning of the time interval are due to the high control action to approach the turnpike. Further, the indicators for the cost functional decay at the end of the horizon due to the terminal condition of the adjoint, which requires the control to approach zero, leading to a more regular state by diffusion and thus less need to refine. In the lower row of Figure 5.7, the resulting space grids for three different numbers of maximal spatial degrees of freedom (DOFs) are depicted. It is clearly visible that for $I^\tau(x, u)$ a refinement only takes place at the beginning of the time horizon and the majority of the space grids are unrefined. In contrast to that, the spatial refinement for $J(x, u)$ takes place on the whole horizon.

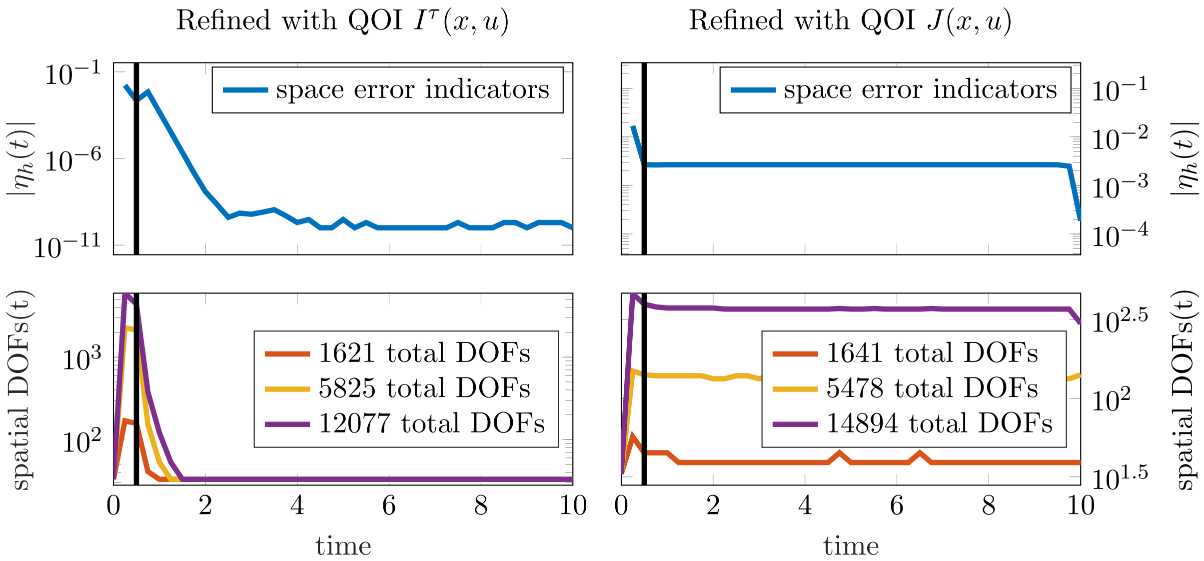


Figure 5.7: Spatial error indicators before refinement and spatial degrees of freedom after last refinement for different maximal numbers of degrees of freedom for an autonomous optimal control problem. The vertical black line indicates the implementation horizon $\tau = 0.5$.

In Figure 5.8, we depict the resulting space grids and the state over time for the intermediate case in Figure 5.7, i.e., the grids enjoy 5825 and 5478 total spatial DOFs, respectively. It is clearly visible that in case of refinement for the full cost functional, the space grids have to capture the steady state turnpike on the majority of the interval. This is not the case for refinement with $I^\tau(x,u)$, where we observe unrefined space grids shortly after the implementation horizon $\tau = 0.5$.

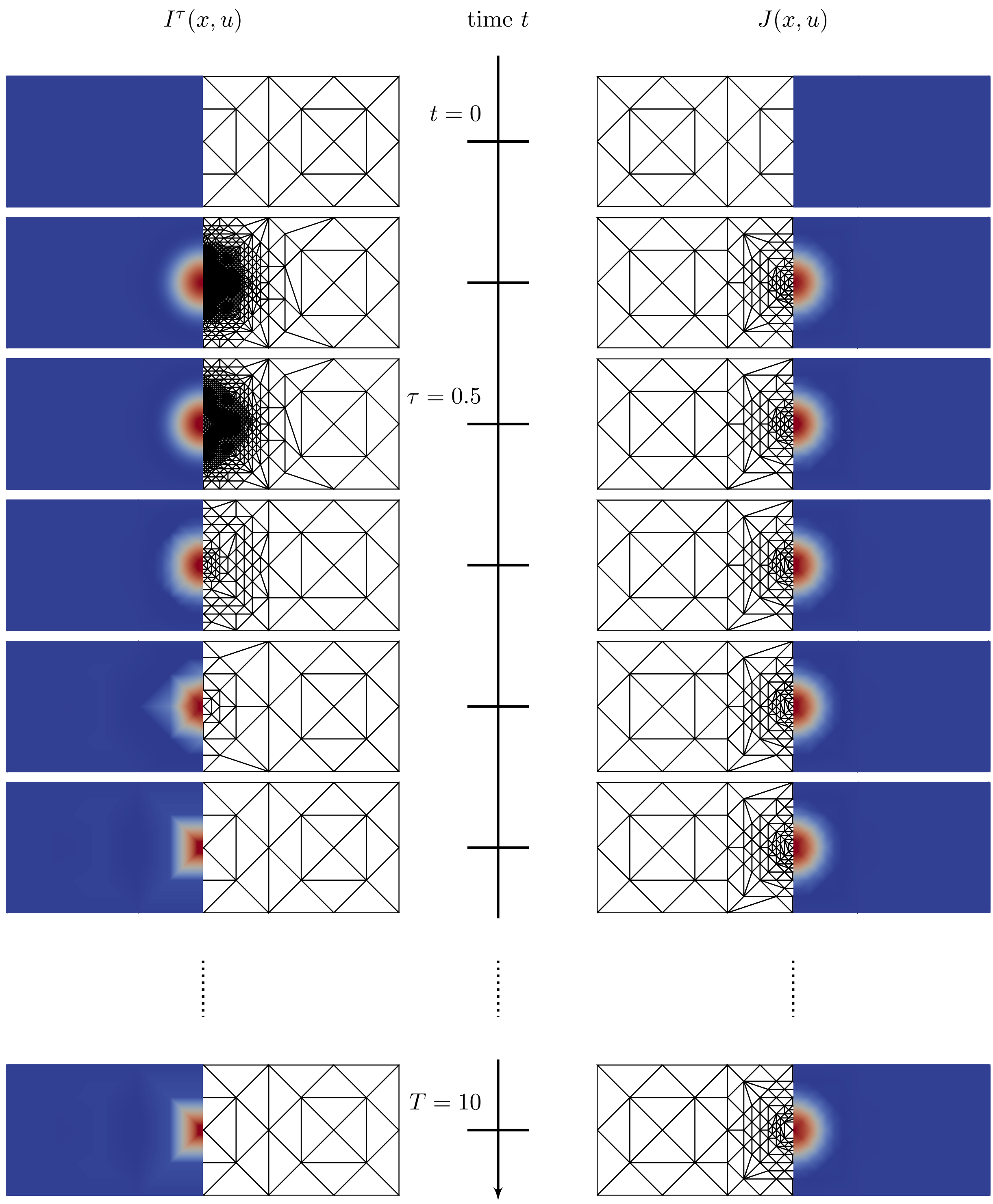


Figure 5.8: Evolution of adaptively refined space grids for $I^\tau(x,u)$ (left) and $J(x,u)$ (right) with 5825 and 5478 total spatial DOFs, respectively.

As a second example for space adaptivity in the linear quadratic setting, we present the case of reference that increases exponentially in time, i.e., $x_{\mathrm{d}}^{\mathrm{exp}}$ as defined in (5.29) with $\omega_{2,\mathrm{peak}} = 0.5$. We choose the Tikhonov parameter $\alpha = 10^{-3}$, $s = 0$, and consider the distributed dynamics (5.31). In Figure 5.9 we show the error indicators before refinement and the spatial DOFs after refinement for different numbers of maximal spatial DOFs. As the reference $x_{\mathrm{d}}^{\mathrm{exp}}$ increases exponentially in time, the solution also increases exponentially in time. This leads to the error indicators for $J(x, u)$ also increasing in time (top right) and in particular to a refinement which is concentrated towards T (bottom right). On the other hand, when refining for $I^\tau(x, u)$ the exponential damping of discretization errors is stronger than the exponential increase of the solution, as the indicators for this truncated cost functional decay exponentially (top left). This, again, similarly to the autonomous case, leads to space grids that are refined on $[0, \tau]$ (bottom left).

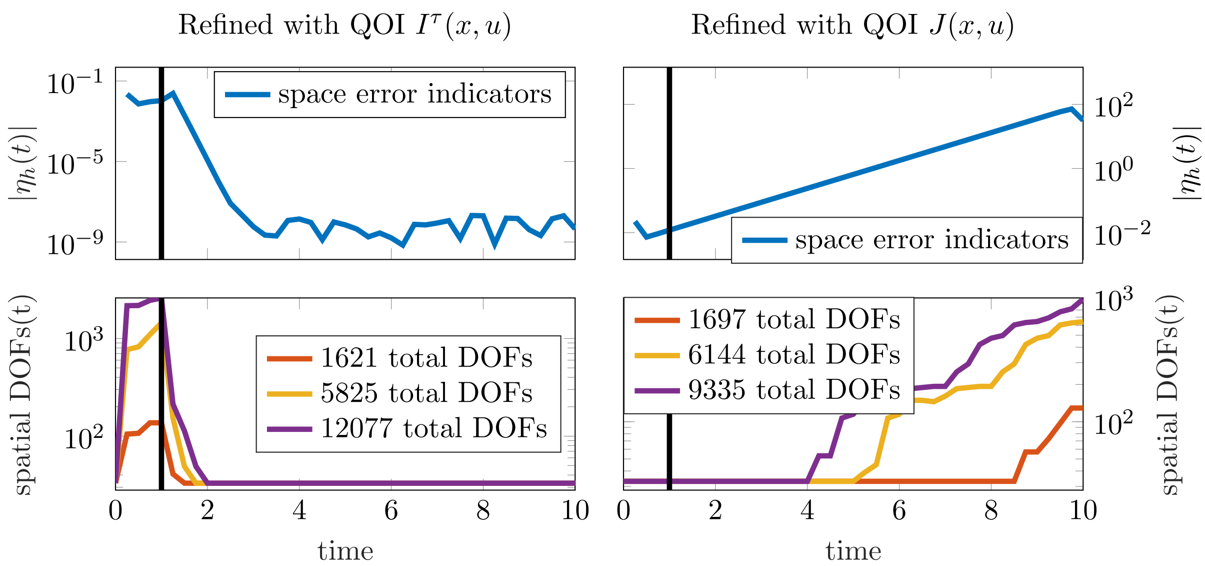


Figure 5.9: Spatial error indicators before refinement and spatial degrees of freedom after last refinement for different maximal numbers of degrees of freedom for a non-autonomous optimal control problem. The vertical black line indicates $\tau = 1$.

Finally we examine the performance gain from using $I^\tau(x, u)$ as a QOI in adaptive MPC again. In Figure 5.10, we observe that for both examples, i.e., the autonomous problem of Figure 5.7 and the non-autonomous problem of Figure 5.9 with exponentially increasing reference, the closed-loop cost is lower when using the specialized QOI $I^\tau(x, u)$ for refinement. In case of the exponentially increasing reference, we further see that increasing the allowance for space refinement does not improve the performance when refining with $J(x, u)$; this is due to the fact that all grid point are used towards T and thus the MPC feedback is not refined at all, cf. the bottom left of Figure 5.9.

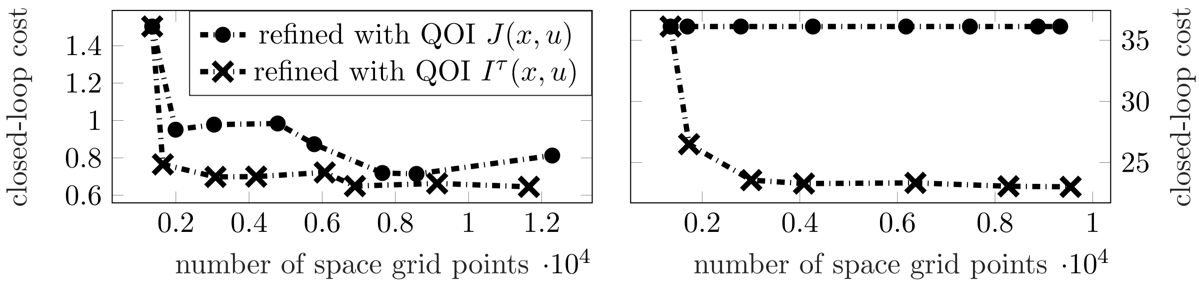


Figure 5.10: Comparison of cost functional values of the MPC closed-loop trajectory for different QOIs used for spatial refinement. Left: Autonomous problem. Right: Non-autonomous problem with exponentially increasing reference.

Space-time adaptivity

We briefly address the subject of space and time adaptivity for the linear dynamics (5.31) with static reference x_d^{stat}, $\alpha = 10^{-3}$, and $s = 0$. After time and space error estimation, we refine either space or time, depending on which is subject to a larger total error. This was chosen due to clarity and simplicity and we note that there are more involved space-time refinement strategies, cf. [100, Section 6.5]. As to be expected, the space and time grid refinement for $I^\tau(x,u)$ primarily takes place on the initial part of the horizon, cf. Figure 5.11.

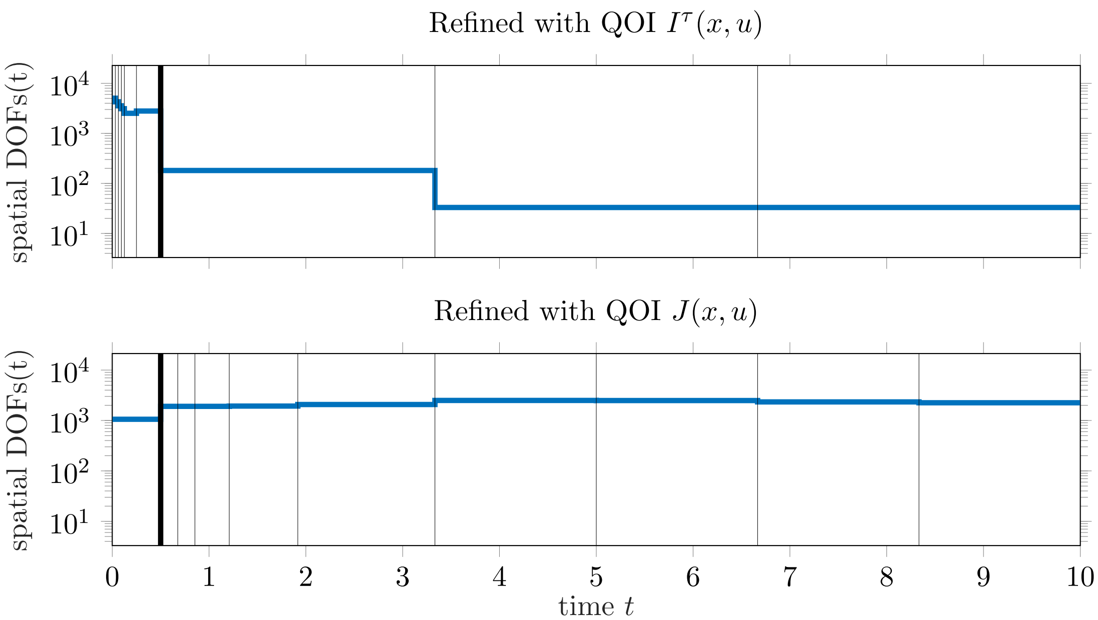


Figure 5.11: Spatial DOFs over time for a total allowance for 20000 degrees of freedom for a fully adaptive space-time refinement.

The adaptive refinement with $I^\tau(x,u)$ terminated with 12 time grid points, whereas the refinement with $J(x,u)$ terminated with 11 time grid points. In Figure 5.12, we clearly observe that again refinement with the truncated cost functional leads to a better performance of the Model Predictive Controller.

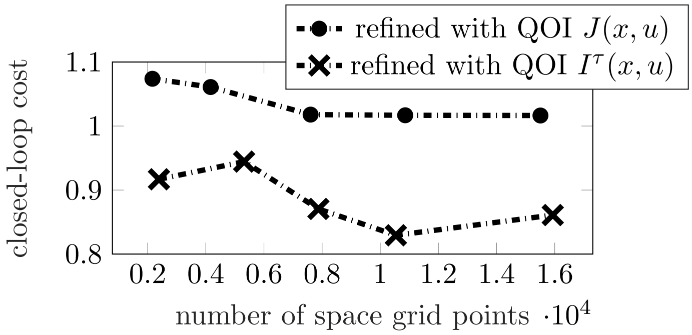


Figure 5.12: Comparison of cost functional values of the MPC closed-loop trajectory for different QOIs used for space-time refinement.

We note that, for this example, employing only time adaptivity for the QOI $I^\tau(x,u)$ with three uniform refinements in space, cf. Table 5.1, leads to a lower closed-loop cost for the same number of total DOFs. This is no longer the case when using two uniform refinements.

5.3.2 Semilinear optimal control problems

In this part, we move from linear dynamics to the semilinear heat equation introduced in Section 4.5.1, i.e.,

$$\begin{aligned} x' - d\Delta x + ex^3 &= u && \text{in } \Omega \times (0,T), \\ x &= 0 && \text{in } \partial\Omega \times (0,T), \\ x(0) &= 0 && \text{in } \Omega, \end{aligned} \tag{5.33}$$

where $d > 0$ is a diffusivity parameter and $e > 0$ is a nonlinearity parameter. We note that the semilinearity has a stabilizing effect due to monotonicity, cf. Section 5.3.2.

Time adaptivity

In Figure 5.13, we depict the error indicators and the resulting time grids for an optimal control problem with cost functional (5.30), dynamic reference $x_\mathrm{d}^\mathrm{dyn}$ defined in (5.28), Tikhonov parameter $\alpha = 10^{-2}$ and the semilinear dynamics described in (5.33) with semilinearity parameter $e = 1$. We choose $\tau = 1$ as implementation horizon. We observe that, similarly to the non-autonomous linear quadratic case depicted in Figure 5.5, the error indicators are concentrated on the beginning of the interval if the objective for refinement is $I^\tau(x,u)$, leading to a fine time grid on $[0,\tau]$. The refinement with the cost functional as QOI shows a time refinement distributed over the whole interval. In particular, the initial part up to the MPC implementation horizon was not refined at all when refining for $J(x, u)$. This different refinement behavior leads to the optimal state and control being fundamentally different.

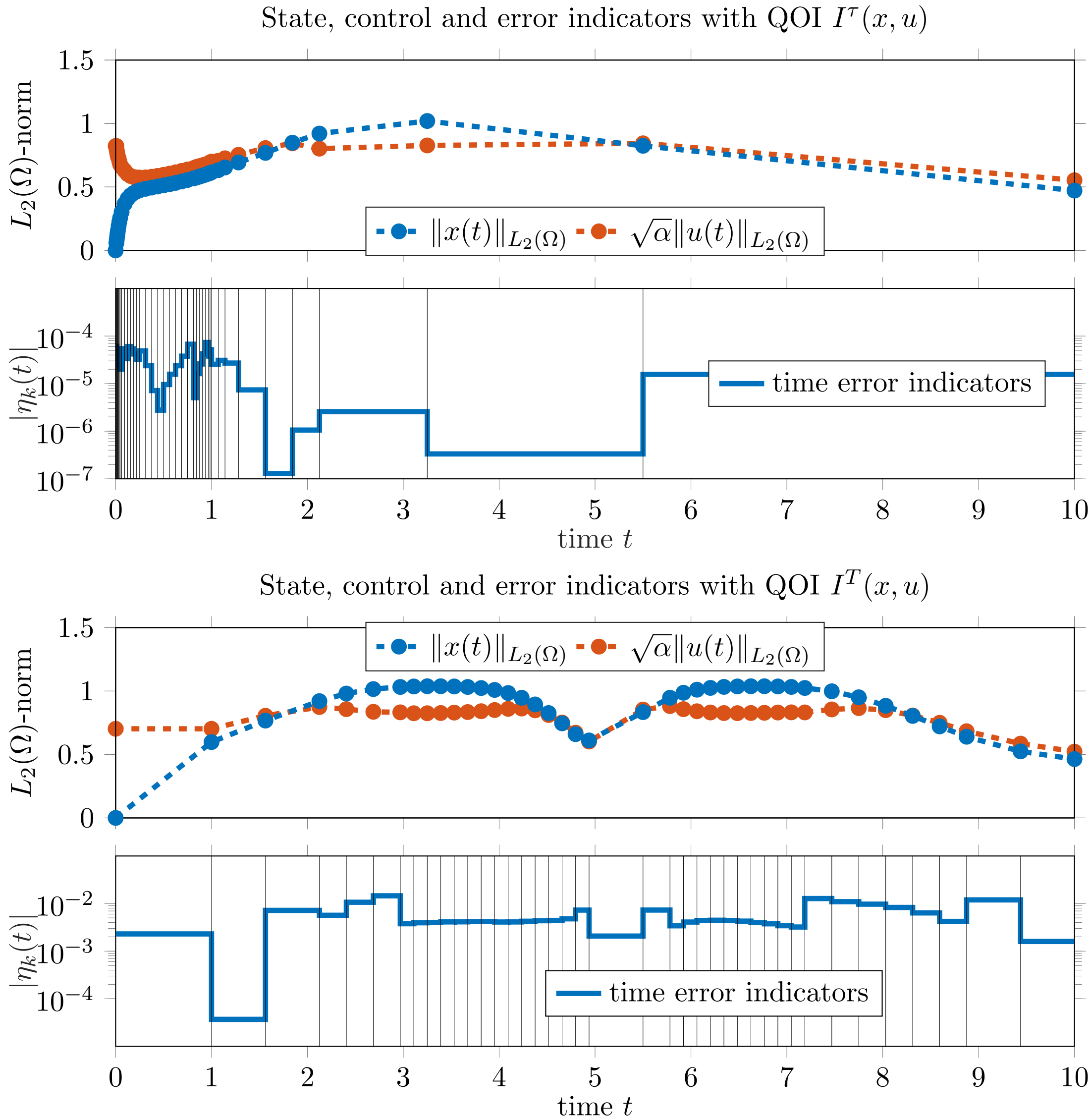


Figure 5.13: Open-loop trajectories and error indicators in the first MPC step after adaptive refinement with 41 time grid points for a semilinear problem. The vertical lines illustrate the adaptively refined time grid.

Correspondingly, in Figure 5.14, we depict the closed-loop cost of the adaptive Model Predictive Controller. We observe that due to the coarse grid on $[0, \tau]$ for the refinement with $J(x, u)$, the performance for increasing number of time grid points remains constant. On the other hand, the performance of the Model Predictive Controller using $I^\tau(x,u)$ as refinement objective is increased when increasing the time grid point allowance.

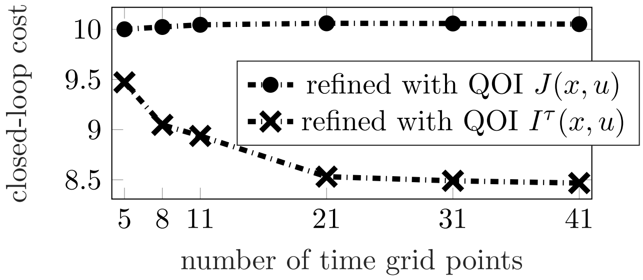


Figure 5.14: Comparison of cost functional values of the MPC closed-loop trajectory for different QOIs used for temporal refinement with semilinear dynamics.

Space adaptivity

We observed that the behavior of the adaptive time refinement algorithm in case of semilinear dynamics is very similar to the case of linear dynamics in the previous part. This is also the case when moving to space adaptivity. In the previous examples with space adaptivity we introduced a necessity for spatial refinement via the reference trajectory. In contrast to that, in this part, we will change the computational domain and consider a reference constant in time and space. We endow the rectangle with a reentrant corner as depicted in Figure 5.15. We note that from a theoretical point of view, regularity theory is often formulated for convex domains, which is not the case here. In the case of a nonconvex domain, as depicted in Figure 5.15, one usually obtains regularity results coupled with the angle at the reentrant corner, cf. [60, Section 8.4].

Figure 5.15: Rectangular domain $[0,3] \times [0,1]$ with a reentrant corner at $(1.5, 0.5)$ with corresponding angle of approximately 2.4 degrees.

In this part, we consider a reference constant in time and space, i.e., $x_d \equiv 1$ and Tikhonov parameter $\alpha = 10^{-3}$ in the cost functional (5.30) and choose $e = 0.1$ as nonlinearity parameter. In Figure 5.16, the error indicators and the corresponding space grids are depicted. Again, we see that the error indicators for the objective $I^\tau(x,u)$ decay exponentially in time, whereas the indicators for $J(x,u)$ stay constant. Correspondingly, the adaptive space refinement for $I^\tau(x,u)$

is concentrated on the implementation horizon $[0, 0.5]$, whereas the refinement for $J(x, u)$ is distributed evenly over the whole horizon $[0, 10]$. The latter is, yet again, due to the turnpike property which can also clearly be observed in Figure 5.17 as the problem at hand is autonomous.

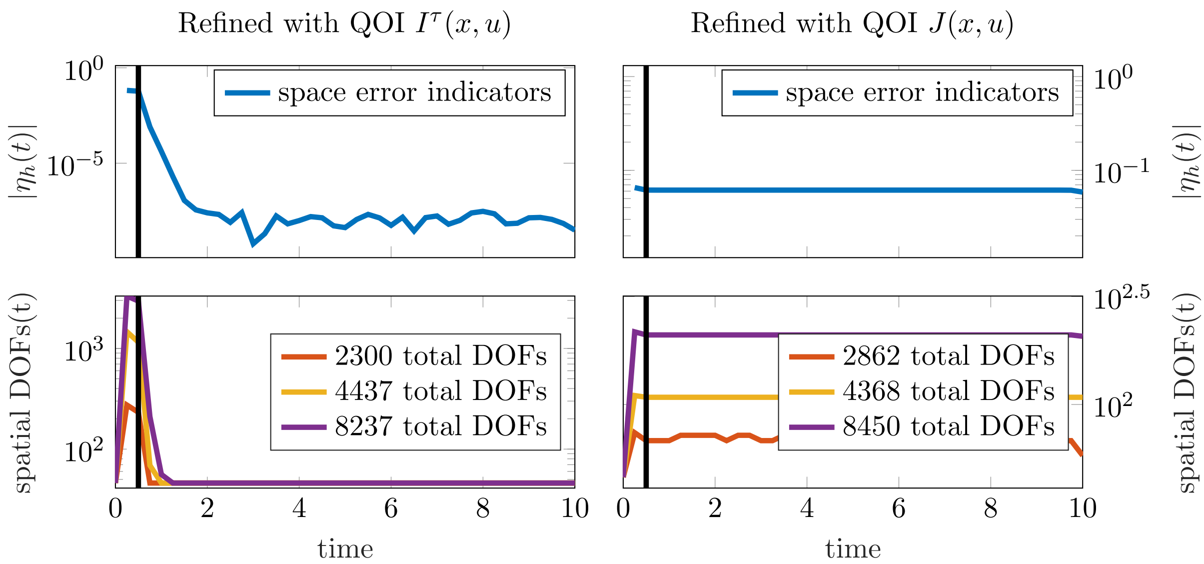


Figure 5.16: Spatial error indicators before refinement and spatial degrees of freedom after last refinement for different maximal numbers of degrees of freedom for an autonomous semilinear optimal control problem. The vertical black line indicates the implementation horizon $\tau = 0.5$.

We depict the space grids over time and the corresponding state in Figure 5.17. Due to the homogeneous Dirichlet boundary conditions, the refinement primarily occurs close to the boundary. In case of refinement for $J(x, u)$ and due to the turnpike property, the space grids are almost identical over time, whereas for the specialized QOI $I^\tau(x, u)$, the refinement happens primarily on the space grids assigned to the implementation horizon $[0, 0.5]$.

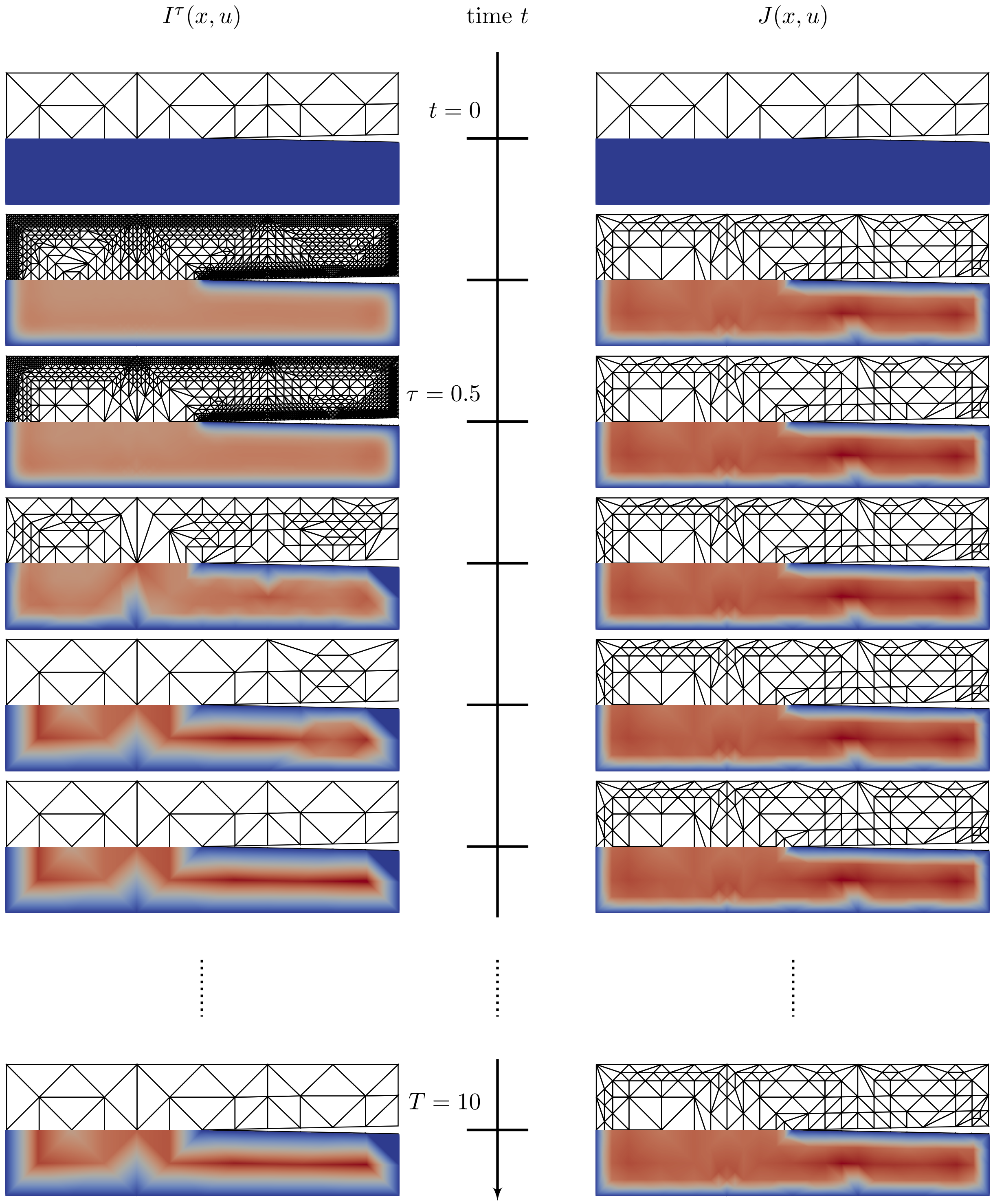


Figure 5.17: Evolution of adaptively refined space grids for $I^\tau(x,u)$ (left) and $J(x,u)$ (right) with 8237 and 8450 total spatial DOFs, respectively.

Finally, in Figure 5.18, we depict the closed-loop cost of the MPC algorithm endowed with goal oriented space adaptivity. Again, the performance is better when using the specialized QOI $I^\tau(x, u)$ for refinement as opposed to adaptivity with respect to the full cost functional.

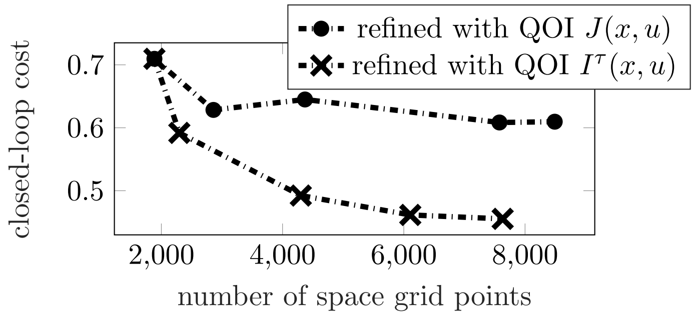


Figure 5.18: Comparison of cost functional values of the MPC closed-loop trajectory for different QOIs used for spatial refinement with semilinear dynamics.

5.3.3 Quasilinear optimal control problems

As a last model problem, we consider optimal control with the boundary controlled quasilinear problem introduced in Section 4.5.2. To this end, we recall the nonlinear heat conduction tensor

$$\kappa(x)(t, \omega) := \big(c|x(t, \omega)|^2 + d\big),$$

where $c, d > 0$ and consider the quasilinear dynamics

$$\begin{aligned} x' - \nabla \cdot (\kappa(x)\nabla x) &= 0 && \text{in } \Omega \times (0, T), \\ \kappa(x)\frac{\partial x}{\partial \nu} &= u && \text{in } \partial\Omega \times (0, T), \\ x(0) &= 0 && \text{in } \Omega. \end{aligned}$$

We note that the theory of Chapter 4 does not cover this boundary controlled quasilinear case. However, the results for exponential decay of the secondary variables, i.e., Theorem 5.2 and Theorem 5.6, still apply. This is due to the fact that the secondary variables are defined as a solution to a linear problem, even if the problem is nonlinear, cf. (5.11).

In Figure 5.19, we depict the time error indicators and corresponding state and control norm over time for a non-autonomous problem with $x_{\mathrm{d}}^{\mathrm{dyn}}$ as reference, $\alpha = 10^{-2}$, $c = d = 0.1$, and implementation horizon $\tau = 1$. The time refinement with the full cost functional as QOI is again, similarly to the case of linear dynamics depicted in Figure 5.5 and the case of semilinear dynamics shown in Figure 5.13 distributed on the whole horizon and the time grid on the implementation horizon $[0, \tau]$ remains unrefined. The refinement for the truncated cost functional $I^\tau(x, u)$ is concentrated on the initial part. Note that the depicted norm is again a scaled $H^1(\Omega)$-norm corresponding to the second derivative of the Lagrange function, i.e., $\|v\|_{\alpha d, H^1(\Omega)} := \|v\|_{L_2(\Omega)} + \sqrt{\alpha d}\|\nabla v\|_{L_2(\Omega)}$.

Time adaptivity

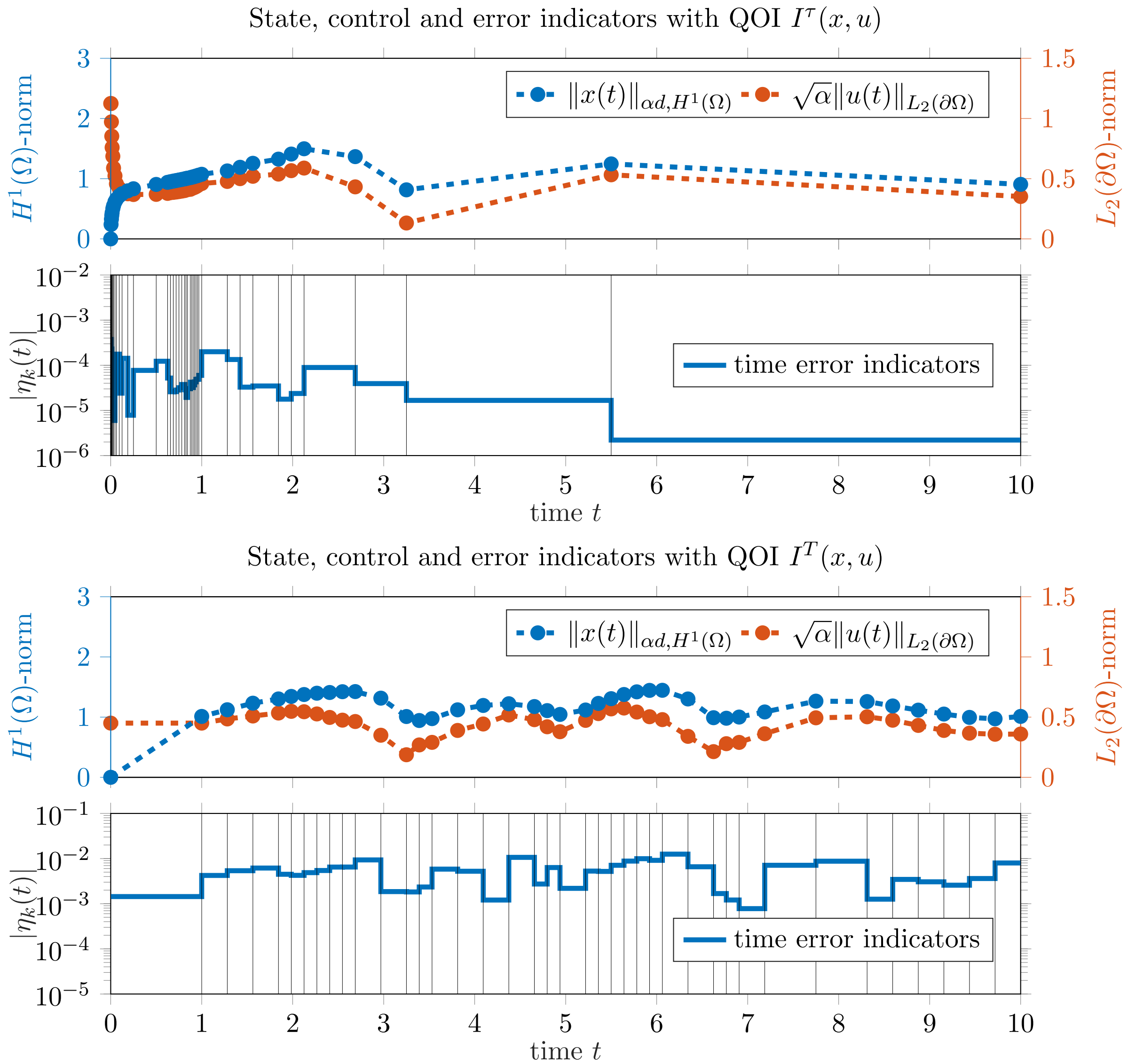


Figure 5.19: Open-loop trajectories and error indicators in the first MPC step after adaptive refinement with 41 time grid points for an autonomous problem with boundary controlled quasi-linear dynamics. The vertical lines illustrate the adaptively refined time grid.

The depiction of the closed-loop cost of the MPC trajectory in Figure 5.20 shows that the cost is again consistently lower when using $I^\tau(x,u)$ as a QOI for adaptive time refinement.

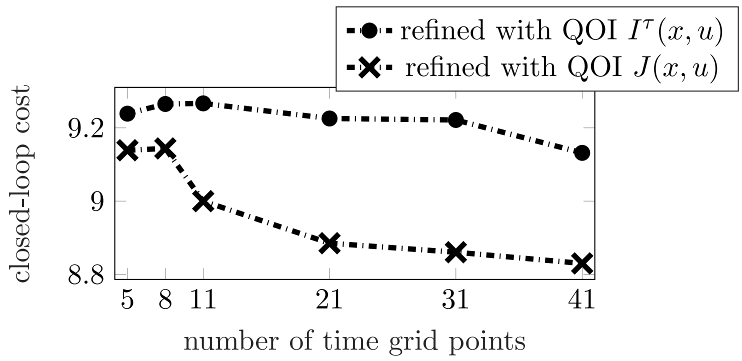


Figure 5.20: Comparison of cost functional values of the MPC closed-loop trajectory for different QOIs used for temporal refinement with quasilinear dynamics.

Space adaptivity

We consider the exponentially increasing reference trajectory $x_{\mathrm{d}}^{\mathrm{exp}}$ with $\omega_{2,\mathrm{peak}} = 1$, implementation horizon $\tau = 1$, Tikhonov parameter $\alpha = 10^{-3}$ and parameters $d = 10^{-1}$ and $c = 10^{-2}$ for the heat conduction tensor. In Figure 5.21, we see that despite the exponentially increasing trajectory, the error indicators for $I^\tau(x,u)$ still decrease exponentially over time. This, yet again shows the damping mechanism with respect to discretization errors. However, as opposed to the example of distributed control of a linear problem in Figure 5.9, the space grids are refined also outside of the implementation horizon. This is due to the fact that the damping mechanism is weaker due to the (less powerful) boundary control. The error indicators and the corresponding spatial DOFs after refinement for the full cost functional $J(x,u)$ are again exponentially increasing.

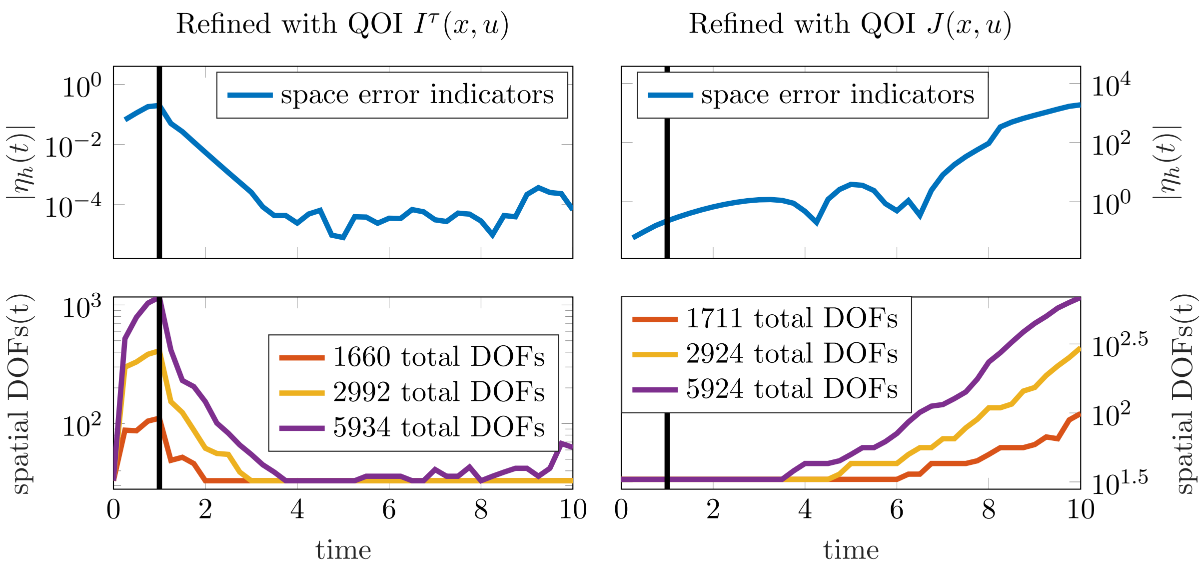


Figure 5.21: Spatial error indicators before refinement and spatial degrees of freedom after last refinement for different maximal numbers of degrees of freedom for a boundary controlled quasilinear problem. The vertical black line indicates the implementation horizon $\tau = 1$.

The state over time and the corresponding space grids are shown in Figure 5.22. Although state and control are relatively small on the initial part, the spatial refinement is most active there when refining for $I^\tau(x,u)$. On the other hand, the spatial grids refined for the full cost functional show no refinement on the whole implementation horizon $[0,1]$, as they are primarily refined towards the end of the horizon, due to the exponentially increasing reference trajectory.

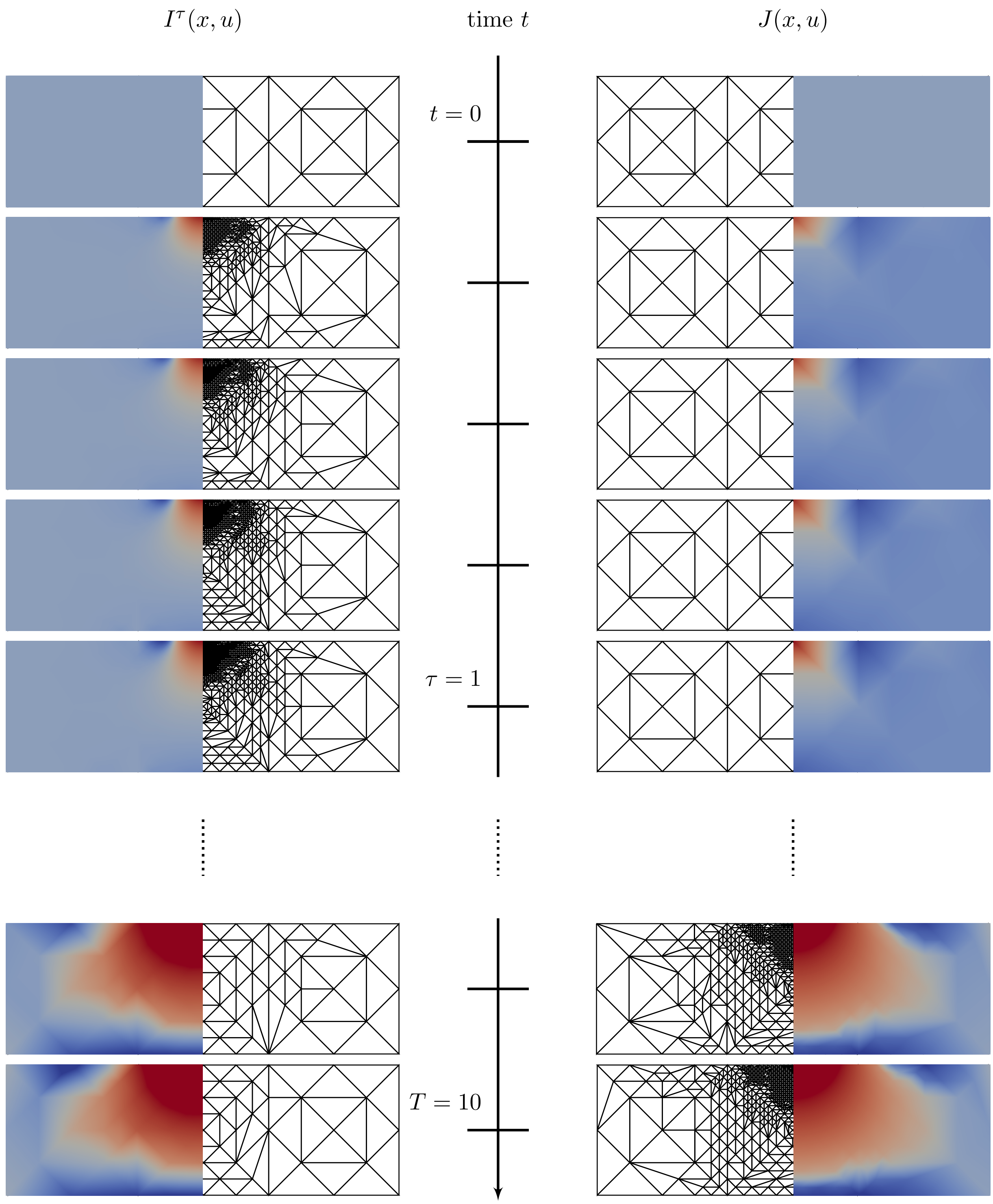


Figure 5.22: Evolution of adaptively refined space grids for the boundary control of a quasilinear equation refined for QOI $I^\tau(x,u)$ (left) and $J(x,u)$ (right) with 5934 and 5924 total spatial DOFs, respectively.

Finally in Figure 5.23 we show the corresponding closed-loop cost of the MPC trajectory. Similar to the linear quadratic example with exponentially increasing reference in Figure 5.10, an increasing number of space grid points does not increase the MPC performance when refining with $J(x, u)$. This is, again, because the error indicators and thus also the refinements are predominant towards T and not on the MPC implementation horizon. Thus, a refinement with the QOI $I^\tau(x, u)$ yields a significantly better controller performance, as can be observed in Figure 5.23.

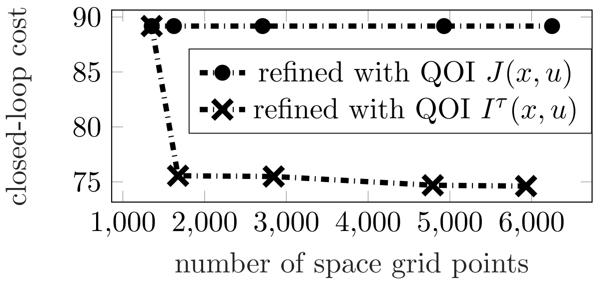


Figure 5.23: Comparison of cost functional values of the MPC closed-loop trajectory for different QOIs used for spatial refinement with quasilinear problem.

5.3.4 Implementation details and particularities for nonlinear problems

We briefly specify implementation details corresponding to the solution of the optimal control problems, the discretization and the error estimation. Further, we discuss various aspects that can be utilized for fast and efficient adaptive MPC methods. All algorithms were implemented in the C++-library for vector space algorithms *Spacy*[1] using the finite element library *Kaskade7* [57] for spatial finite elements, assembly and spatial grid management. For the blockwise factorization in the forward and backward solver of the PDE, we applied the sparse direct solver *UMFPACK* [37].

Evaluation of the error indicators

We first present the numerical realization to evaluate the time and space error indicators defined in (5.14) and (5.15). For the approximation of the weights needed for time error estimation in (5.14), we follow the approach described in [101, Section 5.1] and approximate , e.g., $w_x^k = x - \underline{x}_k$ by $(I_k^{(1)} - I)x_k$, where $I_k^{(1)}$ is the interpolation operator from discontinuous piecewise constant functions to continuous piecewise linear functions. The major advantage of this approach is that almost no additional numerical effort is needed for this approximation. If one computes the integrals weighted by $I_k^{(1)} x_k$ by the trapezoidal rule and the integrals weighted by x_k by the box rule, the error estimator (5.14) can be computed without even computing the interpolant, cf. [100, Section 6.4]. A second aspect of the evaluation of the time error estimator (5.14) is the

[1] https://spacy-dev.github.io/Spacy/

computation of differences of functions of different finite element meshes, i.e., e.g., x_{m-1} and x_m being defined on the grids $\mathcal{K}_h^{m-1}$ and $\mathcal{K}_h^m$, respectively. Directly computing the difference would be computationally infeasible, as one had to loop over vertices of $\mathcal{K}_h^{m-1}$ and evaluate x_m at these points, which includes searching $\mathcal{K}_h^m$ for the triangle containing the evaluation point. However, the common refinement illustrated in Figure 5.2 can be used to compute this difference efficiently: For any cell $K^{m-1/2}$ of the common refinement $\mathcal{K}_h^{m-1/2}$, the *dune grid-glue* module provides direct access to the parent cells $K^{m-1} \in \mathcal{K}_h^{m-1}$ and $K^m \in \mathcal{K}_h^m$. Thus, we can iterate over the cells of the common refinement and efficiently evaluate x_{m-1} and x_m locally on the parent cell K^{m-1} resp. K^m to compute the difference.

For approximating the weights for space error estimation via formula (5.15), one could pursue a similar approach as for time error estimation, i.e., via interpolation. However, in order to define an interpolant it is often assumed that the grid exhibits a patch structure to obtain a higher order solution by, e.g., biquadratic interpolation in case of quadrilaterals, cf. [102, Section 3.2 and Section 3.4]. As we use triangles instead of quadrilaterals and an unstructured mesh, we pursue a different approach and approximate the weights in formula (5.15), i.e., e.g., $w_x^h = x_k - \underline{x}_{kh}$, by a higher order method.

Assume we have a local minimizer $\xi_{kh} = (x_{kh}, u_{kh}, \lambda_{kh}) \in \mathcal{W}_{kh} \times \mathcal{U}_{kh} \times \mathcal{W}_{kh}$ of the nonlinear OCP at hand. Then we have

$$\langle (L^k)'(\xi_{kh}), \varphi \rangle_{(\mathcal{W}_{kh} \times \mathcal{U}_{kh} \times \mathcal{W}_{kh})^* \times \mathcal{W}_{kh} \times \mathcal{U}_{kh} \times \mathcal{W}_{kh}} = 0$$

for all $\varphi \in \mathcal{W}_{kh} \times \mathcal{U}_{kh} \times \mathcal{W}_{kh}$. In order to obtain an approximation of the continuous solution we consider an extension of the finite element space by *bubble functions*, which are bilinear combinations of the standard linear finite elements on the reference triangle, cf. [144, p.62 and Section 5.3]. We denote the resulting finite element space at time step m by $(V_h^m)^{\mathrm{e}}$ resp. $(U_h^m)^{\mathrm{e}}$ where $(V_h^m)^{\mathrm{e}} \cap V_h^m = \{0\}$ and $(U_h^m)^{\mathrm{e}} \cap U_h^m = \{0\}$. Using these ansatz spaces in every time step, we obtain the fully discrete spaces $\mathcal{W}_{kh}^{\mathrm{e}}$ and $\mathcal{U}_{kh}^{\mathrm{e}}$ with $\mathcal{W}_{kh}^{\mathrm{e}} \cap \mathcal{W}_{\mathrm{kh}} = \{0\}$ and $\mathcal{U}_{kh}^{\mathrm{e}} \cap \mathcal{U}_{\mathrm{kh}} = \{0\}$ analogously to (5.9). We evaluate the residual in this function space and perform a Newton step in the higher order space, i.e., we solve

$$\Big((L^k)''(\xi_{kh})\Big)_{\mathrm{ee}} \xi_{kh}^{\mathrm{e}} = -\Big((L^k)'(\xi_{kh})\Big)_{\mathrm{e}}, \tag{5.34}$$

where the subscripts *e* and *ee* denote evaluation in the extension space. We are not interested in a Newton update with a component in $\mathcal{W}_{kh} \times \mathcal{U}_{kh} \times \mathcal{W}_{kh}$, as this term would vanish in (5.15) due to Galerkin orthogonality. Thus, the solution of (5.34) purely on the extension space $\mathcal{W}_{kh}^{\mathrm{e}} \times \mathcal{W}_{kh}^{\mathrm{e}} \times \mathcal{W}_{kh}^{\mathrm{e}}$ is justified. We solve the system (5.4) with a conjugate gradient method and refer the reader to Section 5.4 for a discussion on further work regarding an efficient approximation of the system. Having solved (5.34), we approximate the weights via $\big(x_k - \underline{x}_{kh}, u_k - \underline{u}_{kh}, \lambda_k - \underline{\lambda}_{kh}\big) \approx \xi_{kh}^{\mathrm{e}}$. We apply the same strategy to the secondary variables as they can be characterized to be a critical point for a so called *exterior Lagrangian*, cf. [101, Section 4.2]. The performance of the resulting error estimator is shown in Table 5.2, where we can observe a very accurate estimation of the error in terms of the cost functional.

	1 uni. ref.	2 uni. ref.	3 uni. ref.	4 uni. ref.
$\lvert J(x_h, u_h) - J(x, u)\rvert$	1.30193	0.467676	0.121828	0.0298524
$\lvert\eta_h\rvert$	1.1127	0.441415	0.120613	0.0316621
$\frac{\lvert J(x_h,u_h)-J(x,u)\rvert}{\lvert\eta_h\rvert}$	1.17006	1.05949	1.01007	0.942843

Table 5.2: Performance of space error estimator.

Remark 5.7. *The approach described above is very closely related to hierarchical error estimation, cf. the seminal paper [39], and the respective parts in the monographs [40, Chapter 6] and [2, Chapter 5]. More recently, goal oriented error estimation for optimal control of elliptic PDEs using hierarchical error estimation techniques was discussed in [143, Section 2.4]. In that context, one usually aims to solve an error system of the form*

$$\begin{pmatrix} L''_{hh} & L''_{he} \\ L''_{eh} & L''_{ee} \end{pmatrix} \begin{pmatrix} e_h \\ e_e \end{pmatrix} = - \begin{pmatrix} 0 \\ L'_e \end{pmatrix},$$

A common approach for efficiently solving the system is to simplify the equation above by dropping a block of the operator, cf. [143, Section 2.2]. This yields, e.g.,

$$\begin{pmatrix} L''_{hh} & L''_{he} \\ 0 & L''_{ee} \end{pmatrix} \begin{pmatrix} e_h \\ e_e \end{pmatrix} = - \begin{pmatrix} 0 \\ L'_e \end{pmatrix}.$$

As in the context of goal oriented error estimation the influence of e_h vanishes in the evaluation of the error estimator (5.15) due to Galerkin orthogonality, one only has to solve the lower equation which corresponds to (5.34).

Localization strategies

We briefly recall localization strategies to localize the error estimator of (5.14) and (5.15) to a cell-wise level. All considerations are valid for space and time discretization. The most straightforward approach for localization is to use the cell contributions of the time and space integrals occurring in the definition of the error estimators as local cell indicators. However, in case of space error estimation, it was shown in [29] that this can lead to overestimation of the total error. Hence, more advanced strategies were developed in the literature. First, under a regularity assumption of the optimal triple one can apply integration by parts in space to formulate local error contributions, cf. [17, 18]. Second, assuming a patch structure on the underlying mesh, a filtering approach was introduced in [22, 102]. Last, and more recently, a strategy using a partition of unity was introduced in [121, Section 4.3], which leads to nodal error contributions. In this work, we use the first methodology due to its simplicity, i.e., we use the cell-wise contributions of the error estimator as local indicators. Further, we are not primarily interested in the total error but rather in the relative behavior of the error indicators over time. Accordingly, we mainly compare the refinement for two different QOIs for a given number of maximal grid points, where the relative size of error indicators with respect to other indicators is more important than the absolute error.

Refinement strategies

Assume we have cell-wise error indicators $\eta_i, i \in \mathcal{I}$ for the space or time error at hand, where $\mathcal{I}$ the index set of the time or space grid. We present two refinement strategies that can applied to either space or time adaptivity. Further, we assume that the error indicators are ordered by absolute value, i.e. w.l.o.g. η_i, $i \in \mathcal{I}$ are ordered in decreasing order. The strategy depicted in Algorithm 3 refines all cells with error indicators above a fixed fraction of the maximal error.

Algorithm 3 Refinement strategy for percentage of maximal error

Given: maximal number of cells and $0 < c < 1$
while #cells < maximal number of cells **do**
    Solve OCP
    Compute error indicators η_i for all $i \in \mathcal{I}$
    for $i \in \mathcal{I}$ **do**
        if $|\eta_i| \geq c|\eta_0|$ **then**
            Mark cell i for refinement
        end if
    end for
    Refine
end while

Second, we present a refinement strategy first discussed by Dörfler in [42, Section 4.2], which aims to reduce the error by a certain percentage.

Algorithm 4 Dörfler strategy

Given: maximal number of cells and $0 < c < 1$
while #cells < maximal number of cells **do**
    Solve OCP
    Compute error indicators η_i for all $i \in \mathcal{I}$
    $p = 0$
    for $i \in \mathcal{I}$ **do**
        Mark cell i for refinement
        $p = p + |\eta_i|$
        if $p \geq c \sum_i |\eta_i|$ **then**
            break
        end if
    end for
    Refine
end while

In all numerical experiments we used Algorithm 4 with $c = 0.5$ for time refinement and Algorithm 3 with $c = 0.3$ for space refinement. This choice is due to the fact that the refinement via Algorithm 3 is more aggressive if a lot of space grids need to be refined evenly, e.g., in case the variables are close to a turnpike for several time discretization points. With a Dörfler criterion, i.e., Algorithm 4, it can happen that the refinement procedure terminates after refining only some of the space grids, despite the error indicators being of the same size. For a more advanced refinement strategy, where the number of refined cells is optimized, the interested reader is referred to [100, Section 6.5] and the references therein.

OCP solver

For the numerical solution of the optimal control problem in every MPC step, we use a composite step method [98, 126] which is particularly well suited for strongly nonlinear optimal control problems. Roughly speaking, this method can be seen as a globalized Newton method for the first-order optimality system, splitting the total step into a tangential step for optimality and a normal step for feasibility. This allows us to also solve quasilinear problems with a very strong nonlinearity. The arising linear systems are solved with a projected preconditioned conjugate gradient method, where the dynamics are solved exactly by blockwise factorization in the preconditioner.

Solution warm starts

An important component for nonlinear problems is to fully utilize initial guesses for the nonlinear iteration, whenever they are available. In the context of grid adaptivity, one has to solve the nonlinear problem again after refinement. Interpolating the solution on the old grid onto the new grid serves as a good starting guess. We will illustrate this by the example of spatial refinement. We consider the cost functional (5.30) and the quasilinear equation of Section 4.5.2 and replace the Neumann boundary control by a distributed control. Further we set the nonlinearity parameter to $c = 10000$, the Tikhonov parameter to $\alpha = 10^{-2}$ and use the dynamic reference (5.28). We run the nonlinear algorithm on a coarse mesh, starting with 33 vertices, cf. Table 5.1, uniformly refine the space grid and use the interpolated solution on the fine grid as a starting guess for the next nonlinear iteration. Then the process is repeated. We compare the iteration numbers with choosing zero as initial guess in each nonlinear iteration in Table 5.3. We observe that significantly less iterations are needed when using the interpolated solution as starting guess.

refinement loop	0	1	2	3
warm start (it.)	20	10	7	6
no warm start (it.)	20	18	18	21

Table 5.3: Iterations of the nonlinear solver with and without solution warm start after grid refinement for a strongly quasilinear problem.

A second application of initial guesses is to use the shifted solution of the OCP in the previous MPC step as an initial guess for the nonlinear OCP solve in the current MPC step. We will illustrate this by means of an example with the control variable u for initialization of the second MPC iteration and describe how to obtain a starting guess for the optimal control defined on $[\tau, T+\tau]$ from the previously obtained optimal control defined on $[0,T]$. First, we restrict the optimal control of the previous MPC step defined on $[0,T]$ onto $[\tau,T]$. Second, we perform a constant extrapolation, i.e., $u(t) = u(T)$ for $t \in [T, T+\tau]$, to obtain a function defined on $[\tau, T+\tau]$. We will briefly evaluate this by methodology means of two problems with different reference trajectories, i.e., the static reference $x_{\mathrm{d}}^{\mathrm{stat}}$ defined in (5.27) and the dynamic reference $x_{\mathrm{d}}^{\mathrm{dyn}}$ defined in (5.28) and the quasilinear problem mentioned above. The first nonlinear solve for the first MPC step took 16 iterations for the autonomous problem and 20 iterations for the non-autonomous problem. We choose the time horizon $T = 10$ and compare the number of iterations needed for the second MPC step with and without solution warm start for different implementation horizons in Table 5.4. In the autonomous case depicted on the left, despite the length of the implementation horizon the nonlinear solver with initial guess needed only three iterations to converge. This is mostly due to the presence of a steady state turnpike property that is approached by the open-loop solution of the first MPC step. The second MPC step then only remains at the turnpike, leading to both open-loop solutions being very similar. Without initial guess, significantly more iterations were needed. This performance gain when using a good initial guess is dampened when considering larger implementation horizons τ for a non-autonomous problem, as the open-loop solutions of two succeeding MPC steps can be fundamentally different.

τ	1	3	5	10
warm start (it.)	3	3	3	3
no warm start (it.)	13	14	13	15

τ	1	3	5	10
warm start (it.)	6	15	13	12
no warm start (it.)	16	22	19	11

Table 5.4: Iterations of the nonlinear solver with or without solution warm start in the second MPC step for a strongly quasilinear problem. Left: autonomous problem. Right: non-autonomous problem

Grid warm starts

Another aspect is the grid refinements and their reuse in an MPC scheme. One can use the adaptive grids computed in MPC step i for starting grids in MPC step $i+1$. Due to the forward stepping of MPC, one does not need to incorporate a coarsening algorithm. If the time or space grid at time instance t_i with $t_i > t_0 + \tau$ is refined in one MPC step due to a significant influence on the MPC feedback, its influence will be even higher on the MPC feedback computed in MPC step $i+1$. Further, if one wants to utilize warm starts of the solution in the MPC scheme, grid warm starts further allow to use the computed solution without interpolation.

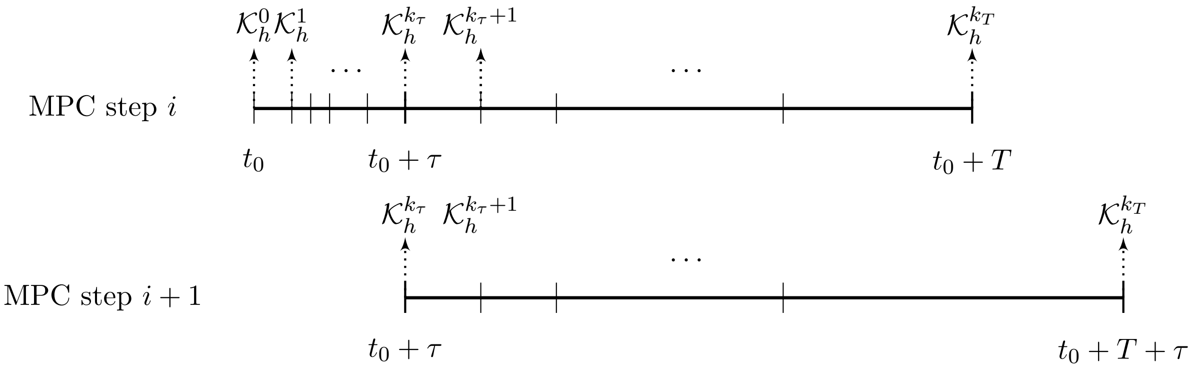


Figure 5.24: Possible reuse of time and space grid in an adaptive Model Predictive Controller.

Parallelization

We briefly compare the effect of parallelization for the two most time consuming tasks in the solution of the nonlinear OCP, i.e., the factorization of the block diagonal for each Euler step and the computation of the common refinements. We show the computation times for sequential and parallel execution of these two tasks in Table 5.5 and Table 5.6 and observe that a parallelization with four CPU cores led to a speedup of approximately two in both cases. For both test cases, we used eleven time step points.

total space DOFs	1243	4587	17611	69003	273163
time sequential (ms)	3	14	68	289	1472
time parallel (ms)	3	9	40	151	701
speedup factor	1	1.6	1.7	1.91	2.1

Table 5.5: Comparison of sequential and parallel computation time for block-diagonal factorization of the differential operator.

total space DOFs	1243	4587	17611	69003	273163
time sequential (ms)	292	1225	5093	20771	86806
time parallel (ms)	140	669	2394	10371	44528
speedup factor	2.1	1.8	2.1	2.0	1.9

Table 5.6: Comparison of sequential and parallel computation time for common refinement and transfer matrix assembly.

5.4 Outlook

We conclude this chapter with several research perspectives.

- A direction of further research could be to utilize model order reduction combined with grid adaptivity to obtain fast MPC methods. We refer to recent works combining grid adaptivity and proper orthogonal decomposition [58, 59] and works employing proper orthogonal decomposition in an MPC context, cf. [64, 99]. In that context, the turnpike property can turn out useful as it reveals a lot of structure of the dynamic problem and, in case of a steady state turnpike, can be used to construct a reduced basis of high approximation quality after solution of an elliptic OCP, which then can be enlarged by classical methods.

- We considered only parabolic problems in this chapter. An adaption to hyperbolic problems is straightforward, cf. [86] for a posteriori goal oriented methods for hyperbolic problems.

- A possible extension could be to not perform grid refinement and solution of the nonlinear problem separately, but to blend both into an adaptive algorithm. To this end, one could apply the techniques of goal oriented error estimation to the defining equations of the updates for the nonlinear algorithm. In the spirit of inexact Newton methods, one could start the refinement procedure as soon as the region of fast local convergence is entered. In order to obtain an efficient algorithm, the refinement needs to be just as aggressive to render the solution on the coarse grid interpolated to the new grid in the region of fast local convergence of the refined problem. Additionally, one can couple the estimated discretization error and the tolerance of the underlying linear solvers to render the algebraic and discretization error to be of the same order of magnitude. For an introduction to Newton algorithms with adaptive finite element methods, the reader is referred to [38, Chapter 8].

- One could investigate the use of an approximation of the system for computing the weights for spatial error estimation, i.e., (5.34), affects the error estimation. A possible approximation could be a constraint preconditioner, cf. [33] and [126, Section 7], i.e., dropping $L_{xx}(x, u, \lambda)$ and replacing $L_{uu}(x, u, \lambda)$ by its diagonal. Further, in the spirit of the DLY methodology [39], one could additionally approximate the discretized differential operator by taking the diagonal blockwise, which leads to a very efficient solution of the above system without any considerable additional effort. We expect that this does not have any considerable impact on the error estimation.

List of Figures

List of Tables

List of Algorithms

Bibliography

[1] R. A. Adams. *Sobolev Spaces.* Academic Press, Amsterdam, Boston, 1975.

[2] M. Ainsworth and J. T. Oden. *A posteriori error estimation in finite element analysis*, volume 37. John Wiley & Sons, 2011.

[3] N. Altmüller. *Model Predictive Control for Partal Differential Equations.* PhD thesis, Universität Bayreuth, 2014.

[4] H. Amann. Dual semigroups and second order linear elliptic boundary value problems. *Israel Journal of Mathematics*, 45(2):225–254, 1983.

[5] H. Amann. Linear parabolic problems involving measures. *Rev. R. Acad. Cien. Serie A. Mat.(RACSAM)*, 95:85–119, 2001.

[6] H. Amann. Maximal regularity for nonautonomous evolution equations. *Advanced Nonlinear Studies*, 4(4):417–430, 2004.

[7] B. D. Anderson and P. V. Kokotovic. Optimal control problems over large time intervals. *Automatica*, 23(3):355–363, 1987.

[8] J. Appell and P. P. Zabrejko. *Nonlinear Superposition Operators.* Cambridge Tracts in Mathematics. Cambridge University Press, 1990.

[9] J. Ball. Strongly continuous semigroups, weak solutions, and the variation of constants formula. *Proceedings of the American Mathematical Society*, 63:370–373, 1977.

[10] W. Bangerth and R. Rannacher. *Adaptive Finite Element Methods for Differential Equations.* Springer Basel AG, 2003.

[11] R. Bank, A. Sherman, and A. Weiser. Some refinement algorithms and data structures for regular local mesh refinement. *Scientific Computing, Applications of Mathematics and Computing to the Physical Sciences*, 1:3–17, 1983.

[12] V. Barbu. Controllability of parabolic and Navier-Stokes equations. *Scientiae Mathematicae Japonicae*, 56(1):143–211, 2002.

[13] C. Bardos, G. Lebeau, and J. Rauch. Sharp sufficient conditions for the observation, control, and stabilization of waves from the boundary. *SIAM Journal on Control and Optimization*, 30(5):1024–1065, 1992.

[14] P. Bastian, G. Buse, and O. Sander. Infrastructure for the coupling of dune grids. In *Numerical Mathematics and Advanced Applications 2009*, pages 107–114. Springer, 2010.

[15] R. Becker, H. Kapp, and R. Rannacher. Adaptive finite element methods for optimal control of partial differential equations: Basic concept. *SIAM Journal on Control and Optimization*, 39(1):113–132, 2000.

[16] R. Becker, D. Meidner, and B. Vexler. Efficient numerical solution of parabolic optimization problems by finite element methods. *Optimization Methods and Software*, 22(5):813–833, 2007.

[17] R. Becker and R. Rannacher. A feed-back approach to error control in finite element methods: basic analysis and examples. *East-West Journal of Numerical Mathematics*, 4(4):237–264, 1996.

[18] R. Becker and R. Rannacher. An optimal control approach to a posteriori error estimation in finite element methods. *Acta Numerica*, 10:1–102, 2001.

[19] A. Bensoussan, G. Da Prato, M. C. Delfour, and S. Mitter. *Representation and control of infinite dimensional systems*. Springer Science & Business Media, 2007.

[20] M. Blatt, A. Burchardt, A. Dedner, C. Engwer, J. Fahlke, B. Flemisch, C. Gersbacher, C. Gräser, F. Gruber, C. Grüninger, D. Kempf, R. Klöfkorn, T. Malkmus, S. Müthing, M. Nolte, M. Piatkowski, and O. Sander. The Distributed and Unified Numerics Environment, Version 2.4. *Archive of Numerical Software*, 4(100):13–29, 2016.

[21] L. Bonifacius and I. Neitzel. Second order optimality conditions for optimal control of quasilinear parabolic equations. *Mathematical Control & Related Fields*, 8(1):1–34, 2018.

[22] M. Braack and A. Ern. A posteriori control of modeling errors and discretization errors. *Multiscale Modeling & Simulation*, 1(2):221–238, 2003.

[23] D. Braess. *Finite Elements*. Cambridge University Press, 1997.

[24] T. Breiten and L. Pfeiffer. On the turnpike property and the receding-horizon method for linear-quadratic optimal control problems. *SIAM Journal on Control and Optimization*, 58(2):1077–1102, 2020.

[25] J. A. E. Bryson and Y.-C. Ho. *Applied Optimal Control*. Routledge, 1975.

[26] E. F. Camacho and C. Bordons. *Model Predictive Control*. Springer-Verlag London, 2 edition, 2007.

[27] T. Carraro and M. Geiger. Direct and indirect multiple shooting for parabolic optimal control problems. In *Multiple Shooting and Time Domain Decomposition Methods*, pages 35–67. Springer, 2015.

[28] T. Carraro, M. Geiger, and R. Rannacher. Indirect multiple shooting for nonlinear parabolic optimal control problems with control constraints. *SIAM Journal on Scientific Computing*, 36(2):A452–A481, 2014.

[29] C. Carstensen and R. Verfürth. Edge residuals dominate a posteriori error estimates for low order finite element methods. *SIAM Journal on Numerical Analysis*, 36(5):1571–1587, 1999.

[30] E. Casas. Boundary control of semilinear elliptic equations with pointwise state constraints. *SIAM Journal on Control and Optimization*, 31(4):993–1006, 1993.

[31] E. Casas, L. A. Fernández, and J. Yong. Optimal control of quasilinear parabolic equations. *Proceedings of the Royal Society of Edinburgh Section A: Mathematics*, 125(3):545–565, 1995.

[32] P. G. Ciarlet. *The Finite Element Method for Elliptic Problems.* Society for Industrial and Applied Mathematics, 2002, https://epubs.siam.org/doi/pdf/10.1137/1.9780898719208.

[33] T. F. Coleman. Linearly constrained optimization and projected preconditioned conjugate gradients. In *Proceedings of the Fifth SIAM Conference on Applied Linear Algebra*, pages 118–122, 1994.

[34] J.-M. Coron. *Control and nonlinearity.* American Mathematical Soc., 2007.

[35] R. F. Curtain and H. Zwart. *An introduction to infinite-dimensional linear systems theory*, volume 21. Springer Science & Business Media, 1995.

[36] T. Damm, L. Grüne, M. Stieler, and K. Worthmann. An exponential turnpike theorem for dissipative discrete time optimal control problems. *SIAM Journal on Control and Optimization*, 52(3):1935–1957, 2014.

[37] T. Davis and I. Duff. An unsymmetric-pattern multifrontal method for sparse LU factorization. *SIAM Journal on Matrix Analysis and Applications*, 18(1):140–158, 1997.

[38] P. Deuflhard. *Newton Methods for Nonlinear Problems: Affine Invariance and Adaptive Algorithms.* Springer Verlag Berlin Heidelberg, 2011.

[39] P. Deuflhard, P. Leinen, and H. Yserentant. Concepts of an adaptive hierarchical finite element code. *IMPACT of Computing in Science and Engineering*, 1(1):3–35, 1989.

[40] P. Deuflhard and M. Weiser. *Adaptive Numerical Solution of PDEs.* de Gruyter, 2012.

[41] T. Deutschen, S. Gasser, M. Schaller, and J. Siehr. Modeling the self-discharge by voltage decay of a NMC/graphite lithium-ion cell. *Journal of Energy Storage*, 19:113–119, 2018.

[42] W. Dörfler. A convergent adaptive algorithm for Poisson's equation. *SIAM Journal on Numerical Analysis*, 33(3):1106–1124, 1996.

[43] R. Dorfman, P. A. Samuelson, and R. M. Solow. *Linear programming and economic analysis*. Courier Corporation, 1958.

[44] K.-J. Engel and R. Nagel. *One-Parameter Semigroups for Linear Evolution Equations*. Springer Verlag New York, 2000.

[45] K. Eriksson and C. Johnson. Adaptive finite element methods for parabolic problems i: A linear model problem. *SIAM Journal on Numerical Analysis*, 28(1):43–77, 1991.

[46] K. Eriksson, C. Johnson, and V. Thomé. Time discretization of parabolic problems by the discontinuous Galerkin method. *ESAIM: Mathematical Modelling and Numerical Analysis - Modélisation Mathématique et Analyse Numérique*, 19(4):611–643, 1985.

[47] C. Esteve, B. Geshkovski, D. Pighin, and E. Zuazua. Large-time asymptotics in deep learning, 2020, arXiv:2008.02491.

[48] C. Esteve, H. Kouhkouh, D. Pighin, and E. Zuazua. The turnpike property and the long-time behavior of the Hamilton-Jacobi equation, 2020, arXiv:2006.10430.

[49] L. C. Evans. *Partial differential equations*. American Mathematical Society, Providence, R.I., 2010.

[50] L. C. Evans and R. F. Gariepy. *Measure theory and fine properties of functions*. Chapman and Hall/CRC, 2015.

[51] T. Faulwasser, K. Flaßkamp, S. Ober-Blöbaum, and K. Worthmann. A dissipativity characterization of velocity turnpikes in optimal control problems for mechanical systems, 2020, arXiv:2002.04388.

[52] T. Faulwasser, L. Grüne, J.-P. Humaloja, and M. Schaller. The interval turnpike property for adjoints, 2020, arXiv:2005.12120.

[53] T. Faulwasser, M. Korda, C. N. Jones, and D. Bonvin. On turnpike and dissipativity properties of continuous-time optimal control problems. *Automatica*, 81:297—304, 2017.

[54] A. Fursikov and O. Imanuvilov. *Controllability of evolution equations*. Suhak kangŭirok. Seoul National University, 1996.

[55] H. Gajewski, K. Gröger, and K. Zacharias. *Nichtlineare Operatorgleichungen und Operatordifferentialgleichungen*, volume 67. Akademie-Verlag Berlin, 1974.

[56] H. Goldberg, W. Kampowsky, and F. Tröltzsch. On Nemytskij operators in L^p-spaces of abstract functions. *Mathematische Nachrichten*, 155(1):127–140, 1992.

[57] S. Götschel, A. Schiela, and M. Weiser. Kaskade 7–A flexible finite element toolbox. *Computers & Mathematics with Applications*, 2020.

[58] C. Gräßle. *Adaptivity in Model Order Reduction with Proper Orthogonal Decomposition*. PhD thesis, University of Hamburg, 2019.

[59] C. Gräßle and M. Hinze. POD reduced-order modeling for evolution equations utilizing arbitrary finite element discretizations. *Advances in Computational Mathematics*, 44(6):1941–1978, 2018.

[60] P. Grisvard. *Elliptic Problems in Nonsmooth Domains*. Society for Industrial and Applied Mathematics, 2011, https://epubs.siam.org/doi/pdf/10.1137/1.9781611972030.

[61] L. Grüne. Approximation properties of receding horizon optimal control. *Jahresbericht Deutsche Mathematiker-Vereinigung*, 118(1):3–37, 2016.

[62] L. Grüne and R. Guglielmi. Turnpike properties and strict dissipativity for discrete time linear quadratic optimal control problems. *SIAM Journal on Control and Optimization*, 56(2):1282–1302, 2018.

[63] L. Grüne and R. Guglielmi. On the relation between turnpike properties and dissipativity for continuous time linear quadratic optimal control problems. *Mathematical Control and Related Fields*, Online First, June 2020.

[64] L. Grüne, L. Mechelli, S. Pirkelmann, and S. Volkwein. Performance estimates for economic model predictive control and their application in proper orthogonal decomposition-based implementations. 2019.

[65] L. Grüne and M. A. Müller. On the relation between strict dissipativity and turnpike properties. *Systems & Control Letters*, 90:45–53, 2016.

[66] L. Grüne and J. Pannek. *Nonlinear Model Predictive Control: Theory and Algorithms*. Springer Verlag London, 2016.

[67] L. Grüne and S. Pirkelmann. Economic model predictive control for time-varying system: Performance and stability results. *Optimal Control Applications and Methods*, 41(1):42–64, 2020.

[68] L. Grüne, S. Pirkelmann, and M. Stieler. Strict dissipativity implies turnpike behavior for time-varying discrete time optimal control problems. In *Control Systems and Mathematical Methods in Economics: Essays in Honor of Vladimir M. Veliov*, volume 687 of *Lecture Notes in Economics and Mathematical Systems*, pages 195–218. Springer, Cham, 2018.

[69] L. Grüne, M. Schaller, and A. Schiela. Sensitivity analysis of optimal control motivated by model predictive control. In *Proceedings of the 23rd International Symposium on Mathematical Theory of Networks and Systems. - Hong Kong*, 2018.

[70] L. Grüne, M. Schaller, and A. Schiela. Sensitivity analysis of optimal control for a class of parabolic PDEs motivated by model predictive control. *SIAM Journal on Control and Optimization*, 57(4):2753–2774, 2019.

[71] L. Grüne, M. Schaller, and A. Schiela. Exponential sensitivity and turnpike analysis for linear quadratic optimal control of general evolution equations. *Journal of Differential Equations*, 268(12):7311–7341, 2020.

[72] L. Grüne, M. Schaller, and A. Schiela. Sensitivity analysis of optimal control motivated by model predictive control, 2020, to appear in: *Proceedings of the 24th International Symposium on Mathematical Theory of Networks and Systems. - Cambridge.*

[73] L. Grüne, M. Schaller, and A. Schiela. Abstract nonlinear sensitivity and turnpike analysis and an application to semilinear parabolic pdes. *ESAIM: Control, Optimisation & Calculus of Variations*, 2021.

[74] L. Grüne, M. Schaller, and A. Schiela. Efficient MPC for parabolic PDEs with goal oriented error estimation, 2020, arXiv:2007.14446.

[75] M. Gugat and F. Hante. On the turnpike phenomenon for optimal boundary control problems with hyperbolic systems. *SIAM Journal on Control and Optimization*, 57(1):264–289, 2019.

[76] M. Gugat, E. Trélat, and E. Zuazua. Optimal Neumann control for the 1D wave equation: Finite horizon, infinite horizon, boundary tracking terms and the turnpike property. *Systems & Control Letters*, 90:61–70, 2016.

[77] W. Hackbusch. *Theorie und Numerik elliptischer Differentialgleichungen.* B.G. Teubner, 1986.

[78] R. Hartmann. *Adaptive Finite Element Methods for the Compressible Euler Equations.* PhD thesis, University of Heidelberg, 2002.

[79] A. Haurie. Optimal control on an infinite time horizon: The turnpike approach. *Journal of Mathematical Economics*, 3(1):81–102, 1976.

[80] H. K. Hesse. *Multiple shooting and mesh adaptation for PDE constrained optimization problems.* PhD thesis, Universität Heidelberg, 2008.

[81] H. K. Hesse and G. Kanschat. Mesh adaptive multiple shooting for partial differential equations. Part i: linear quadratic optimal control problems. *Journal of Numerical Mathematics*, 17(3):195–217, 2009.

[82] M. Hinze, R. Pinnau, M. Ulbrich, and S. Ulbrich. *Optimization with PDE constraints*, volume 23 of *Mathematical Modelling: Theory and Applications*. Springer, New York, 2009.

[83] B. Jacob and H. J. Zwart. *Linear port-Hamiltonian systems on infinite-dimensional spaces*, volume 223. Springer Science & Business Media, 2012.

[84] T. Kato. *Perturbation theory for linear operators*, volume 132. Springer Science & Business Media, 2013.

[85] R. Kress. *Linear integral equations*, volume 17. Springer, 2014.

[86] A. Kröner. Adaptive finite element methods for optimal control of second order hyperbolic equations. *Computational Methods in Applied Mathematics*, 11(2):214–240, 2011.

[87] O. A. Ladyzhenskaia, V. A. Solonnikov, and N. N. Ural'ceva. *Linear and quasi-linear equations of parabolic type*, volume 23. American Mathematical Soc., 1968.

[88] G. Lance, E. Trélat, and E. Zuazua. Turnpike in optimal shape design. *IFAC-PapersOnLine*, 52(16):496–501, 2019.

[89] I. Lasiecka, J.-L. Lions, and R. Triggiani. Nonhomogeneous boundary value problems for second order hyperbolic operators. *Journal de mathématiques pures end appliquées*, 65(9):149–192, 1986.

[90] I. Lasiecka and R. Triggiani. *Control Theory for Partial Differential Equations: Volume 1, Abstract Parabolic Systems: Continuous and approximation theories*, volume 1. Cambridge University Press, 2000.

[91] I. Lasiecka and R. Triggiani. *Control Theory for Partial Differential Equations: Volume 2, Abstract Hyperbolic-Like Systems Over a Finite Time Horizon: Continuous and Approximation Theories*, volume 2. Cambridge University Press, 2000.

[92] P. D. Lax. *Linear Algebra and its applications*. Wiley-Interscience, second edition, 2007.

[93] D. Leykekhman and B. Vexler. Discrete maximal parabolic regularity for Galerkin finite element methods. *Numerische Mathematik*, 135(3):923–952, 2017.

[94] D. Leykekhman and B. Vexler. Discrete maximal parabolic regularity for Galerkin finite element methods for nonautonomous parabolic problems. *SIAM Journal on Numerical Analysis*, 56(4):2178–2202, 2018.

[95] X. Li and J. Yong. *Optimal Control Theory for Infinite Dimensional Systems*. Birkhäuser, 1995.

[96] J.-L. Lions. *Optimal Control of Systems Governed by Partial Differential Equations*. Springer Verlag Berlin Heidelberg, 1971.

[97] J.-L. Lions. Exact controllability, stabilization and perturbations for distributed systems. *SIAM review*, 30(1):1–68, 1988.

[98] L. Lubkoll, A. Schiela, and M. Weiser. An affine covariant composite step method for optimization with PDEs as equality constraints. *Optimization Methods and Software*, 32(5):1132–1161, 2017.

[99] L. Mechelli and S. Volkwein. POD-based economic model predictive control for heat-convection phenomena. In *European Conference on Numerical Mathematics and Advanced Applications*, pages 663–671. Springer, 2017.

[100] D. Meidner. *Adaptive Space-Time Finite Element Methods for Optimization Problems Governed by Nonlinear Parabolic Systems.* PhD thesis, Universität Heidelberg, 2008.

[101] D. Meidner and B. Vexler. Adaptive space-time finite element methods for parabolic optimization problems. *SIAM Journal on Control and Optimization*, 46(1):116–142, 2007.

[102] D. Meidner and B. Vexler. A priori error estimates for space-time finite element discretization of parabolic optimal control problems. Part i: Problems without control constraints. *SIAM Journal on Control and Optimization*, 47(3):1150–1177, 2008.

[103] D. Meidner and B. Vexler. Optimal error estimates for fully discrete Galerkin approximations of semilinear parabolic equations. *ESAIM: M2AN*, 52(6):2307–2325, 2018.

[104] H. Meinlschmidt. *Analysis and Optimal Control of Quasilinear Parabolic Evolution Equations in Divergence Form on Rough Domains.* PhD thesis, Universität Darmstadt, 2017.

[105] S. Na, S. Shin, M. Anitescu, and V. M. Zavala. Overlapping Schwarz decomposition for nonlinear optimal control, 2020, arXiv:2005.06674.

[106] I. Neitzel and B. Vexler. A priori error estimates for space–time finite element discretization of semilinear parabolic optimal control problems. *Numerische Mathematik*, 120(2):345–386, 2012.

[107] J. Nečas. *Direct Methods in the Theory of Elliptic Equations (Springer Monographs in Mathematics)*. Springer, 2012.

[108] N. S. Papageorgiou. On the optimal control of strongly nonlinear evolution equations. *Journal of Mathematical Analysis and Applications*, 164(1):83–103, 1992.

[109] A. Pazy. *Semigroups of Linear Operators and Applications to Partial Differential Equations.* Springer Verlag New York, 1983.

[110] D. Pighin. The turnpike property in semilinear control, 2020, arXiv:2004.03269.

[111] D. Pighin and N. Sakamoto. The turnpike with lack of observability, 2020, arXiv:2007.14081.

[112] A. Porretta. On the turnpike property for mean field games. *Minimax Theory and its Applications*, 3(2):285–312, 2018.

[113] A. Porretta and E. Zuazua. Long time versus steady state optimal control. *SIAM Journal on Control and Optimization*, 51(6):4242–4273, 2013.

[114] A. Porretta and E. Zuazua. Remarks on long time versus steady state optimal control. In *Mathematical paradigms of climate science*, pages 67–89. Springer, 2016.

[115] S. J. Qin and T. A. Badgwell. A survey of industrial model predictive control technology. *Control engineering practice*, 11(7):733–764, 2003.

[116] R. Rannacher, B. Vexler, and W. Wollner. *A Posteriori Error Estimation in PDE-constrained Optimization with Pointwise Inequality Constraints*, pages 349–373. Springer Basel, Basel, 2012.

[117] J. Rauch, M. Taylor, and R. Phillips. Exponential decay of solutions to hyperbolic equations in bounded domains. *Indiana University Mathematics Journal*, 24(1):79–86, 1974.

[118] J. B. Rawlings, D. Q. Mayne, and M. Diehl. *Model Predictive Control: Theory, Computation, and Design*, volume 2. Nob Hill Publishing Madison, WI, 2017.

[119] J. P. Raymond and H. Zidani. Hamiltonian Pontryagin's principles for control problems governed by semilinear parabolic equations. *Applied Mathematics and Optimization*, 39(2):143–177, 1999.

[120] T. Richter, A. Springer, and B. Vexler. Efficient numerical realization of discontinuous Galerkin methods for temporal discretization of parabolic problems. *Numerische Mathematik*, 124(1):151–182, 2012.

[121] T. Richter and T. Wick. Variational localizations of the dual weighted residual estimator. *Journal of Computational and Applied Mathematics*, 279:192–208, 2015.

[122] F. Rothe. Uniform bounds from bounded L_p-functionals in reaction-diffusion equations. *Journal of Differential Equations*, 45(2):207–233, 1982.

[123] N. Sakamoto, D. Pighin, and E. Zuazua. The turnpike property in nonlinear optimal control–a geometric approach. In *2019 IEEE 58th Conference on Decision and Control (CDC)*, pages 2422–2427, 2019.

[124] D. Salamon. Infinite-dimensional linear systems with unbounded control and observation: a functional analytic approach. *Transactions of the American Mathematical Society*, 300(2):383–431, 1987.

[125] D. Salamon. Realization theory in Hilbert space. *Mathematical Systems Theory*, 21(1):147–164, 1988.

[126] M. Schaller, A. Schiela, and M. Stöcklein. A composite step method with inexact step computations for PDE constrained optimization. Preprint SPP1962-098, 2018.

[127] A. Schiela. A concise proof for existence and uniqueness of solutions of linear parabolic PDEs in the context of optimal control. *System & Control Letters*, 62(10):895–901, 2013.

[128] M. Schmich and B. Vexler. Adaptivity with dynamic meshes for space-time finite element discretizations of parabolic equations. *SIAM Journal on Scientific Computing*, 30(1):369–393, 2008.

[129] F. Schwenninger. Input-to-state stability for parabolic boundary control: Linear and semi-linear systems, 2019, arXiv:1908.08317.

[130] E. D. Sontag. *Mathematical Control Theory.* Springer New York, 1998.

[131] A. Springer and B. Vexler. Third order convergent time discretization for parabolic optimal control problems with control constraints. *Computational Optimization and Applications*, 57(1):205–240, 2013.

[132] O. Staffans. *Well-posed linear systems*, volume 103. Cambridge University Press, 2005.

[133] H. Tanabe. *Equations of Evolution*, volume 6 of *Monographs and studies in mathematics.* Pitman (Advanced Publishing Program), London, 1979.

[134] E. Trélat and C. Zhang. Integral and measure-turnpike properties for infinite-dimensional optimal control systems. *Mathematics of Control, Signals, and Systems*, 30(1):3, 2018.

[135] E. Trélat, C. Zhang, and E. Zuazua. Steady-state and periodic exponential turnpike property for optimal control problems in Hilbert spaces. *SIAM Journal on Control and Optimization*, 56(2):1222–1252, 2018.

[136] E. Trélat and E. Zuazua. The turnpike property in finite-dimensional nonlinear optimal control. *Journal of Differential Equations*, 258(1):81–114, 2015.

[137] H. Triebel. *Interpolation theory, function spaces, differential operators.* North-Holland, 1978.

[138] F. Tröltzsch. *Optimal control of partial differential equations: theory, methods, and applications*, volume 112. American Mathematical Soc., 2010.

[139] M. Tucsnak and G. Weiss. *Observation and control for operator semigroups.* Springer Science & Business Media, 2009.

[140] M. Tucsnak and G. Weiss. Well-posed systems–the LTI case and beyond. *Automatica*, 50(7):1757–1779, 2014.

[141] B. Vexler. *Adaptive finite element methods for parameter identification problems.* PhD thesis, Universität Heidelberg, 2004.

[142] M. Warma and S. Zamorano. Exponential turnpike property for fractional parabolic equations with non-zero exterior data, 2020, arXiv:2008.03773.

[143] M. Weiser. On goal-oriented adaptivity for elliptic optimal control problems. *Optimization Methods & Software*, 28(5):969–992, 2013.

[144] M. Weiser. *Inside Finite Elements*. Walter de Gruyter GmbH & Co KG, 2016.

[145] G. Weiss. Regular linear systems with feedback. *Mathematics of Control, Signals and Systems*, 7(1):23–57, 1994.

[146] G. Weiss, O. J. Staffans, and M. Tucsnak. Well-posed linear systems–a survey with emphasis on conservative systems. *International Journal of Applied Mathematics and Computer Science*, 11:7–33, 2001.

[147] D. Werner. *Funktionalanalysis*. Springer Verlag Berlin Heidelberg, 2011.

[148] J. C. Willems. Dissipative dynamical systems part i: General theory. *Archive for rational mechanics and analysis*, 45(5):321–351, 1972.

[149] J. C. Willems. Dissipative dynamical systems part ii: Linear systems with quadratic supply rates. *Archive for rational mechanics and analysis*, 45(5):352–393, 1972.

[150] J. Wloka. *Partial Differential Equations*. Cambridge University Press, 1987.

[151] M.-Y. Wu. A note on stability of linear time-varying systems. *IEEE Transactions on Automatic Control*, 19(2):162–162, 1974.

[152] A. Yagi. *Abstract parabolic evolution equations and their applications*. Springer Science & Business Media, 2009.

[153] K. Yosida. *Functional Analysis*. Grundlehren der mathematischen Wissenschaften. Springer Verlag Berlin Heidelberg, 1978.

[154] J. Zabczyk. *Mathematical Control Theory*. Birkhäuser, 1992.

[155] S. Zamorano. Turnpike property for two-dimensional Navier–Stokes equations. *Journal of Mathematical Fluid Mechanics*, 20(3):869–888, 2018.

[156] M. Zanon, L. Grüne, and M. Diehl. Periodic optimal control, dissipativity and MPC. *IEEE Transactions on Automatic Control*, 62(6):2943–2949, 2017.

[157] A. J. Zaslavski. *Turnpike theory of continuous-time linear optimal control problems*. Springer, 2016.

[158] E. Zeidler. *Nonlinear Functional Analysis and its Applications- II/A: Linear Monotone Operators*. Springer Verlag New York, 1990.

[159] E. Zuazua. Controllability and observability of partial differential equations: some results and open problems. *Handbook of Differential Equations: Evolutionary Equation*, 3:527–621, 2007.

Publications

This thesis contains results from the following publications:

- L. Grüne, M. Schaller, and A. Schiela. Sensitivity analysis of optimal control motivated by model predictive control. In *Proceedings of the 23rd International Symposium on Mathematical Theory of Networks and Systems. - Hong Kong*, 2018
- L. Grüne, M. Schaller, and A. Schiela. Sensitivity analysis of optimal control for a class of parabolic PDEs motivated by model predictive control. *SIAM Journal on Control and Optimization*, 57(4):2753–2774, 2019
- L. Grüne, M. Schaller, and A. Schiela. Exponential sensitivity and turnpike analysis for linear quadratic optimal control of general evolution equations. *Journal of Differential Equations*, 268(12):7311–7341, 2020
- L. Grüne, M. Schaller, and A. Schiela. Sensitivity analysis of optimal control motivated by model predictive control, 2020, to appear in: *Proceedings of the 24th International Symposium on Mathematical Theory of Networks and Systems. - Cambridge*
- L. Grüne, M. Schaller, and A. Schiela. Efficient MPC for parabolic PDEs with goal oriented error estimation, 2020, arXiv:2007.14446
- L. Grüne, M. Schaller, and A. Schiela. Abstract nonlinear sensitivity and turnpike analysis and an application to semilinear parabolic pdes. *ESAIM: Control, Optimisation & Calculus of Variations*, 2021

The content of the first two publications is included in Chapter 3, whereas the results of the third publication are contained in Chapter 2. The fourth publication partly announced results of Chapters 4 and 5 in the format of an extended abstract. The content of the fifth publication is included in Chapter 5. The last publication is included in Chapter 4.

Further, the author contributed to the following publications, the results of which are not included in this thesis:

- T. Deutschen, S. Gasser, M. Schaller, and J. Siehr. Modeling the self-discharge by voltage decay of a NMC/graphite lithium-ion cell. *Journal of Energy Storage*, 19:113–119, 2018

- M. Schaller, A. Schiela, and M. Stöcklein. A composite step method with inexact step computations for PDE constrained optimization. Preprint SPP1962-098, 2018

- T. Faulwasser, L. Grüne, J.-P. Humaloja, and M. Schaller. The interval turnpike property for adjoints, 2020, arXiv:2005.12120

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.

Ich erkläre zudem, dass ich keine früheren Promotionsversuche unternommen habe.

Ich bin damit einverstanden, dass die elektronische Fassung dieser Dissertation unter Wahrung meiner Urheberrechte und des Datenschutzes einer gesonderten Prüfung unterzogen werden kann und dass bei Verdacht wissenschaftlichen Fehlverhaltens Ermittlungen durch universitätsinterne Organe der wissenschaftlichen Selbstkontrolle stattfinden können.

Ilmenau, den ____________ ________________________

(Manuel Schaller)