

**ABSTRACT NONLINEAR SENSITIVITY AND TURNPIKE ANALYSIS
AND AN APPLICATION TO SEMILINEAR PARABOLIC PDES***LARS GRÜNE¹, MANUEL SCHALLER² AND ANTON SCHIELA¹

Abstract. We analyze the sensitivity of the extremal equations that arise from the first order necessary optimality conditions of nonlinear optimal control problems with respect to perturbations of the dynamics and of the initial data. To this end, we present an abstract implicit function approach with scaled spaces. We will apply this abstract approach to problems governed by semilinear PDEs. In that context, we prove an exponential turnpike result and show that perturbations of the extremal equation's dynamics, e.g., discretization errors decay exponentially in time. The latter can be used for very efficient discretization schemes in a Model Predictive Controller, where only a part of the solution needs to be computed accurately. We showcase the theoretical results by means of two examples with a nonlinear heat equation on a two-dimensional domain.

1991 Mathematics Subject Classification. 49K20, 49K40, 93D20, 35K55, 35Q93.

March 16, 2021.

1. INTRODUCTION

In this paper we provide an abstract framework for exponential sensitivity analysis of nonlinear optimal control problems with respect to perturbations of the right-hand side of the first-order necessary optimality conditions. We extend previous results regarding linear quadratic optimal control problems, where an exponential damping property was proven for problems governed by non-autonomous parabolic equations in [25] and for problems governed by autonomous general evolution equations in [27]. The main tool in these works is a bound on the solution operator of the first order optimality conditions that is independent of the time horizon, which can be deduced under stabilizability and detectability assumptions. In this work considering nonlinear

Keywords and phrases: Nonlinear Optimal Control, Sensitivity Analysis, Turnpike Property, Model Predictive Control

* *This work was supported by the DFG Grants GR 1569/17-1 and SCHI 1379/5-1. Some results were partially announced in [26] in the form of an extended abstract.*

¹ Mathematical Institute, University of Bayreuth, 95447 Bayreuth, Germany,
(e-mail: lars.gruene@uni-bayreuth.de & anton.schiela@uni-bayreuth.de).

² Institute for Mathematics, Technische Universität Ilmenau, 98693 Ilmenau, Germany,
(e-mail: manuel.schaller@tu-ilmenau.de).

problems, we derive an implicit function theorem that allows for estimates in scaled spaces, where, if the solution operator to the linearized system is bounded independently of the time horizon, all involved neighborhoods and constants are independent of the time horizon. As a consequence, we obtain exponential sensitivity results for nonlinear problems, stating that perturbations of the first order optimality conditions decay exponentially in time.

This sensitivity analysis has several important applications. First, exponential decay of perturbations gives rise to efficient numerical methods. In [25, 28, 48] efficient discretization methods for Model Predictive Control (MPC) were presented. MPC is a feedback control technique, where an optimal control problem on a infinite or indefinite horizon is approximated by a series of optimal control problems on finite horizons $T > 0$. In every iteration of the MPC algorithm, an optimal control for a problem on the time interval $[0, T]$ is computed. Then, an initial part of the control up to time $\tau > 0$ is used as a feedback, where often $\tau \ll T$. The resulting state $x(\tau)$ is measured or estimated and the process is repeated on the horizon $[\tau, T + \tau]$ with initial datum $x(\tau)$. For an in-depth introduction, its approximation properties and treatment of various aspects related to MPC the interested reader is referred to [20, 23, 43]. As only an initial part of the computed solution is of relevance for the feedback, an exponential decay of perturbation allows for efficient solution of the corresponding OCPs on space and time grids that are finer on $[0, \tau]$ than on the remainder $[\tau, T]$. This specialized discretization can be done in a a priori fashion [25, 36] or with goal oriented a posteriori methods [28]. Further, in [36], a Schwarz decomposition method was considered which also strongly leverages the exponential decay of perturbations.

A second application of the proposed sensitivity analysis is the derivation of turnpike properties. A particular turnpike property is the steady state turnpike property, which is a feature of solutions to autonomous optimal control problems. In a nutshell, it states that the solution of the optimal control problem on a long time horizon resides near a steady-state of the dynamics, the so-called turnpike, for the majority of the time. We briefly recall some of the existing literature on turnpike analysis. The linear quadratic case for control of evolution equations was considered in [7, 11, 25, 27, 29, 30]. A turnpike property for shape optimization was introduced in [34, 50]. Nonlinear finite dimensional problems were considered in [52], including the case of nonlinear initial and terminal conditions. This was extended to a Hilbert space setting in [41, 42, 51]. A turnpike result for the two-dimensional Navier-Stokes equations was obtained in [58]. These works analyze the turnpike property via the extremal equations and are of local nature, i.e., the initial resp. terminal value for state and adjoint need to be close to the turnpike. Further, in [39], 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 [46]. Another approach that leads to global turnpike properties is stability analysis based on a dissipativity concept. Motivated by the seminal papers by Willems [56, 57], a notion of dissipativity for optimal control problems can be defined, where the supply rate is defined via the cost functional. Assuming this dissipativity property, a global turnpike result for states and controls was deduced in, e.g., [17, 22, 24] or [23, 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 [16]. The connection of dissipativity and the turnpike property is also discussed in [21, 49]. Recently, turnpike properties for non-observable systems [15, 40], for problems arising in deep learning [13] and for fractional parabolic problems [54] were presented. Further, the connection of turnpike properties and long-time behavior of the Hamilton-Jacobi equation was analyzed in [14].

In this work, we provide a framework to lift the approach of analyzing the extremal equations' solution operator that was considered in [25, 27, 54] to the nonlinear case. Thus, our approach and the turnpike results in this paper are similar to [51, 52] in the sense that we also consider stabilizability conditions to derive local results via analysis of the extremal equations. A particular novelty of this work is the implicit function theorem Theorem 3.1 that provides a flexible and rigorous tool to conclude a local nonlinear turnpike result by means of analyzing the solution operator of the linearized equations. In that context, turnpike properties in norms particularly tailored to the corresponding regularity properties of the underlying problem can be deduced. The choices could range from, e.g., uniform norms in a general semigroup context, to integral norms with values in Sobolev spaces for parabolic problems. Here, we provide a particular application to semilinear parabolic problems.

Our approach is conceptually simple: the nonlinear problem is considered as a perturbation of a linearized problem and an implicit function theorem is applied. Then exponential sensitivity and turnpike behavior are inherited from the linearized problem to the nonlinear problem under a smallness condition. What makes the analysis delicate is that substantial results of this form require independence of the involved quantities from the length of the time interval T under consideration. For example, the exponential rate of decay μ should not degenerate as $T \rightarrow \infty$. It is thus necessary to check all estimates for uniformity in T .

This work is organized as follows. After introducing the problem of interest in Section 2, we present an implicit function theorem in Section 3 that allows for scaled estimates for solutions of the nonlinear first-order necessary optimality conditions if the linearization of the latter is T -uniformly continuous and T -uniformly invertible, where T is the time horizon of the optimal control problem. The former assumption will be analyzed in Section 3.2, whereas we will verify the latter for a class of problems involving a semilinear heat equation under stabilizability conditions in Section 4. Consequently, we present in Section 5 the main results for semilinear problems: We provide a turnpike result in Corollary 5.2 and a sensitivity result with respect to perturbations of the dynamics in Corollary 5.3. Finally, we present numerical examples that illustrate the theoretical findings for semilinear equations and further provide an example with a boundary controlled quasilinear equation that is not covered by theoretical results yet, motivating directions of further research. We will illustrate the turnpike property and propose and evaluate an a priori discretization method specialized for MPC.

2. SETTING AND PRELIMINARIES

We briefly define the nonlinear optimal control problem of interest and formally derive the optimality conditions. Let $\Omega \subset \mathbb{R}^n$, $n \geq 2$ be a bounded domain with smooth boundary. Further, suppose that $(V, \|\cdot\|_V)$ is a separable and reflexive Banach space such that $V \hookrightarrow L_2(\Omega) \cong L_2(\Omega)^* \hookrightarrow V^*$ form a Gelfand triple. We will abbreviate $\langle \cdot, \cdot \rangle := \langle \cdot, \cdot \rangle_{L_2(\Omega)}$ and $\|\cdot\| := \|\cdot\|_{L_2(\Omega)}$. Further we denote

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

where the time derivative is meant in a weak sense. The control space U will be assumed to be a Hilbert space with scalar product denoted by $\langle \cdot, \cdot \rangle_U$ and induced norm $\|\cdot\|_U$. We consider the

following nonlinear optimal control problem.

$$\begin{aligned} \min_{(x,u)} J(x, u) &:= \int_0^T \bar{J}(t, x(t)) + \|R(u(t) - u_d(t))\|_U^2 dt \\ \text{s.t. } x'(t) &= \bar{A}(x(t)) + \bar{B}u(t) + d(t) \\ x(0) &= x_0, \end{aligned} \quad (1)$$

where $x_0 \in H$, $d \in L_2(0; T; V^*)$, $J(x, u)$ is a sufficiently smooth functional, $\bar{B}: U \rightarrow V^*$ is a continuous and linear operator, and $\bar{A}: V \rightarrow V^*$ is a sufficiently smooth operator. We define $B: L_2(0, T; U) \rightarrow L_2(0, T; V^*)$ via

$$\langle Bu, \lambda \rangle_{L_2(0, T; V^*) \times L_2(0, T; V)} := \int_0^T \langle \bar{B}u(t), \lambda(t) \rangle_{V^* \times V} dt$$

and $A: L_2(0, T; V) \rightarrow L_2(0, T; V^*)$ by

$$\langle A(x), \lambda \rangle_{L_2(0, T; V^*) \times L_2(0, T; V)} := \int_0^T \langle \bar{A}(x(t)), \lambda(t) \rangle_{V^* \times V} dt.$$

for $\lambda \in L_2(0, T; V)$. We will assume that the optimal control problem (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-semi-continuity and coercivity properties of the objective functional. A second factor can be to establish the existence of a continuous control to state map. In the linear case, i.e., if $A(x) = Ax$ this follows if $-A$ satisfies a Gårding inequality, cf. [47, Theorem 3.4]:

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

For solvability of semilinear equations with globally Lipschitz semilinearities we refer to [38, Chapter 6] and [53, Chapter 5]. Locally Lipschitz semilinearities were treated in [44], where global existence of solutions was ensured by sufficiently regular data $(Bu + d, x_0)$, such that the solution is bounded, i.e., $x \in L_\infty(0, T; L_\infty(\Omega))$. For an in-depth analysis of optimal control problems governed by quasilinear parabolic equations, the interested reader is referred to [6, 10, 33, 35, 37].

We introduce a Lagrange multiplier $(\lambda, \lambda_0) \in L_2(0, T; V) \times L_2(\Omega)$ and define the Lagrange function via

$$\begin{aligned} L(x, u, (\lambda, \lambda_0)) &:= J(x, u) + \langle x' - Ax, \lambda \rangle_{(L_2(0, T; V) \times L_2(\Omega))^* \times (L_2(0, T; V) \times L_2(\Omega))} \\ &\quad - \langle Bu + d, \lambda \rangle_{L_2(0, T; V^*) \times L_2(0, T; V)} + \langle x(0) - x_0, \lambda_0 \rangle. \end{aligned}$$

Proceeding formally we obtain the first-order necessary optimality conditions

$$L'(x, u, \lambda) = \begin{pmatrix} J_x(x, u) - \lambda' - A'(x)^* \lambda \\ \lambda(T) \\ R^* Ru - B^* \lambda \\ x' - A(x) - Bu - d \\ x(0) - x_0 \end{pmatrix} = 0. \quad (2)$$

with $\lambda \in W([0, T])$ as $J_x(x, u) \in L_2(0, T; V^*)$, cf. [47, Proposition 3.8]. Setting $Q := R^*R$ and $L_r(x, \lambda) := L(x, Q^{-1}B^*\lambda + u_d, \lambda)$, the reduced extremal equations read

$$L'_r(x, \lambda) = \begin{pmatrix} J_x(x, u) - \lambda' - A'(x)^*\lambda \\ \lambda(T) \\ x' - A(x) - BQ^{-1}B^*\lambda - Bu_d - d \\ x(0) - x_0 \end{pmatrix} = 0. \quad (3)$$

We present two perturbations of the extremal equations (3) that we aim to analyze in this work. In order to obtain a sensitivity result with respect to perturbations of the dynamics, we introduce a perturbation $\varepsilon = (\varepsilon_1, \varepsilon_T, \varepsilon_2, \varepsilon_0) \in (L_2(0, T; V^*) \times L_2(\Omega))^2$, which could result from, e.g., temporal or spatial discretization errors. We denote by $(\tilde{x}, \tilde{u}, \tilde{\lambda}) \in W([0, T]) \times L_2(0, T; U) \times W([0, T])$ the solution of this perturbed system, i.e.,

$$L'_r(\tilde{x}, \tilde{\lambda}) = \begin{pmatrix} J_x(\tilde{x}, \tilde{u}) - \tilde{\lambda}' - A'(\tilde{x})^*\tilde{\lambda} \\ \tilde{\lambda}(T) \\ \tilde{x}' - A(\tilde{x}) - BQ^{-1}B^*\tilde{\lambda} - Bu_d - d \\ \tilde{x}(0) - x_0 \end{pmatrix} = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_T \\ \varepsilon_2 \\ \varepsilon_0 \end{pmatrix} \quad (4)$$

and $\tilde{u} = Q^{-1}B^*\tilde{\lambda} + u_d$. Further, to derive a turnpike result, i.e., a sensitivity result with respect to perturbations of the initial and terminal condition, we will consider the steady-state problem as a perturbation of the first order conditions of the dynamic problem. In that context, we will always assume that the cost functional is given by $J(x, u) = \int_0^T \bar{J}(x(t)) + \frac{1}{2}\|R(u(t) - u_d)\|^2 dt$, i.e., \bar{J} itself does not explicitly depend on time and $u_d \in U$, $R \in L(U, U)$, $d \in V^*$ are independent of time. To indicate this time-independence denote $\bar{R} := R$, $\bar{u}_d := u_d$ and $\bar{d} := d$. The corresponding steady-state problem then reads

$$\begin{aligned} \min_{(\bar{x}, \bar{u})} \bar{J}(\bar{x}) + \frac{1}{2}\|\bar{R}(\bar{u} - \bar{u}_d)\|_U^2 \\ \text{s.t. } -\bar{A}(\bar{x}) = \bar{B}\bar{u} + \bar{d}, \end{aligned} \quad (5)$$

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 $\bar{L}(\bar{x}, \bar{u}, \bar{\lambda}) := \bar{J}(\bar{x}, \bar{u}) - \langle \bar{A}(\bar{x}) - \bar{B}\bar{u}, \bar{\lambda} \rangle_{V^* \times V}$, which leads to the first order conditions

$$\bar{L}'(\bar{x}, \bar{u}, \bar{\lambda}) = \begin{pmatrix} \bar{J}_x(\bar{x}) - \bar{A}'(\bar{x})^*\bar{\lambda} \\ \bar{R}^*\bar{R}\bar{u} - \bar{B}^*\bar{\lambda} \\ -\bar{A}(\bar{x}) - \bar{B}\bar{u} - \bar{d} \end{pmatrix} = 0. \quad (6)$$

Eliminating the control via $\bar{u} = \bar{Q}^{-1}\bar{B}^*\bar{\lambda} + u_d$, where $\bar{Q} = \bar{R}^*\bar{R}$ and defining the reduced static Lagrangian $\bar{L}_r(\bar{x}, \bar{\lambda}) := \bar{L}(\bar{x}, \bar{Q}^{-1}\bar{B}^*\bar{\lambda} + u_d, \bar{\lambda})$, 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 by adding $\bar{\lambda}' = \bar{x}' = 0$ and initial resp. terminal value to the equations, i.e.,

$$L_r'(\bar{x}, \bar{\lambda}) = \begin{pmatrix} J_x(\bar{x}, \bar{u}) - \bar{\lambda}' - A'(\bar{x})^* \bar{\lambda} \\ \bar{\lambda}(T) \\ \bar{x}' - A(\bar{x}) - BQ^{-1}B^* \bar{\lambda} - Bu_d - \bar{d} \\ \bar{x}(0) - x_0 \end{pmatrix} = \begin{pmatrix} 0 \\ \bar{\lambda} \\ 0 \\ \bar{x} - x_0 \end{pmatrix}. \quad (7)$$

To obtain localized estimates in time, we consider a smooth scaling function $s: \mathbb{R}^{\geq 0} \rightarrow \mathbb{R}$ with $s(t) > 0$ for all $t \in \mathbb{R}^{\geq 0}$ and endow $L_p(0, T; V)$ with the scaled norm

$$\|x\|_{L_p^s(0, T; V)} := \|sx\|_{L_p(0, T; V)} \quad (8)$$

for any $1 \leq p \leq \infty$. The equivalence of this norm to the standard $L_p(0, T; V)$ -norm follows from the positivity of s as we get for $1 \leq p < \infty$ that

$$\min_{t \in [0, T]} s(t) \left(\int_0^T \|x(t)\|_V^p dt \right)^{\frac{1}{p}} \leq \left(\int_0^T \|s(t)x(t)\|_V^p dt \right)^{\frac{1}{p}} \leq \max_{t \in [0, T]} s(t) \left(\int_0^T \|x(t)\|_V^p dt \right)^{\frac{1}{p}} \quad (9)$$

and

$$\min_{t \in [0, T]} s(t) \operatorname{ess\,sup}_{t \in [0, T]} \|x(t)\|_V \leq \operatorname{ess\,sup}_{t \in [0, T]} \|s(t)x(t)\|_V \leq \max_{t \in [0, T]} s(t) \operatorname{ess\,sup}_{t \in [0, T]} \|x(t)\|_V. \quad (10)$$

As $L_p(0, T; V)$ with the standard norm is a Banach space, by the equivalence of the norms above, $(L_p(0, T; V), \|\cdot\|_{L_p^s(0, T; V)})$ is also a Banach space. Note that the equivalence of norms can deteriorate for $T \rightarrow \infty$ depending on the scaling function. Later in Section 4, we will consider a closed operator $A: D(A) \subset L_2(\Omega) \rightarrow L_2(\Omega)$ that is a generator of an analytic semigroup in $L_2(\Omega)$, where $D(A)$ is the domain of A endowed by the graph norm $\|\cdot\| + \|A \cdot\|$. We will impose either homogeneous Neumann or homogeneous Dirichlet boundary conditions and thus the domain will, in our case, be either $D(A) = \{v \in H^2(\Omega) \mid \frac{\partial}{\partial \nu_A} v = 0\}$, where $\frac{\partial}{\partial \nu_A}$ is the conormal derivative corresponding to A , or $D(A) = H^2(\Omega) \cap H_0^1(\Omega)$, i.e., $V = D((-A)^{\frac{1}{2}})$. Correspondingly, we will set $V = H^1(\Omega)$ or $V = H_0^1(\Omega)$ depending on the choice of boundary conditions. We will denote

$$W^{1,2}(0, T, D(A), L_2(\Omega)) := \{v \in L_2(0, T; D(A)) \mid v' \in L_2(0, T; L_2(\Omega))\},$$

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

where the time derivative is meant in a weak sense. We have the T -independent embedding $W^{1,2}(0, T, D(A), L_2(\Omega)) \hookrightarrow C(0, T; V)$, cf. [5, Part II-1, Remark 4.1, Remark 4.2]. For this vector-valued Sobolev space, we will also utilize a scaled norm, i.e.,

$$\|v\|_{W_s^{1,2}(0, T, D(A), L_2(\Omega))} := \|sv\|_{W^{1,2}(0, T, D(A), L_2(\Omega))}.$$

For the scaling terms we have in mind, i.e., exponential functions, one can straightforwardly show that the norm $\|v\|_{W_s^{1,2}(0, T, D(A), L_2(\Omega))} = \|sv\|_{L_2(0, T; D(A))} + \|(sv)'\|_{L_2(0, T; L_2(\Omega))}$ is equivalent to $\|sv\|_{L_2(0, T; D(A))} + \|s v'\|_{L_2(0, T; L_2(\Omega))} = \|v\|_{L_2^s(0, T; D(A))} + \|v'\|_{L_2^s(0, T; L_2(\Omega))}$. Thus, by the equivalence of scaled and unscaled L_2 -norms shown above, $\|\cdot\|_{W_s^{1,2}(0, T, D(A), L_2(\Omega))}$ is equivalent

to the standard norm $\|\cdot\|_{W^{1,2}(0,T,D(A),L_2(\Omega))}$ with constants strongly depending on T . Hence, $\left(W^{1,2}(0,T,D(A),L_2(\Omega)), \|\cdot\|_{W_s^{1,2}(0,T,D(A),L_2(\Omega))}\right)$ is a Banach space. Finally, whenever we write $V^{s(t)}$ for either $V = H^1(\Omega)$ or $V = H_0^1(\Omega)$ and $t \in [0, T]$, we mean V endowed with the equivalent norm $s(t)\|\cdot\|_V$. If X is a Banach space we denote by $L(X) := L(X, X)$ the space of bounded linear operators from X to X .

3. AN ABSTRACT FRAMEWORK FOR SENSITIVITY ANALYSIS

Considering the nonlinear equation (3) and the perturbations (4) resp. (7), the question we aim to answer is the following: How do $\tilde{z} = (\tilde{x}, \tilde{u})$ and $\bar{z} = (\bar{x}, \bar{u})$ differ from $z = (x, u)$ 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 by means of the scaled norms introduced above. To this end, denoting by Z a solution space and by E a perturbation space, we introduce a nonlinear operator

$$G: Z \times E \rightarrow E$$

defined by

$$G(z, \varepsilon) := L_r'(z) - \varepsilon \quad \forall (z, \varepsilon) \in Z \times E. \quad (11)$$

It is clear that

- $G(z, 0) = 0$ for any solution $z = (x, \lambda)$ of the dynamic problem (3),
- $G(\bar{z}, (0, \bar{\lambda}, 0, \bar{x} - x_0)) = 0$ for any solution $\bar{z} = (\bar{x}, \bar{\lambda})$ of the static problem (6),
- $G(\tilde{z}, (\varepsilon_1, \varepsilon_T, \varepsilon_2, \varepsilon_0)) = 0$ for any solution $\tilde{z} = (\tilde{x}, \tilde{\lambda})$ of the perturbed dynamic problem (4).

In Section 4 we apply the abstract theory of this chapter to a class of semilinear parabolic problems. In that context, we will utilize the smoothing effect of parabolic equations and we will obtain estimates in

$$\begin{aligned} Z_s &= \left(L_p(0, T; L_p(\Omega)) \cap W^{1,2}(0, T; D(A), L_2(\Omega)), \|\cdot\|_{L_p^s(0, T; L_p(\Omega)) \cap W_s^{1,2}(0, T; D(A), L_2(\Omega))} \right)^2, \\ E_s &= \left(L_2(0; T; L_2(\Omega)), \|\cdot\|_{L_2^s(0, T; L_2(\Omega))} \right) \times V^{s(T)} \\ &\quad \times \left(L_2(0; T; L_2(\Omega)), \|\cdot\|_{L_2^s(0, T; L_2(\Omega))} \right) \times V^{s(0)} \end{aligned}$$

to derive sensitivity and turnpike results. 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 . This regularity of the data leads to solutions with values a.e. in $D(A)$ that have a weak time derivative with values a.e. in $L_2(\Omega)$ by maximal parabolic regularity, cf. [5, Part II-1, Section 3]. In order to obtain exponential sensitivity estimates, we equip both the solution space and the space of right-hand sides by a scaled norm.

3.1. An implicit function theorem

We now present an implicit function theorem that allows for estimates in scaled norms in a very general setting. A particular feature of the following implicit function theorem is the tracking of dependencies of the neighborhoods of perturbations and solutions in scaled and unscaled norms. This allows us to formulate a criterion that renders these neighborhoods independent of T , namely

T -uniform continuity 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. [25, Corollary 3.16] and [27, Theorem 10]. We will derive a similar property for the linearized system in Section 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 3.1. *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 $G: Z \times E \rightarrow E$ defined in (11) 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

$$\delta_\varepsilon(z^1, z^2) := \frac{\|G(z^1, \varepsilon) - G(z^2, \varepsilon) - G_z(z^0, \varepsilon^0)(z^1 - z^2)\|_E}{\|z^1 - z^2\|_Z} \rightarrow 0,$$

if $z^1, z^2 \rightarrow z^0$ in Z and $\varepsilon \rightarrow \varepsilon^0$ in E .

- iii) It holds that

$$\delta_\varepsilon^s(z^1, z^0) := \frac{\|G(z^1, \varepsilon) - G(z^0, \varepsilon) - G_z(z^0, \varepsilon^0)(z^1 - z^0)\|_{E_s}}{\|z^1 - z^0\|_{Z_s}} \rightarrow 0,$$

if $z^1 \rightarrow z^0$ in Z and $\varepsilon \rightarrow \varepsilon^0$ in E .

Then there is $r_E, r_Z \geq 0$, such that for every ε 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

$$\|z^*(\varepsilon) - z^0\|_{Z_s} \leq c\|\varepsilon - \varepsilon^0\|_{E_s}. \quad (12)$$

Moreover we have the following T -uniformity:

- If the convergence of 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 (12) is independent of T .

Proof. Throughout this proof, we denote $B_{r_Z}^Z(z^0) := \{z \in Z \mid \|z - z^0\|_Z \leq r_Z\}$ and analogously $B_{r_E}^E(\varepsilon^0) := \{\varepsilon \in E \mid \|\varepsilon - \varepsilon^0\|_E \leq r_E\}$. 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

$$\delta z^{k+1} = -G_z(z^0, \varepsilon^0)^{-1} (G(z^{k+1}, \varepsilon) - G(z^k, \varepsilon) - G_z(z^0, \varepsilon^0)(z^{k+1} - z^k))$$

we have with ii) that

$$\|\delta z^{k+1}\|_Z = \|G_z(z^0, \varepsilon^0)^{-1}\|_{L(E, Z)} \delta_\varepsilon(z^{k+1}, z^k) \|\delta z^k\|_Z \quad (13)$$

where $\delta_\varepsilon(z^{k+1}, z^k) \rightarrow 0$, if $z^{k+1}, z^k \rightarrow z^0$ in Z and $\varepsilon \rightarrow \varepsilon^0$ in E . We now choose a neighborhood $B_{r_Z}^Z(z^0) \times B_{r_E}^E(\varepsilon^0)$ such that $\delta_\varepsilon(z_1, z_2) \leq \frac{1}{2\|G_z(z^0, \varepsilon^0)^{-1}\|_{L(E, Z)}}$ for all $\varepsilon \in B_{r_E}^E(\varepsilon^0)$ and $z_1, z_2 \in B_{r_Z}^Z(z^0)$. 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 $\varepsilon \in B_{r_E}^E(\varepsilon^0)$. Thus, we get

$$\|\delta z^{k+1}\|_Z \leq \|G_z(z^0, \varepsilon^0)^{-1}\|_{L(E, Z)} \delta_\varepsilon(z^{k+1}, z^k) \|\delta z^k\|_Z \leq \left(\frac{1}{2}\right)^k \|\delta z^0\|_Z. \quad (14)$$

Hence,

$$\|z^{k+1} - z^0\|_Z \leq \sum_{i=0}^k \|\delta z^i\|_Z \leq \frac{1}{1 - \frac{1}{2}} \|\delta z^0\|_Z \leq r_Z \quad (15)$$

and hence inductively, $z^k \in B_{r_Z}^Z(z^0)$ for all $k \in \mathbb{N}$. Thus, by completeness of Z , the iteration $z^k = z^0 + \sum_{i=0}^k \delta z^i$ converges to an element $z^* \in B_{r_Z}^Z(z^0)$ and as $\delta_\varepsilon(z^*, z^k) \leq \frac{1}{2\|G_z(z^0, \varepsilon^0)^{-1}\|_{L(E, Z)}}$ we get

$$\|G_z(z^0, \varepsilon^0)^{-1}(G(z^*, \varepsilon) - G(z^k, \varepsilon) - G_z(z^0, \varepsilon^0)(z^* - z^k))\|_Z \leq \frac{1}{2}\|z^* - z^k\|_Z.$$

Hence, by the reverse triangle inequality, we get

$$\|G_z(z^0, \varepsilon^0)^{-1}G(z^*, \varepsilon)\|_Z \leq \|\delta z^k + z^* - z^k\|_Z + \frac{1}{2}\|z^* - z^k\|_Z \rightarrow 0$$

for $k \rightarrow \infty$ and thus $G(z^*, \varepsilon) = 0$. To obtain an estimate in the scaled norms, we compute

$$\|z^* - z^0\|_{Z_s} \leq \|z^* - z^1\|_{Z_s} + \|\delta z^0\|_{Z_s}.$$

We further estimate with $z^1 = z^0 + \delta z^0$, $\delta z^0 = -G_z(z^0, \varepsilon^0)^{-1}G(z^0, \varepsilon)$, by *iii*) and $z^* \in B_{r_Z}^Z(z^0)$ after possibly further decreasing r_Z and r_E such that $\delta_\varepsilon^s(z^*, z^0) \leq \frac{1}{2\|G_z(z^0, \varepsilon^0)^{-1}\|_{L(E_s, Z_s)}}$ 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. \square

We have two particular applications of Theorem 3.1 in mind. First, to derive a turnpike result, we set $z^0 = (\bar{x}, \bar{\lambda})$ solving the static extremal equations (7), $\varepsilon^0 = (0, \lambda, 0, \bar{x} - x_0)$, 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 (3), $\varepsilon^0 = (0, 0, 0, 0)$, 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) in scaled norms with scaling function $s(t) = e^{-\mu t}$.

Remark 3.2. Due to its generality, Theorem 3.1 can also be applied to general evolution equations, i.e., hyperbolic equations or alternatively to elliptic problems, where the scaling could act in space. For the latter one can prove an exponential decay property of the influence of right-hand sides in space, a well-known property for elliptic equations, without knowledge of the Greens function.

A crucial point in the proof of the implicit function theorem, i.e., Theorem 3.1, 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, we need 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. We rigorously prove this property for parabolic problems in Section 4.

3.2. Superposition operators and T -uniform continuity

In order to rigorously verify assumptions ii)-iii) in Theorem 3.1, 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 [53, Section 4.3.3] for a short introduction and [4, 18] 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 [53, Section 4.3.1] and [18, Section 2].

Definition 3.3 (Superposition operator). Let W_1 and W_2 be real valued Banach spaces. Consider a mapping $f: W_1 \rightarrow W_2$. Then the mapping Φ defined by

$$\Phi(x)(s) = f(x(s)) \quad \text{for } s \in S$$

assigns to an (abstract) function $x: S \rightarrow W_1$ a new (abstract) function $z: S \rightarrow W_2$ via the relation $z(s) = f(x(s))$ for $s \in S$ and is called an (*abstract*) *Nemytskij* operator or (*abstract*) *superposition operator*.

The image of a superposition operator in L_p -spaces can be characterized under growth and boundedness conditions.

Proposition 3.4. Let W_1 and W_2 be real valued Banach spaces. Let $f: W_1 \rightarrow W_2$ be continuous. For $1 \leq p, q < \infty$ let

$$\|f(w)\|_{W_2} \leq c_1 + c_2 \|w\|_{W_1}^{\frac{p}{q}} \quad \forall w \in W_1 \tag{16}$$

for constants $c_1 \in \mathbb{R}$ and $c_2 \geq 0$. Then the corresponding superposition operator maps $L_p(S; W_1)$ into $L_q(S; W_2)$.

Proof. See [18, Theorem 1]. \square

The following proposition shows that if a superposition operator maps one L_p -space into another, continuity can be derived immediately.

Proposition 3.5 (Continuity of superposition operators). *Let W_1 and W_2 be real valued Banach spaces. Let $f: W_1 \rightarrow W_2$ 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 f is locally Lipschitz, then it is continuous as a map from $L_\infty(S; W_1)$ to $L_\infty(S; W_2)$.*

Proof. For the first part, see [18, Theorem 4]. For the case with $p = q = \infty$, see [18, Theorem 1, Theorem 5] or [53, Lemma 4.11]. \square

We note that in the case $p = q = \infty$, a growth bound of the form (3.5) is not needed. In particular, the property that the superposition Φ operator maps $L_\infty(S; W_1)$ to $L_\infty(S; W_2)$ can be concluded by local Lipschitz continuity of the underlying function f , cf. [18, Theorem 1, Theorem 5] or [53, Lemma 4.11].

Differentiability of superposition operators plays a key role in applying the implicit function theorem. The following result obtained in [18, Theorem 7] gives sufficient conditions for Fréchet differentiability.

Proposition 3.6 (Differentiability of superposition operators). *Let $1 \leq q < p < \infty$. Assume that $f: W_1 \rightarrow W_2$ is continuously Fréchet differentiable. Moreover, let the superposition operator defined by*

$$\Psi(x)(s) = f'(x(s)) \quad \text{for } s \in S$$

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 f is continuously Fréchet differentiable and the Fréchet derivative

$$\Phi': L_p(S; W_1) \rightarrow L(L_p(S; W_1), L_q(S; W_2))$$

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 3.6 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 [18, Section 3.1] and [4, Theorem 3.12].

We briefly illustrate this concept by means of a particular example of a polynomial nonlinearity.

Example 3.7. Consider $W_1 = W_2 = \mathbb{R}$ and $S = \Omega \subset \mathbb{R}^n$ bounded with $n \in \mathbb{N}$. Then the nonlinear function $f: \mathbb{R} \rightarrow \mathbb{R}$, $f(w) = w^d$, $d \in \mathbb{N}$, defines a superposition operator $\underline{\Phi}$ via the relation

$$\underline{\Phi}(x)(\omega) = f(x(\omega)) = x(\omega)^d \quad \text{for } \omega \in \Omega$$

for $x: \Omega \rightarrow \mathbb{R}$. With Proposition 3.4 we have for all $1 \leq p_1 < \infty$ that $\underline{\Phi}: L_{dp_1}(\Omega) \rightarrow L_{p_1}(\Omega)$ and by Proposition 3.5, this mapping is continuous. Further, as $f'(w) = dw^{d-1}$, we can define a continuous superposition operator $\underline{\Psi}: L_{(d-1)p_2}(\Omega) \rightarrow L_{p_2}(\Omega)$ for all $1 \leq p_2 < \infty$ corresponding to

the derivative $f'(w)$. Consequently we can set $(d-1)r = p_1$ in Proposition 3.6, i.e., $p_2 = \frac{p_1}{d}$ which yields $\underline{\Phi}$ to be continuously Fréchet differentiable as a mapping from $L_{dp_1}(\Omega)$ to $L_{p_1}(\Omega)$ and $\underline{\Phi}' = \Psi$ in the sense of Proposition 3.6. Consider now $1 \leq p_3 < \infty$, $T > 0$, and the nonlinear function $\underline{\Phi}: L_{dp_3}(\Omega) \rightarrow L_{p_3}(\Omega)$ defined above. Setting $W_1 = L_{dp_3}(\Omega)$, $W_2 = L_{p_3}(\Omega)$, and $S = [0, T]$ in Definition 3.3, we define a second superposition operator Φ for $x: [0, T] \rightarrow L_{dp}(\Omega)$ via the relation

$$\Phi(x)(t) = \underline{\Phi}(x(t)) \quad \text{for } t \in [0, T].$$

Proceeding analogously as before, we obtain that Φ is continuous and differentiable as operator from $L_{dp_3}(0, T; L_{dp_1}(\Omega))$ to $L_{p_3}(0, T; L_{p_1}(\Omega))$ for $1 \leq p_1, p_3 < \infty$. We thus obtained from a scalar nonlinear function a nonlinear mapping from one space of abstract functions into another one by applying Definition 3.3 twice.

In order to render the radii r_Z and r_E and the estimate (12) 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 2 with norms defined in (8).

Definition 3.8 (T -uniform continuity). Let W_1, W_2 be real-valued Banach spaces. We say that an operator $\Psi: L_p(0, T; W_1) \rightarrow L_q(0, T; W_2)$ 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 3.9. *If the constants c_1 and c_2 of the growth condition (16) are 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. [4, Chapter 3] and [18]. \square

Example 3.10 (Example 3.7 revisited). We briefly illustrate the previous lemma at the example $f(w) = w^d$. In that case it is clear that the growth condition (16), i.e.,

$$|f(w)| \leq c_1 + c_2|w|^{\frac{p}{q}}$$

holds with $p = 3q$, $c_1 = 0$ and $c_2 = 1$, i.e., f induces a T -uniformly continuous superposition operator from $L_{dq}(0, T; L_{dp}(\Omega))$ to $L_q(0, T; L_p(\Omega))$. 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 3.1 can be shown to be T -uniform.

Lemma 3.11. *Let W_1 and W_2 be Banach spaces, $1 \leq p \leq \infty$ and let $\Phi: L_p(0, T; W_1) \rightarrow L_q(0, T; W_2)$ have a T -uniformly continuous Fréchet derivative Φ' . Then,*

$$\frac{\|\Phi(x^1) - \Phi(x^2) - \Phi'(x^0)(x^1 - x^2)\|_{L_q(0, T; W_2)}}{\|x^1 - x^2\|_{L_p(0, T; W_1)}} \rightarrow 0$$

uniformly in T if $x^1, x^2 \rightarrow x^0$ in $L_p(0, T; W_1)$. Moreover,

$$\frac{\|\Phi(x^0 + \delta x) - \Phi(x^0) - \Phi'(x^0)\delta x\|_{L_q^s(0, T; W_2)}}{\|\delta x\|_{L_p^s(0, T; W_1)}} \rightarrow 0$$

uniformly in T if $\delta x \rightarrow 0$ in $L_p(0, T; W_1)$.

Proof. We compute with the fundamental theorem of calculus, cf. [31, p.51], that

$$\begin{aligned} & \|\Phi(x^1) - \Phi(x^2) - \Phi'(x^0)(x^1 - x^2)\|_{L_q(0, T; W_2)} \\ &= \left\| \int_0^1 \Phi'(x^2 + \theta(x^1 - x^2)) - \Phi'(x^0) d\theta(x^1 - x^2) \right\|_{L_q(0, T; W_2)} \\ &\leq \left(\sup_{\theta \in [0, 1]} \|\Phi'(x^2 + \theta(x^1 - x^2)) - \Phi'(x^2)\|_{L(L_p(0, T; W_1), L_q(0, T; W_2))} \right. \\ &\quad \left. + \|(\Phi'(x^0) - \Phi'(x^2))\|_{L(L_p(0, T; W_1), L_q(0, T; W_2))} \right) \|x^1 - x^2\|_{L_p(0, T; W_1)}. \end{aligned}$$

The first claim follows by T -uniform continuity of Φ' . For the second claim in scaled norms with scaling function s , we compute analogously

$$\begin{aligned} & \|\Phi(x^0 + \delta x) - \Phi(x^0) - \Phi'(x^0)\delta x\|_{L_q^s(0, T; W_2)} \\ &= \|s(\Phi(x^0 + \delta x) - \Phi(x^0) - \Phi'(x^0)\delta x)\|_{L_q(0, T; W_2)} \\ &= \sup_{\theta \in [0, 1]} \|\Phi'(x^0) - \Phi'(x^0 + \theta\delta x)\|_{L(L_p(0, T; W_1), L_q(0, T; W_2))} \|s\delta x\|_{L_p(0, T; W_1)} \\ &\leq \sup_{\theta \in [0, 1]} \|\Phi'(x^0) - \Phi'(x^0 + \theta\delta x)\|_{L(L_p(0, T; W_1), L_q(0, T; W_2))} \|\delta x\|_{L_p^s(0, T; W_1)}, \end{aligned}$$

which concludes the proof. \square

Hence, it turns out that whenever the superposition operators are differentiable with T -uniformly continuous derivative, the uniform convergence needed in Theorem 3.1 ii) and iii) to obtain T -uniform neighborhoods holds true. The last thing to prove to apply the implicit function theorem is the T -uniform estimate on the solution operator to 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 semilinear problems that provides flexibility in the norms to match the functional analytic framework where one established T -uniform continuity.

4. A T -INDEPENDENT BOUND FOR THE EXTREMAL EQUATIONS' SOLUTION OPERATOR FOR SEMILINEAR PARABOLIC PROBLEMS

In this part we will rigorously verify the assumptions of the abstract implicit function theorem, i.e., Theorem 3.1, for a class of semilinear heat equations. The analysis in this part is heavily motivated by the approach taken in [44], 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 [44]. To this end, we assume that the PDE of interest is semilinear parabolic, i.e., $\bar{A}(x) = \mathcal{A}x - f(x)$ with $f'(x) \geq c_0$ for $c_0 \in \mathbb{R}$ and that $B \in L(L_2(\Omega_c), L_2(\Omega))$ for a control domain $\Omega_c \subset \Omega$ which includes, e.g., the case of distributed control. The operator $-\mathcal{A}$ is considered to be an elliptic differential operator of second order, i.e.,

$$\mathcal{A}x := \sum_{i,j=0}^n D_i(a_{ij}D_jx), \quad (17)$$

where $k \geq 0$, $a_{ij} = a_{ji} \in C(\bar{\Omega}, \mathbb{R})$ and $a_{i,j}(\omega)v \cdot v > 0$ for all $\omega \in \Omega$ and $v \in \mathbb{R}^n$. By $\frac{\partial x}{\partial \nu_{\mathcal{A}}}(t, s) = \sum_{i,j=0}^n a_{ij}(s)D_jx(t, s)\nu_i(s)$ 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

$$D(\mathcal{A}) = \{v \in C^2(\Omega) \mid v = 0 \text{ on } \partial\Omega\} \quad \text{or} \quad D(\mathcal{A}) = \{v \in C^2(\Omega) \mid \frac{\partial v}{\partial \nu_{\mathcal{A}}} = 0 \text{ on } \partial\Omega\} \quad (18)$$

for either homogeneous Dirichlet or homogeneous Neumann boundary conditions. We assume w.l.o.g. that there is $\alpha > 0$ such that

$$- \int_{\Omega} \mathcal{A}vv \, d\omega \geq \frac{\alpha}{2} \|v\|_{H^1(\Omega)}^2 \quad (19)$$

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 $(\mathcal{A} - kI)x$ for any $k > 0$ by $\bar{A}(x) = \mathcal{A}x - f(x) = (\mathcal{A} - kI)x + kx - f(x)$ and redefine $f(x) := f(x) - kx$.

It can be shown that for all $1 \leq l < \infty$ the closure A_l of \mathcal{A} in $L_l(\Omega)$ generates an analytic semigroup in $L_l(\Omega)$ and we abbreviate $A = A_2$. For $1 < l < \infty$, the domain is given by $D(A_l) = \{v \in W^{2,l}(\Omega) \mid v = 0 \text{ on } \partial\Omega\}$ or $D(A_l) = \{v \in W^{2,l}(\Omega) \mid \frac{\partial v}{\partial \nu_{\mathcal{A}}} = 0 \text{ on } \partial\Omega\}$, depending on the choice in (18). Additionally, the spectrum of A_l does not depend on $1 \leq l < \infty$. For details we refer to [45] and [44, Section 3].

Correspondingly, depending on the choice of boundary conditions above, we will set $V = H^1(\Omega)$ in the case of homogeneous Neumann boundary conditions or $V = H_0^1(\Omega)$ in the case of homogeneous Dirichlet boundary conditions.

Assumption 4.1. *Assume that $(x, \lambda) \in L_{\infty}(0, T; L_{\infty}(\Omega))$ for any solution (x, u, λ) of (2) and $(\bar{x}, \bar{\lambda}) \in L_{\infty}(\Omega)$ for any solution $(\bar{x}, \bar{u}, \bar{\lambda})$ of (6).*

As stated before, in order to render this assumption satisfied, one usually assumes that the data of (1) and (5) are sufficiently smooth and that the nonlinearity is monotone. Under these assumptions, boundedness of solutions in time and space for parabolic problems was proven in [44]. Similarly, for semilinear elliptic equations, a proof for continuity of solutions can be found in [9]. The interested reader is also referred to the respective parts in the monograph [53]. Under Assumption 4.1 we can conclude by smoothness of $J(x, u)$, $\bar{J}(x)$, and $f(x)$ with Proposition 3.5 that the corresponding superposition operators are continuous from $L_{\infty}(\Omega)$ to $L_{\infty}(\Omega)$ and from $L_{\infty}(0, T; L_{\infty}(\Omega))$ to $L_{\infty}(0, T; L_{\infty}(\Omega))$, respectively. We further introduce a square root property for the second derivative of the reduced Lagrange function with respect to the state.

Lemma 4.2. Let $(x^0, \lambda^0) \in L_\infty(0, T; L_\infty(\Omega))^2$ and $(L_r)_{xx}(x, \lambda) = J_{xx}(x^0) + f''(x^0) * \lambda^0 \in L_\infty(0, T; L_\infty(\Omega))$ induce a nonnegative multiplication operator, i.e., for $v: [0, T] \times \Omega \rightarrow \mathbb{R}$

$$((L_r)_{xx}(x^0, \lambda^0)v)(t, \omega) := (L_r)_{xx}(x^0, \lambda^0)(t, \omega) \cdot v(t, \omega)$$

and $(L_r)_{xx}(x^0, \lambda^0)(t, \omega) \geq 0$ for a.e. $t \in [0, T]$ and $\omega \in \Omega$. Then, there is a multiplication operator $C \in L(L_{p_1}(0, T; L_{p_2}(\Omega)))$ for all $1 \leq p_1, p_2 \leq \infty$ defined by

$$(Cv)(t, \omega) := \sqrt{(L_r)_{xx}(x^0, \lambda^0)(t, \omega)} \cdot v(t, \omega) \quad (20)$$

such that

$$(L_r)_{xx}(x^0, \lambda^0) = C^2.$$

Remark 4.3. The assumption of $(L_r)_{xx}(x^0, \lambda^0)$ inducing a multiplication operator is satisfied if the cost functional is of the form

$$J(x, u) = \frac{1}{2} \int_0^T \|x(t) - x_d(t)\|_{L_2(\Omega_o)}^2 + \frac{\alpha}{2} \|u(t)\|_U^2 dt$$

for $\Omega_o \subset \Omega$ and if the nonlinearity is given by $f(x) = x^3$. In that case,

$$L_{xx}(x^0, \lambda^0) = \chi_{\Omega_o} + 6x^0 \lambda^0,$$

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 $L_\infty(0, T; L_\infty(\Omega))$, which, for $(x^0, \lambda^0) = (\bar{x}, \bar{\lambda})$ or $(x^0, \lambda^0) = (\tilde{x}, \tilde{\lambda})$ can be verified by imposing smallness conditions on the data of the underlying steady-state or dynamic OCP, cf. Example 5.5. The assumption that L_{xx} is positive semidefinite, was also mad in [52, Theorem 1] and [51, Theorem 1]. As seen in this example and as stated in [52, Remark 6] this assumption is not standard. 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 in the sense of eq. (20), which itself is necessary to obtain stability results for the linearized system, cf. the proof of Theorem 4.8.

Under Assumption 4.1, we have that $L_{xx}(x, \lambda) \in L(L_{p_1}(0, T; L_{p_2}(\Omega)))$ for $x, \lambda \in L_\infty(0, T, L_\infty(\Omega))$ such that $C \in L(L_{p_1}(0, T; L_{p_2}(\Omega)))$ as defined in (20) for all $1 \leq p_1, p_2 \leq \infty$.

In order to apply Theorem 3.1, we will show a bound on the inverse of

$$L_r''(x^0, \lambda^0): (L_{p_1}(0, T; L_{p_2}(\Omega)) \cap W^{1,2}(0, T, D(A), L_2(\Omega)))^2 \rightarrow (L_2(0, T; L_2(\Omega)) \times V)^2,$$

where $2 \leq p_1, p_2 \leq \infty$, $\frac{n}{2}(\frac{1}{2} - \frac{1}{p_2}) < \frac{1}{p_1} + \frac{1}{2}$ and (x^0, λ^0) either solves the static system (7) or the dynamic system (3). We aim to choose p_1, p_2 as large as possible to render a wide range of nonlinearities continuous and differentiable with T -uniformly continuous derivative in these spaces.

To derive an operator norm we consider the linear system

$$\underbrace{\begin{pmatrix} J_{xx}(x^0) + f''(x^0)\lambda^0 & -\frac{d}{dt} - \mathcal{A}^* + f'(x^0) \\ 0 & E_T \\ \frac{d}{dt} - \mathcal{A} + f'(x^0) & -BQ^{-1}B^* \\ E_0 & 0 \end{pmatrix}}_{L'_r(x^0, \lambda^0)} \begin{pmatrix} \delta x \\ \delta \lambda \end{pmatrix} = \begin{pmatrix} l_1 \\ \delta \lambda_T \\ l_2 \\ \delta x_0 \end{pmatrix} \quad (21)$$

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 f , the terms $J_x(x^0)$, $f'(x^0)$, and $f''(x^0)$ are in $L_\infty(0, T; L_\infty(\Omega))$ because of Proposition 3.5 and hence can be interpreted as pointwise multiplications. With slight abuse of notation, we denote by the same symbol the corresponding superposition operator. We now aim to estimate $(\delta x, \delta \lambda)$ by means of the right-hand side in appropriate norms. To this end, we make the following stabilizability assumptions.

Assumption 4.4. *Let $c_0 \in \mathbb{R}$ be such that $c_0 \leq f'(w)$ for all $w \in \mathbb{R}$ and let $\bar{C} \in L(L_p(\Omega), L_p(\Omega))$ for all $2 \leq p \leq \infty$ be such that $\|\bar{C}v\|_{L_2(0, T; L_2(\Omega))} \leq \|Cv\|_{L_2(0, T; L_2(\Omega))}$ for all $v \in L_2(0, T; L_2(\Omega))$, where C is defined in (20). Additionally assume:*

- i) $(\mathcal{A} - c_0I, \bar{B})$ is exponentially stabilizable in the sense that for all $1 \leq p \leq \infty$ there is $\bar{K}_{\bar{B}} \in L(L_p(\Omega), L_p(\Omega_c))$ satisfying $\bar{B}\bar{K}_{\bar{B}} \in L(L_p(\Omega))$ such that $\mathcal{A} - c_0I + \bar{B}\bar{K}_{\bar{B}}$ satisfies (19).
- ii) $(\mathcal{A} - c_0I, \bar{C})$ is exponentially stabilizable in the sense that for all $1 \leq p \leq \infty$ there is $\bar{K}_{\bar{C}} \in L(L_p(\Omega))$ such that $\mathcal{A} - c_0I + \bar{C}^*\bar{K}_{\bar{C}}$ satisfies (19).
- iii) $(\mathcal{A} - f'(x^0), C)$ and $(\mathcal{A} - f'(x^0), B)$ are exponentially stabilizable for $x^0 = x$ with (x, λ) solving (3) and $x^0 = \bar{x}$ for $(\bar{x}, \bar{\lambda})$ solving (7) 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

$$\begin{aligned} & - \int_0^T \int_\Omega (\mathcal{A} - f'(x^0) + BK_B) vv \, d\omega \, dt \geq \alpha \|v\|_{L_2(0, T; V)}^2 \\ & - \int_0^T \int_\Omega (\mathcal{A} - f'(x^0) + C^*K_C) vv \, d\omega \, dt \geq \alpha \|v\|_{L_2(0, T; V)}^2 \end{aligned}$$

for all $v \in L_2(0, T; D(\mathcal{A}))$.

- 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 $1 \leq p \leq \infty$.

We briefly comment on these assumptions.

Remark 4.5. The first two assumptions ensure that the linearized system is stabilizable and detectable and that the closed-loop operators generate a strongly continuous exponentially stable analytic semigroup in $L_p(\Omega)$ for all $1 \leq p < \infty$. The third assumption allows us to deduce the $W([0, T])$ -bound analogously to [25, Corollary 3.16]. This stabilizability assumption was introduced for non-autonomous parabolic problems in [25, Definition 3.6]. The last assumption ensures that the coefficients in the linearized systems 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. [39, Theorem 0.2] under a smallness assumption on the reference state in case of a tracking type cost functional.

In the following we denote by \mathcal{A}_{cl} an operator of the form (17) satisfying (19) endowed with a domain defined in (18). By $(\mathcal{T}_{cl}^l(t))_{t \geq 0}$ we will denote the analytic semigroup on $L_l(\Omega)$ that is generated by the closure of \mathcal{A}_{cl} in $L_l(\Omega)$ for $1 \leq l < \infty$.

Proposition 4.6. *There is $\mu_0 > 0$ and a constant $c > 0$ independent of t , such that*

$$\|\mathcal{T}_{cl}^l(t)\psi_0\|_{L_q(\Omega)} \leq c \frac{e^{-\mu_0 t}}{t^{\frac{n}{2}(\frac{1}{l} - \frac{1}{q})}} \|\psi_0\|_{L_l(\Omega)} \quad \forall t > 0$$

for all $\psi_0 \in L_l(\Omega)$ and $1 \leq l \leq q \leq \infty$ with $l < \infty$.

Proof. See [45, Lemma 1] or [2, Proposition 12.5]. \square

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 theorem. As a consequence, 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 3.4 independent of T , cf. Example 3.10. We will again pick up this issue in Remark 5.6.

A central tool in the following will be a convolution estimate, similar to the proof of [27, Lemma 8]. The proof is motivated by the approach of [44, Proposition 3.1].

Theorem 4.7. *Let Assumption 4.4 hold and let $(\delta x, \delta \lambda) \in W([0, T])^2$ solve (21). Then, for all $p_1, p_2 \geq 2$ satisfying $\frac{n}{2}(\frac{1}{2} - \frac{1}{p_2}) < \frac{1}{p_1} + \frac{1}{2}$ with $p_2 < \frac{2n}{n-2}$, there is a constant $c > 0$ independent of T , such that*

$$\begin{aligned} & \|(\delta x, \delta \lambda)\|_{W^{1,2}(0, T, D(A), L_2(\Omega))^2} + \|(\delta x, \delta \lambda)\|_{L_{p_1}(0, T; L_{p_2}(\Omega))^2} \\ & \leq c(\|C\delta x\|_{L_2(0, T; L_2(\Omega))} + \|B^*\delta \lambda\|_{L_2(0, T; L_2(\Omega_c))} + \|r\|_{(L_2(0, T; L_2(\Omega)) \times V)^2}), \end{aligned}$$

where $r := (l_1, \delta \lambda_T, l_2, \delta x_0)$.

Proof. We will first show the $W^{1,2}(0, T, D(A), L_2(\Omega))$ -estimate. To this end, we consider the state equation of eq. (21), i.e.,

$$\delta x' + (-\mathcal{A} + f'(x^0))\delta x - BQ^{-1}B^*\delta \lambda = l_2$$

with initial condition $\delta x(0) = \delta x_0$. Adding the stabilizing feedback C^*K_C from Assumption 4.4 iii), we obtain

$$\delta x' + (-\mathcal{A} + f'(x^0) - C^*K_C)\delta x = BQ^{-1}B^*\delta \lambda + l_2 - C^*K_C\delta x.$$

and testing the equation with δx , using the coercivity of Assumption 4.4 iii) we get

$$\|\delta x\|_{L_2(0, T; V)} \leq c(\|\delta x_0\|_V + \|B^*\delta \lambda\|_{L_2(0, T; L_2(\Omega))} + \|l_2\|_{L_2(0, T; L_2(\Omega))} + \|C\delta x\|_{L_2(0, T; L_2(\Omega))}). \quad (22)$$

As \mathcal{A} generates an exponentially stable analytic semigroup in $L_2(\Omega)$ by applying the maximal regularity result [5, Part II-1, Theorem 3.1] to

$$\delta x' - \mathcal{A}\delta x = -f'(x^0)\delta x + BQ^{-1}B^*\delta \lambda + l_2$$

we obtain

$$\begin{aligned} & \|\delta x'\|_{L_2(0,T;L_2(\Omega))} + \|\mathcal{A}\delta x\|_{L_2(0,T;L_2(\Omega))} \\ & \leq c \left(\|\delta x\|_{L_2(0,T;L_2(\Omega))} + \|B^*\delta\lambda\|_{L_2(0,T;L_2(\Omega))} + \|l_2\|_{L_2(0,T;L_2(\Omega))} \right). \end{aligned}$$

Together with (22) we conclude

$$\begin{aligned} & \|\delta x\|_{W^{1,2}(0,T,D(A),L_2(\Omega))} \\ & \leq c \left(\|l_2\|_{L_2(0,T;L_2(\Omega))} + \|B^*\delta\lambda\|_{L_2(0,T;L_2(\Omega))} + \|C\delta y\|_{L_2(0,T;L_2(\Omega))} + \|\delta x_0\|_V \right). \end{aligned}$$

Proceeding analogously for the adjoint yields the first part of the estimate. To obtain the estimate in $L_{p_1}(0,T;L_{p_2}(\Omega))$, we proceed similarly to [44, Proof of Proposition 3.1]. Let $\psi_0 \in L_l(\Omega)$, $1 \leq l < \infty$ and ψ solve the auxiliary problem

$$\psi' = (\mathcal{A} - c_0I + \bar{C}^* \bar{K}_{\bar{C}}) \psi, \quad \psi(0) = \psi_0,$$

where $\bar{K}_{\bar{C}}$ is a stabilizing feedback for $(\mathcal{A} - c_0I, \bar{C})$ in the sense of Assumption 4.4 ii). Thus, by Proposition 4.6 for all $1 \leq l \leq q \leq \infty$ with $l < \infty$ and $t > \tau \geq 0$ we have the estimate

$$\|\psi(t - \tau)\|_{L_q(\Omega)} \leq c \frac{e^{-\mu_0(t-\tau)}}{(t - \tau)^{\frac{\alpha}{2}(\frac{1}{l} - \frac{1}{q})}} \|\psi_0\|_{L_l(\Omega)}. \quad (23)$$

We compute

$$\int_{\Omega} \psi_0(\omega) \delta x(t, \omega) d\omega = \underbrace{\int_0^t \left(\frac{d}{d\tau} \int_{\Omega} \psi(t - \tau, \omega) \delta x(\tau, \omega) d\omega \right) d\tau}_{\textcircled{I}} + \underbrace{\int_{\Omega} \psi(t, \omega) \delta x_0(\omega) d\omega}_{\textcircled{II}}. \quad (24)$$

For the first part of (24) we obtain with $c_0 - f'(x^0) \leq 0$ and as \mathcal{A} is self-adjoint that

$$\begin{aligned} \textcircled{I} &= \int_0^t \left(\int_{\Omega} -\psi'(t - \tau, \omega) \delta x(\tau, \omega) + \psi(t - \tau, \omega) \delta x'(\tau, \omega) d\omega \right) d\tau \\ &= \int_0^t \left(\int_{\Omega} -\mathcal{A}\psi(t - \tau, \omega) \delta x(\tau, \omega) - \bar{C}^* K \psi(t - \tau, \omega) \delta x(\tau, \omega) + c_0 \psi(t - \tau, \omega) \delta x(\tau, \omega) \right. \\ &\quad \left. + \psi(t - \tau, \omega) \mathcal{A} \delta x(\tau, \omega) - f'(x^0) \psi(t - \tau, \omega) \delta x(\tau, \omega) + l_2(t, \omega) \psi(t - \tau, \omega) \right. \\ &\quad \left. + \bar{B}Q^{-1} \bar{B}^* \delta\lambda(\tau, \omega) \psi(t - \tau, \omega) d\omega \right) d\tau \\ &\leq \int_0^t \left(\int_{\Omega} -\bar{C}^* K \psi(t - \tau, \omega) \delta x(\tau, \omega) + \psi(t - \tau, \omega) l_2(t, \omega) + \psi(t - \tau, \omega) \bar{B}Q^{-1} \bar{B}^* \delta\lambda(\tau, \omega) d\omega \right) d\tau. \end{aligned}$$

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 (23) and setting $l = p'_2$ and $q = 2$, we obtain for the first summand of (24)

that

$$\textcircled{I} \leq c \|\psi_0\|_{L_{p'_2}(\Omega)} \int_0^t \frac{e^{-\mu_0(t-\tau)}}{(t-\tau)^{\frac{n}{2}(\frac{1}{p'_2}-\frac{1}{2}+\delta)}} \left(\|\bar{C}\delta x(\tau)\|_{L_2(\Omega)} + \|\bar{B}^*\delta\lambda(\tau)\|_{L_2(\Omega_c)} + \|l_2(\tau)\|_{L_2(\Omega)} \right) d\tau.$$

For the second part of (24) we use Hölder's inequality and (23) with $q = l = p'_2$ and obtain that

$$\textcircled{II} \leq \|\psi(t)\|_{L_{p'_2}(\Omega)} \|\delta x_0\|_{L_{p_2}(\Omega)} \leq c e^{-\mu_0 t} \|\psi_0\|_{L_{p'_2}(\Omega)} \|\delta x_0\|_{L_{p_2}(\Omega)}. \quad (25)$$

Taking the supremum over all $\psi_0 \in L_{p'_2}(\Omega)$ yields for any $t \in [0, T]$ that

$$\begin{aligned} \|\delta x(t)\|_{L_{p_2}(\Omega)} &\leq c \int_0^t \frac{e^{-\mu_0(t-\tau)}}{(t-\tau)^{\frac{n}{2}(\frac{1}{p'_2}-\frac{1}{2})}} \left(\|\bar{C}\delta x(\tau)\|_{L_2(\Omega)} + \|\bar{B}^*\delta\lambda(\tau)\|_{L_2(\Omega_c)} + \|l_2(\tau)\|_{L_2(\Omega)} \right) d\tau \\ &\quad + c e^{-\mu_0 t} \|x_0\|_{L_{p_2}(\Omega)}. \end{aligned}$$

We now integrate this inequality over time. To this end, we recall Young's convolution inequality, cf. [55, Theorem II.4.4], which states for $\frac{1}{p_1} + 1 = \frac{1}{2} + \frac{1}{h}$ that

$$\|w * g\|_{L_{p_1}(\mathbb{R})} \leq \|w\|_{L_h(\mathbb{R})} \|g\|_{L_2(\mathbb{R})}.$$

We apply this convolution inequality to

$$\begin{aligned} g(\tau) &:= \|\bar{C}\delta x(\tau)\|_{L_2(\Omega)} + \|\bar{B}^*\delta\lambda(\tau)\|_{L_2(\Omega_c)} + \|l_2(\tau)\|_{L_2(\Omega)}, \\ w(\tau) &:= \frac{e^{-\mu_0\tau}}{\tau^{\frac{n}{2}(\frac{1}{p'_2}-\frac{1}{2})}} \end{aligned}$$

for any $\tau \in [0, T]$ and extended by zero otherwise. Additionally, we require that $\frac{n}{2}(\frac{1}{p'_2} - \frac{1}{2}) < \frac{1}{h}$ and $p_1 < \infty$ to ensure $\|w\|_{L_h(\mathbb{R})} < \infty$. Then we conclude that

$$\begin{aligned} &\|\delta x\|_{L_{p_1}(0, T; L_{p_2}(\Omega))} \\ &\leq c \left(\|\bar{C}\delta x\|_{L_2(0, T; L_2(\Omega))} + \|\bar{B}^*\delta\lambda\|_{L_2(0, T; L_2(\Omega_c))} + \|l_2\|_{L_2(0, T; L_2(\Omega))} + \|e^{-\mu t}\|_{L_{p_1}(\mathbb{R})} \|\delta x_0\|_{L_{p_2}(\Omega)} \right) \\ &\leq c \left(\|C\delta x\|_{L_2(0, T; L_2(\Omega))} + \|B^*\delta\lambda\|_{L_2(0, T; L_2(\Omega_c))} + \|l_2\|_{L_2(0, T; L_2(\Omega))} + \|\delta x_0\|_V \right) \end{aligned}$$

where the last estimate follows from Assumption 4.4, i.e., $\|\bar{C}v\|_{L_2(0, T; L_2(\Omega))} \leq \|Cv\|_{L_2(0, T; L_2(\Omega))}$ for all $v \in L_2(0, T; L_2(\Omega))$ and by the classical Sobolev embedding theorem $H^1(\Omega) \hookrightarrow L_{p_2}(\Omega)$ for $p_2 < \frac{2n}{n-2}$, cf. [1, Theorem 5.4]. \square

This stability estimate can be used to derive a T -uniform estimate for the solution operators norm. The latter 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 3.1. The following theorem states the main result of this section.

Theorem 4.8. *Let $z^0 = (x^0, \lambda^0)$ be such that Assumption 4.1 and Assumption 4.4 hold. Then, for all $2 \leq p_1, p_2$ satisfying $\frac{n}{2}(\frac{1}{2} - \frac{1}{p_2}) < \frac{1}{p_1} + \frac{1}{2}$ and $p_2 < \frac{2n}{n-2}$ there is a constant $c \geq 0$ independent of*

T such that

$$\|L_r''(z^0)^{-1}\|_{L((L_2(0,T;L_2(\Omega)) \times V)^2, (L_{p_1}(0,T;L_{p_2}(\Omega)) \cap W^{1,2}(0,T,D(A),L_2(\Omega)))^2)} \leq c \quad (26)$$

Moreover for all $\mu > 0$ satisfying

$$\mu < \frac{1}{\|L_r''(z^0)^{-1}\|_{L((L_2(0,T;L_2(\Omega)) \times V)^2, (L_{p_1}(0,T;L_{p_2}(\Omega)) \cap W^{1,2}(0,T,D(A),L_2(\Omega)))^2)}} \quad (27)$$

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

$$\|L_r''(z^0)^{-1}\|_{L((L_2^s(0,T;L_2(\Omega)) \times V^{s(T)} \times L_2^s(0,T;L_2(\Omega)) \times V^{s(0)}), (L_{p_1}^s(0,T;L_{p_2}(\Omega)) \cap W_s^{1,2}(0,T,D(A),L_2(\Omega)))^2)} \leq c \quad (28)$$

for the scaling functions $s(t) = e^{-\mu t}$ or $s(t) = \frac{1}{e^{-\mu t} + e^{-\mu(T-t)}}$.

Proof. Using the bound derived in Theorem 4.7, it only remains to estimate $\|C\delta x\|_{L_2(0,T;L_2(\Omega))}^2 + \|B^*\delta\lambda\|_{L_2(0,T;L_2(\Omega_c))}^2$. This follows by a classical estimate in optimal control, i.e., by testing in (21) the adjoint equation with the state, the state equation with the adjoint, integrating by parts and subtracting, which yields

$$\begin{aligned} & \|C\delta x\|_{L_2(0,T;L_2(\Omega))}^2 + \|B^*\delta\lambda\|_{L_2(0,T;L_2(\Omega_c))}^2 \leq \\ & |\langle l_1, \delta x \rangle_{L_2(0,T;L_2(\Omega))}| + |\langle l_2, \delta\lambda \rangle_{L_2(0,T;L_2(\Omega))}| + |\langle \delta x_0, \delta\lambda(0) \rangle_{L_2(\Omega)}| + |\langle \delta\lambda_T, \delta x(T) \rangle_{L_2(\Omega)}|. \end{aligned}$$

The bound eq. (26) then follows. To prove the bound in the scaled spaces we proceed analogously to the proof of [25, Theorem 3.1 and Theorem 5.2]. Hence we define $M := L_r''(z^0)$ and set $Z := (L_{p_1}(0,T;L_{p_2}(\Omega)) \cap W^{1,2}(0,T;D(A),L_2(\Omega)))^2$ 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 $P := \begin{pmatrix} 0 & F \\ -F & 0 \\ 0 & 0 \end{pmatrix}$ 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. [32, 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 eq. (28) for $s(t) = \frac{1}{e^{-\mu t} + e^{-\mu(T-t)}}$ follows. Completely analogously we conclude eq. (28) for $s(t) = e^{-\mu t}$ with the same argumentation and $P := \begin{pmatrix} 0 & -I \\ 0 & 0 \\ I & 0 \\ 0 & 0 \end{pmatrix}$. \square

We briefly comment on the estimates of Theorem 4.7 and Theorem 4.8.

Remark 4.9. In the case of $n = 2$, the restriction for p_1, p_2 includes 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 for non-autonomous equations, cf. [3]. 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.

Remark 4.10. In our setting, the differential operator \mathcal{A} gives rise to an analytic semigroup on $L_l(\Omega)$ for $1 \leq l < \infty$. If one only has analyticity for $1 < l < \infty$, Theorem 4.7 still holds under the additional restriction that $p_2 < \infty$. This restriction stems from the fact that we can not choose test functions ψ that are subject to an analytic semigroup on $L_1(\Omega)$ which would be needed to obtain an $L_\infty(\Omega)$ -estimate.

5. EXPONENTIAL TURNPIKE AND SENSITIVITY RESULTS FOR SEMILINEAR PARABOLIC PROBLEMS

We can now combine the results of Section 3.2 regarding superposition operators and the bound on the solution operator to the linearized problem of Section 4 to apply the implicit function theorem (Theorem 3.1) 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 4.8 on $p_1 = p_2 = p$ simplify to $p < \frac{2n}{n-2}$. This choice of p represents the highest exponent of all $L_p(\Omega)$ -spaces in which $H^1(\Omega)$ is embedded into.

Theorem 5.1. *Let $f(x)$ 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. Suppose the assumptions of Theorem 4.8 hold with $p_1 = p_2 = p$, i.e., $2 \leq p < \frac{2n}{n-2}$. Then $G(z, \varepsilon) := L'_r(z) - \varepsilon$ with L'_r given in (3) satisfies the assumptions of Theorem 3.1 uniformly in T for any $\mu > 0$ satisfying (27) and setting either*

- i) $z^0 = (\bar{x}, \bar{\lambda})$ solving the steady-state problem (7) 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*
- ii) $z^0 = (x, \lambda)$ solving the dynamic problem (3) and $\varepsilon^0 = 0$ and the scaling $s(t) = e^{-\mu t}$*

in the spaces $Z = (L_p(0, T; L_p(\Omega)) \cap W^{1,2}(0, T, D(A), L_2(\Omega)))^2$ and $E = (L_2(0, T; L_2(\Omega)) \times V)^2$ endowed with the scaled norms $\|\cdot\|_{Z_s} = \|\cdot\|_{(L_p^s(0, T; L_p(\Omega)) \cap W_s^{1,2}(0, T, D(A), L_2(\Omega)))^2}$ 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 3.1 follow by T -uniform continuity of the derivative of the superposition operator via Lemma 3.11. 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 4.8. \square

This result can now be used to deduce a local turnpike result, stating that solutions of the dynamic problem eq. (2) are close to solutions of the static problem eq. (7) for the majority of the time under the assumption that initial resp. terminal values are close enough at the turnpike.

Corollary 5.2. *Let the assumptions of Theorem 5.1 hold with $2 \leq p < \frac{2n}{n-2}$. Consider (x, u, λ) solving the nonlinear dynamic problem (1) and $(\bar{x}, \bar{u}, \bar{\lambda})$ solving the nonlinear static problem (5).*

Define $(\delta x, \delta u, \delta \lambda) := (x - \bar{x}, u - \bar{u}, \lambda - \bar{\lambda})$. Then there are $r_E > 0$, $\mu > 0$ (satisfying eq. (27)) and $c \geq 0$ independent of T such that if

$$\|x_0 - \bar{x}\|_V + \|\bar{\lambda}\|_V \leq r_E$$

it holds that

$$\begin{aligned} & \left\| \frac{1}{e^{-\mu t} + e^{-\mu(T-t)}} \delta x(t) \right\|_{L_p(0,T;L_p(\Omega)) \cap W^{1,2}(0,T,D(A),L_2(\Omega))} + \left\| \frac{1}{e^{-\mu t} + e^{-\mu(T-t)}} \delta u(t) \right\|_{L_\infty(0,T;L_2(\Omega_c))} \\ & + \left\| \frac{1}{e^{-\mu t} + e^{-\mu(T-t)}} \delta \lambda(t) \right\|_{L_p(0,T;L_p(\Omega)) \cap W^{1,2}(0,T,D(A),L_2(\Omega))} \leq cr_E. \end{aligned}$$

Proof. The proof follows by Theorem 5.1 with the choice i) for the scaling function and applying Theorem 3.1. \square

Second, we can conclude a sensitivity result, which states that perturbations of the extremal equations' dynamics that are small at 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 eq. (4) are close to the solutions of the unperturbed dynamic problem eq. (2) 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 5.3. *Let the assumptions of Theorem 5.1 hold with $2 \leq p < \frac{2n}{n-2}$. Let (x, λ) solve the nonlinear extremal equations (3) and $(\tilde{x}, \tilde{\lambda})$ the perturbed extremal equations (4). Define $(\delta x, \delta \lambda) := (\tilde{x} - x, \tilde{\lambda} - \lambda)$ and $\delta u = Q^{-1}B^*\delta \lambda$. Then there are $r_E > 0$, $\mu > 0$ satisfying (27) and $c \geq 0$ 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

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

it holds that

$$\begin{aligned} & \left\| e^{-\mu t} \delta x(t) \right\|_{L_p(0,T;L_p(\Omega)) \cap W^{1,2}(0,T,D(A),L_2(\Omega))} + \left\| e^{-\mu t} \delta u(t) \right\|_{L_\infty(0,T;L_2(\Omega_c))} \\ & + \left\| e^{-\mu t} \delta \lambda(t) \right\|_{L_p(0,T;L_p(\Omega)) \cap W^{1,2}(0,T,D(A),L_2(\Omega))} \leq c\rho. \end{aligned}$$

Proof. The proof follows by Theorem 5.1 with the choice ii) for the scaling function and applying Theorem 3.1. \square

Remark 5.4. 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 4.6. Perturbations of analytic semigroups can be analyzed with the notion of A -boundedness or A -compactness, cf. [12, Chapter III].

Example 5.5. 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 a static reference $x_d = \bar{x}_d \in L_2(\Omega)$. Let the cost functional be given by $J(x, u) = \frac{1}{2} \int_0^T \|x - x_d\|_{L_2(\Omega)}^2 + \|u\|_{L_2(\Omega_c)}^2$, $d = 0$ and consider the nonlinearity $f(x) = x^3 - c_0 x$ 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. Moreover, using maximal elliptic regularity, cf. [9] 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 4.1. 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 4.2. The square root \bar{C} of this operator in the sense of (20) can thus be defined pointwise for $v: \Omega \rightarrow \mathbb{R}$ 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 eq. (20) is time-independent and we can set $\bar{C} = C$ in Assumption 4.4. The remaining parts of Assumption 4.4 can be verified with the generalized Poincaré inequality, cf. [53, Lemma 2.5], analogously to [25, Example 3.7]. Hence, Assumption 4.4 is satisfied and we can apply the turnpike result of Corollary 5.2.

In order to apply the sensitivity result of Corollary 5.3 we need to analyze

$$(L_r)_{xx}(x, \lambda) = I - x\lambda$$

where (x, λ) solves (3). In [44], the authors deduce a T -dependent bound

$$\begin{aligned} & \|x\|_{L_\infty(0,T;L_\infty(\Omega))} + \|\lambda\|_{L_\infty(0,T;L_\infty(\Omega))} \\ & \leq c(T) \left(\|x_d\|_{L_{2+\delta}(0,T;L_{2+\delta}(\Omega))} + \|u_d\|_{L_{2+\delta}(0,T;L_{2+\delta}(\Omega_c))} + \|x_0\|_{L_\infty(\Omega)} \right). \end{aligned}$$

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. [39, 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. [39, 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 4.2 and we can define the square root C via (20) 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 4.4, where $\underline{c} \leq \sqrt{1 - \underline{m}^2}$ yields for $v: (0, T) \times \Omega \rightarrow \mathbb{R}$ and a.e. $(t, \omega) \in (0, T) \times \Omega$ the estimate

$$\underline{c}^2 |v(t, \omega)| \leq |1 - \underline{m}| |v(t, \omega)| = (1 - \|\lambda x\|_{L_\infty(0,T;L_\infty(\Omega))}) |v(t, \omega)| \leq \|1 - \lambda x\|_{L_\infty(0,T;L_\infty(\Omega))} |v(t, \omega)|.$$

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 4.4 is satisfied and we can apply the sensitivity result of Corollary 5.3.

We will briefly discuss the smallness assumptions made in Example 5.5 for the application of the turnpike theorem Corollary 5.2. First, in order to obtain a square root of $(L_r)_{xx}(\bar{x}, \bar{\lambda})$, we hinge on smallness of \bar{x} and $\bar{\lambda}$ in $L_\infty(\Omega)$. Second, for the application of Corollary 5.2, we have to assume smallness of $\|x_0 - \bar{x}\|_V$ and $\|\bar{\lambda}\|_V$ where V is not necessarily embedded in $L_\infty(\Omega)$. The combination of smallness in these norms allows to conclude a turnpike result.

We briefly discuss the relation of the established turnpike result and Example 5.5 to the turnpike results of [42] and [51]. Under smallness assumptions on $\|y_0 - \bar{y}\|_{L_\infty(\Omega)}$ and $\|\bar{\lambda}\|_{L_\infty(\Omega)}$, a turnpike result in $C(0, T; L_\infty(\Omega))$ for semilinear heat equations was given in [42, Theorem 1] for space dimension $n \leq 3$. Further in [51, Theorem 1], for a Hilbert space V , $C(0, T; V)$ -estimates are concluded via smallness assumptions of the initial and terminal distance to the turnpike in V . In this case, positive semi-definiteness of $(L_r)_{xx}$ is assumed in order to obtain a square root. This was then verified a posteriori for an example by assuming smallness of the optimal steady state, similar to our approach in Example 5.5. Here, we assume in Corollary 5.2 smallness in $H^1(\Omega)$ (which does not necessarily imply smallness in $L_\infty(\Omega)$ if $n \geq 2$). However, as seen in the previous example, we additionally need smallness in $L_\infty(\Omega)$ in order to obtain a square root of $(L_r)_{xx}$. Under these assumptions, by invoking Corollary 5.2 we obtain $L_p(0, T; L_p(\Omega))$ -estimates for $2 \leq p < \frac{2n}{n-2}$ which are weaker than the $L_\infty(\Omega)$ estimates of [42]. However, we also obtain estimates in the Sobolev norm $W^{1,2}(0, T; D(A), L_2(\Omega))$ which yields turnpike of time and space derivatives. This is a novelty in this work regarding turnpike properties. Compared to [51], we obtain the same estimates, i.e., smallness in V implies turnpike in $W^{1,2}(0, T; D(A), L_2(\Omega)) \hookrightarrow C(0, T; V)$.

Remark 5.6. We briefly discuss the case of nonlinearities that are sums of monotone polynomials, e.g., $f(x) = x^3 + x^5$. 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 $f(x) = x^3 + x^5$ 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 $x \leq 1$. 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 3.10. As a remedy, one has to invoke Theorem 3.1 with the space $Z = (L_{10}(0, T; L_{10}(\Omega)) \cap L_6(0, T; L_6(\Omega)) \cap W^{1,2}(0, T; D(A), L_2(\Omega)))^2$, where the bound on the solution operator follows by Theorem 4.8 if $n = 2$. This means that the turnpike estimate in the Sobolev spaces remains unchanged. However, one obtains additionally estimates in other $L_p(0, T; L_p(\Omega))$ -spaces. Note that these estimates are not equivalent independently of T , i.e., we can not conclude a turnpike in $L_6(0, T; L_6(\Omega))$ from a turnpike in $L_{10}(0, T; L_{10}(\Omega))$ as the constant would depend on T .

6. NUMERICAL EXAMPLES

In this part, we will showcase the theoretical results from Section 4 by means of an example of a semilinear heat equation. We further present numerical results for the boundary control of a quasilinear heat equation. All numerical examples are performed with the C++-library for vector space algorithms *Spacy*¹ using the finite element library *Kaskade7* [19].

We consider $T = 10$, $\Omega = [0, 3] \times [0, 1]$, and the cost functional

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

¹<https://spacy-dev.github.io/Spacy/>

where x_d is either of the following: To illustrate the turnpike property, we will use the static reference defined by

$$x_d^{\text{stat}}(\omega) := g\left(\frac{10}{3} \left\| \omega - \begin{pmatrix} 1.5 \\ 0.5 \end{pmatrix} \right\|\right), \quad (29)$$

$$\text{where } g(s) := \begin{cases} 10e^{1-\frac{1}{1-s^2}} & s < 1 \\ 0 & \text{else.} \end{cases} \quad (30)$$

We proved in Corollary 5.3 that perturbations that occur far in the future have a negligible effect on the initial part of the optimal triple. Motivated by this, we will suggest two a priori discretization schemes tailored to an MPC context. To evaluate the performance of these schemes, we will also consider non-autonomous problems. To this end, we define, with $g(s)$ as above,

$$x_d^{\text{dyn}}(t, \omega) := g\left(\frac{10}{3} \left\| \omega - \begin{pmatrix} \omega_{1,\text{peak}}(t) \\ \omega_{2,\text{peak}}(t) \end{pmatrix} \right\|\right) \quad (31)$$

and

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

We will consider the MPC Algorithm 6.1 and evaluate the performance of different a priori time grids used in every solution of the OCP.

Algorithm 6.1 (Standard MPC Algorithm).

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

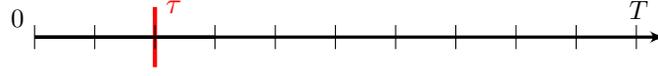
We now present two specialized a priori discretization techniques motivated by the exponential decay of perturbations deduced in Corollary 5.3. First, we will evaluate an exponential distribution of grid points. To this end, we compute vertices t_i , $i \in \{0, \dots, N-1\}$ such that

$$\int_{t_i}^{t_{i+1}} e^{-ct} dt = \frac{I}{N-1} \quad \forall i \in \{0, \dots, N-2\}, \quad (32)$$

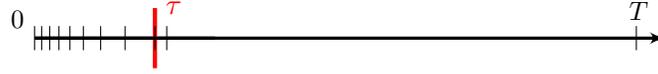
where $I := \int_0^T e^{-ct} dt = \frac{1}{c}(1 - e^{-cT})$. For all computations considered in this work, we chose $c = 1$.

As a second specialized refinement procedure we propose to use the same number of grid points on $[0, \tau]$ as on $[\tau, T]$. If $\tau \ll T$, this naturally leads to a finer mesh on the initial part. As a reference, we use a standard uniform grid. All three gridding schemes are depicted in the following for eleven grid points.

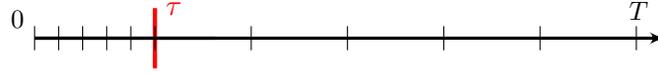
- i) Uniform:



ii) Exponential:



iii) Piecewise uniform:



6.1. Distributed control of a semilinear heat equation

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 $e \geq 0$ is a nonlinearity parameter. We first illustrate the turnpike property and hence choose a static reference trajectory, i.e., x_d^{stat} . In fig. 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 high 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 forces the solution of the state equation to 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

$$\|x(t)\|^2 \leq -c(\Omega) \int_0^t \|x(s)\|^2 + e\|x(s)\|^4 ds + \int_0^t u(s)x(s) ds.$$

We note that the depicted plots only show the turnpike property for a space-time discretized problem. For further numerical experiments that indicate that the turnpike also holds in function space, the interested reader is referred to [28].

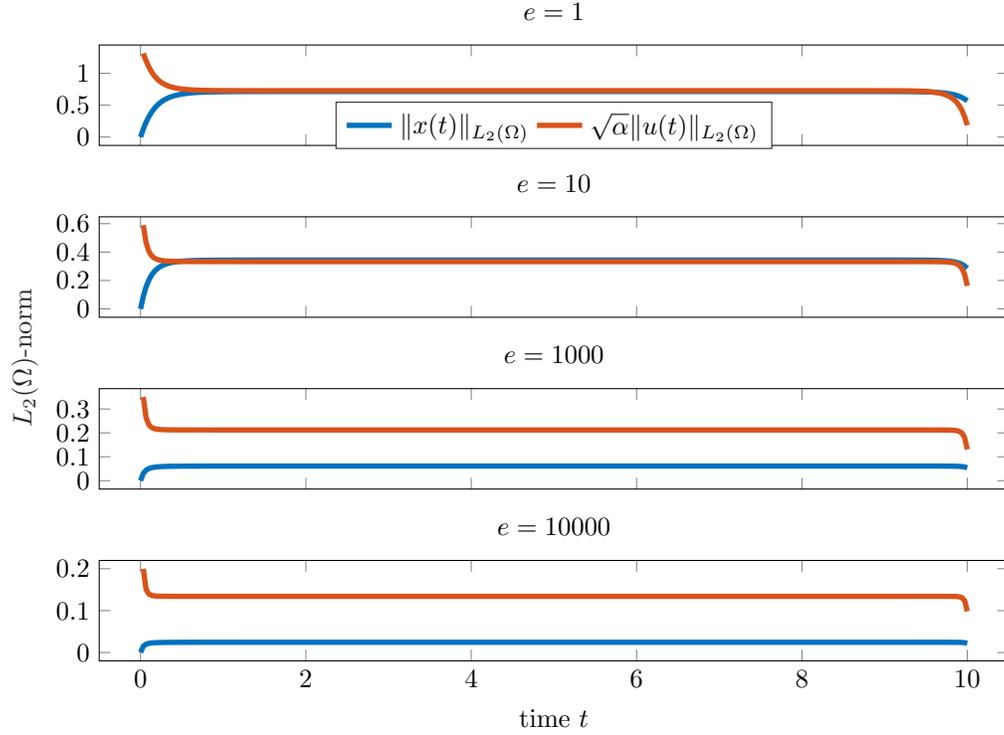


FIGURE 1. Spatial norm of open loop state and control over time with the static reference x_d^{stat} and $\alpha = 10^{-1}$ for different nonlinearity parameters.

Second, we apply 4 steps of the MPC algorithm Algorithm 6.1 to the optimal control problem above. We set the implementation horizon $\tau = 1$ and choose the dynamic reference x_d^{dyn} defined in (31). The simulation of the closed-loop trajectory emerging from the MPC feedback is again computed on three uniform refinements of the initial grid. fig. 2 shows the closed-loop cost for different a priori time discretization regimes. 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, as the grids are generated a priori.

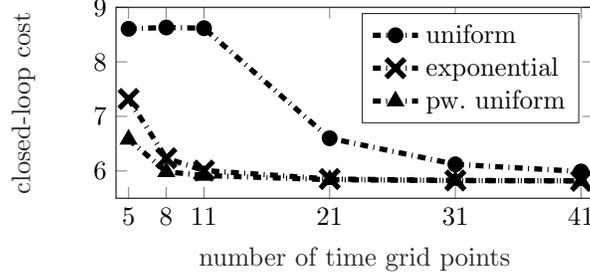


FIGURE 2. Comparison of MPC closed-loop cost for a different priori time discretization regimes with dynamic reference x_d^{dyn} and parameters $e = 1$ and $\alpha = 10^{-2}$.

6.2. Boundary control of a quasilinear equation

As a second numerical example, we consider a heat equation with heat conductivity depending on the temperature. Again, we refer the reader to [28] for various numerical examples with space-time grid adaptivity that would go beyond the scope of this work. To this end, we introduce the heat conduction tensor

$$\kappa(x)(t, \omega) := (c|x(t, \omega)|^2 + 0.1),$$

where $c \geq 0$ is a nonlinearity parameter and consider the quasilinear dynamics

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

Our theoretical results of Section 4 do not cover the case of a quasilinear equation. However, the turnpike property can be observed in fig. 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 effect of the nonlinearity forcing the state to zero. We depict the turnpike property in a norm that is motivated by the second derivative of the Lagrangian, i.e., a scaled $H^1(\Omega)$ -norm $\|v\|_{\alpha d, H^1(\Omega)} := \|v\|_{L_2(\Omega)} + \sqrt{d\alpha} \|\nabla v\|_{L_2(\Omega)}$ with $d = 0.1$.

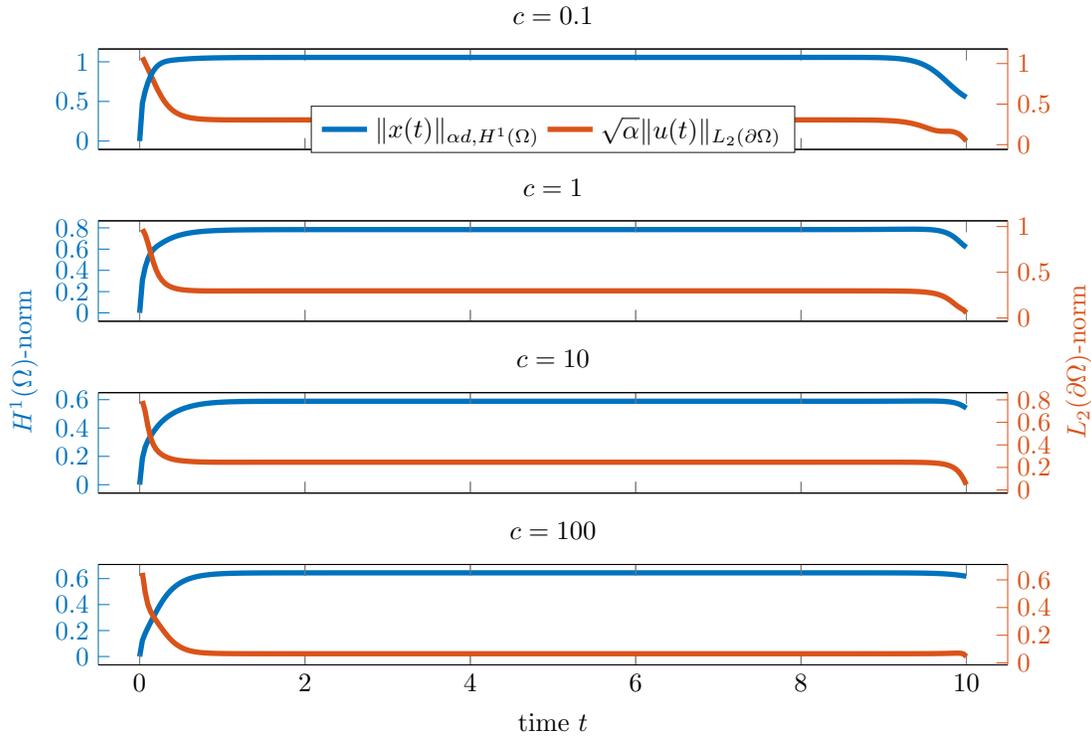


FIGURE 3. Spatial norm of open loop state and control over time with the static reference x_d^{stat} for different nonlinearity parameters and $\alpha = 10^{-1}$.

In fig. 4 we compare the closed-loop cost of different a priori time discretization regimes. Similar to the semilinear example investigated before, we observe that exponential and piecewise uniform a priori time grids outperform the conventional uniform grid.

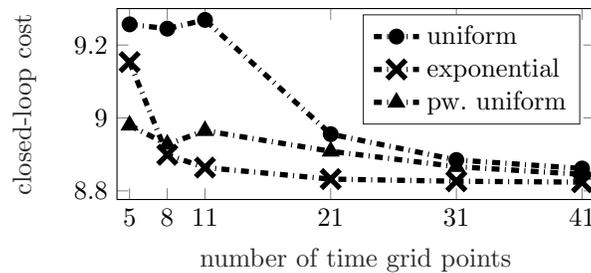


FIGURE 4. Comparison of MPC closed-loop cost for a different priori time discretization regimes with dynamic reference x_d^{dyn} for different priori time discretization regimes with parameters $c = 0.1$ and $\alpha = 10^{-2}$.

7. CONCLUSION AND OUTLOOK

We have proposed an abstract approach to analyze the sensitivity of necessary optimality conditions arising in optimal control. To this end we presented an implicit function theorem that, under T -independent continuity and invertibility assumptions on the linearization, allows to deduce estimates in scaled norms. We further applied the abstract theory to the case of optimal control of a class of semilinear heat equations to derive a turnpike property and a sensitivity result stating that perturbations of the dynamics decay exponentially in time. Finally we presented numerical examples that illustrate these theoretical results. We conclude this paper by presenting some possible directions of further research.

First, a nonlinear dependence on the control could be introduced. In our case, the quadratic dependence of the cost functional on the control allowed for a direct elimination of the control. Additionally, we did not have deal with superposition operators for the control, where improved regularity of the optimal control might be needed. In particular cases this improved regularity can be established by classical bootstrapping. If the system depends nonlinearly on the control, a standard assumption is the existence of $\alpha > 0$ such that $J_{uu}(x, u)(\delta u, \delta u) \geq \alpha \|\delta u\|_{L_2(0, T; U)}^2$ for all $\delta u \in L_2(0, T; U)$ in order to represent the optimal control by the adjoint state arising in the first order necessary conditions. This property is sometimes referred to as the strengthened Legendre-Clebsch condition, cf. [8, Chapter 6].

Second, as discussed in Remark 5.4, another natural extension would be the case of boundary control.

Finally, the abstract approach of Section 3 could be used to derive turnpike or sensitivity estimates for nonlinear hyperbolic or elliptic systems.

REFERENCES

- [1] R. A. Adams. *Sobolev Spaces*. Academic Press, Amsterdam, Boston, 1975.
- [2] H. Amann. Dual semigroups and second order linear elliptic boundary value problems. *Israel J. Math*, 45(2):225–254, 1983.
- [3] H. Amann. Maximal regularity for nonautonomous evolution equations. *Adv. Nonlinear Stud.*, 4(4):417–430, 2004.
- [4] J. Appell and P. P. Zabrejko. *Nonlinear Superposition Operators*. Cambridge Tracts in Mathematics. Cambridge University Press, 1990.
- [5] A. Bensoussan, G. Da Prato, M. C. Delfour, and S. Mitter. *Representation and control of infinite dimensional systems*. Springer Science & Business Media, 2007.
- [6] 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.
- [7] T. Breiten and L. Pfeiffer. On the turnpike property and the receding-horizon method for linear-quadratic optimal control problems. *SIAM J. Control Optim.*, 58(2):1077–1102, 2020.
- [8] J. A. E. Bryson and Y.-C. Ho. *Applied Optimal Control*. Routledge, 1975.
- [9] E. Casas. Boundary control of semilinear elliptic equations with pointwise state constraints. *SIAM J. Control Optim.*, 31(4):993–1006, 1993.
- [10] E. Casas, L. A. Fernández, and J. Yong. Optimal control of quasilinear parabolic equations. *Proc. Roy. Soc. Edinburgh Sect. A*, 125(3):545–565, 1995.
- [11] T. Damm, L. Grüne, M. Stieler, and K. Worthmann. An exponential turnpike theorem for dissipative discrete time optimal control problems. *SIAM J. Control Optim.*, 52(3):1935—1957, 2014.
- [12] K.-J. Engel and R. Nagel. *One-Parameter Semigroups for Linear Evolution Equations*. Springer Verlag New York, 2000.
- [13] C. Esteve, B. Geshkovski, D. Pighin, and E. Zuazua. Large-time asymptotics in deep learning, 2020.
- [14] C. Esteve, H. Kouhkouh, D. Pighin, and E. Zuazua. The turnpike property and the long-time behavior of the Hamilton-Jacobi equation, 2020.

- [15] T. Faulwasser, K. Flaßkamp, S. Ober-Blöbaum, and K. Worthmann. A dissipativity characterization of velocity turnpikes in optimal control problems for mechanical systems, 2020.
- [16] T. Faulwasser, L. Grüne, J.-P. Humaloja, and M. Schaller. The interval turnpike property for adjoints, 2020.
- [17] T. Faulwasser, M. Korda, C. N. Jones, and D. Bonvin. On turnpike and dissipativity properties of continuous-time optimal control problems. *Automatica J. IFAC*, 81:297–304, 2017.
- [18] H. Goldberg, W. Kampowsky, and F. Tröltzsch. On Nemytskij operators in L_p -spaces of abstract functions. *Math. Nachr.*, 155(1):127–140, 1992.
- [19] S. Götschel, A. Schiela, and M. Weiser. Kaskade 7—A flexible finite element toolbox. *Comput. Math. Appl.*, 2020.
- [20] L. Grüne. Approximation properties of receding horizon optimal control. *Jahresber. Dtsch. Math.-Ver.*, 118(1):3–37, 2016.
- [21] L. Grüne and R. Guglielmi. On the relation between turnpike properties and dissipativity for continuous time linear quadratic optimal control problems. *Math. Control Relat. Fields*, Online First, June 2020.
- [22] L. Grüne and M. A. Müller. On the relation between strict dissipativity and turnpike properties. *Systems Control Lett.*, 90:45–53, 2016.
- [23] L. Grüne and J. Pannek. *Nonlinear Model Predictive Control: Theory and Algorithms*. Springer Verlag London, 2016.
- [24] 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 Econom. and Math. Systems*, pages 195–218. Springer, Cham, 2018.
- [25] L. Grüne, M. Schaller, and A. Schiela. Sensitivity analysis of optimal control for a class of parabolic PDEs motivated by model predictive control. *SIAM J. Control Optim.*, 57(4):2753–2774, 2019.
- [26] L. Grüne, M. Schaller, and A. Schiela. Exponential sensitivity analysis for model predictive control of PDEs, 2020.
- [27] L. Grüne, M. Schaller, and A. Schiela. Exponential sensitivity and turnpike analysis for linear quadratic optimal control of general evolution equations. *J. Differential Equations*, 268(12):7311–7341, 2020.
- [28] L. Grüne, M. Schaller, and A. Schiela. Efficient MPC for parabolic PDEs with goal oriented error estimation, 2020.
- [29] M. Gugat and F. Hante. On the turnpike phenomenon for optimal boundary control problems with hyperbolic systems. *SIAM J. Control Optim.*, 57(1):264–289, 2019.
- [30] 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 Lett.*, 90:61–70, 2016.
- [31] 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.
- [32] R. Kress. *Linear integral equations*, volume 17. Springer, 2014.
- [33] O. A. Ladyzhenskaja, V. A. Solonnikov, and N. N. Ural'ceva. *Linear and quasi-linear equations of parabolic type*, volume 23. American Mathematical Soc., 1968.
- [34] G. Lance, E. Trélat, and E. Zuazua. Turnpike in optimal shape design. *IFAC-PapersOnLine*, 52(16):496–501, 2019.
- [35] H. Meinlschmidt. *Analysis and Optimal Control of Quasilinear Parabolic Evolution Equations in Divergence Form on Rough Domains*. PhD thesis, Universität Darmstadt, 2017.
- [36] S. Na, S. Shin, M. Anitescu, and V. M. Zavala. Overlapping schwarz decomposition for nonlinear optimal control, 2020.
- [37] N. S. Papageorgiou. On the optimal control of strongly nonlinear evolution equations. *J. Math. Anal. Appl.*, 164(1):83–103, 1992.
- [38] A. Pazy. *Semigroups of Linear Operators and Applications to Partial Differential Equations*. Springer Verlag New York, 1983.
- [39] D. Pighin. The turnpike property in semilinear control, 2020.
- [40] D. Pighin and N. Sakamoto. The turnpike with lack of observability, 2020.
- [41] A. Porretta and E. Zuazua. Long time versus steady state optimal control. *SIAM J. Control Optim.*, 51(6):4242–4273, 2013.
- [42] 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.
- [43] J. B. Rawlings, D. Q. Mayne, and M. Diehl. *Model Predictive Control: Theory, Computation, and Design*, volume 2. Nob Hill Publishing Madison, WI, 2017.

- [44] J. P. Raymond and H. Zidani. Hamiltonian Pontryagin's principles for control problems governed by semilinear parabolic equations. *Appl. Math. Optim.*, 39(2):143–177, 1999.
- [45] F. Rothe. Uniform bounds from bounded L_p -functionals in reaction-diffusion equations. *J. Differential Equations*, 45(2):207–233, 1982.
- [46] 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.
- [47] A. Schiela. A concise proof for existence and uniqueness of solutions of linear parabolic PDEs in the context of optimal control. *Systems Control Lett.*, 62(10):895–901, 2013.
- [48] S. Shin and V. M. Zavala. Diffusing-horizon model predictive control, 2020.
- [49] E. Trélat and C. Zhang. Integral and measure-turnpike properties for infinite-dimensional optimal control systems. *Math. Control Signals Systems*, 30(1):3, 2018.
- [50] E. Trélat, C. Zhang, and E. Zuazua. Optimal shape design for 2d heat equations in large time. *Journal of Applied Functional Analysis*, 3:255–269, 2018.
- [51] E. Trélat, C. Zhang, and E. Zuazua. Steady-state and periodic exponential turnpike property for optimal control problems in Hilbert spaces. *SIAM J. Control Optim.*, 56(2):1222–1252, 2018.
- [52] E. Trélat and E. Zuazua. The turnpike property in finite-dimensional nonlinear optimal control. *J. Differential Equations*, 258(1):81–114, 2015.
- [53] F. Tröltzsch. *Optimal control of partial differential equations: theory, methods, and applications*, volume 112. American Mathematical Soc., 2010.
- [54] M. Warma and S. Zamorano. Exponential turnpike property for fractional parabolic equations with non-zero exterior data. *ESAIM: Control Optim. Calc. Var.*, 27:1, 2021.
- [55] D. Werner. *Funktionalanalysis*. Springer Verlag Berlin Heidelberg, 2011.
- [56] J. C. Willems. Dissipative dynamical systems part i: General theory. *Arch. Ration. Mech. Anal.*, 45(5):321–351, 1972.
- [57] J. C. Willems. Dissipative dynamical systems part ii: Linear systems with quadratic supply rates. *Arch. Ration. Mech. Anal.*, 45(5):352–393, 1972.
- [58] S. Zamorano. Turnpike property for two-dimensional Navier–Stokes equations. *J. Math. Fluid Mech.*, 20(3):869–888, 2018.