Local stabilization of an unstable parabolic equation via saturated controls

Andrii Mironchenko, Christophe Prieur and Fabian Wirth

***Abstract*—We derive a saturated feedback control, which locally stabilizes a linear reaction-diffusion equation. In contrast to most other works on this topic, we do not assume the Lyapunov stability of the uncontrolled system and consider general unstable systems. Using Lyapunov methods, we provide estimates for the region of attraction for the closed-loop system, given in terms of linear and bilinear matrix inequalities. We show that our results can be used with distributed as well as scalar boundary control, and with different types of saturations. The efficiency of the proposed method is demonstrated by means of numerical simulations.**

***Index Terms*—PDE control, reaction-diffusion equation, saturated control, stabilization, attraction region.**

I. Introduction

In applications of control technology, physical inputs (like force, torque, thrust, stroke, etc.) are often limited in size [3]. If such input limitations are neglected, this may result in undesirable oscillations of the closed-loop system, lack of global stabilizability and in a dramatic reduction of the region of attraction of the closed-loop system, see e.g. [41], [45], [33] for an introduction to the nonlinear behavior induced by input limitations.

This leads to the problem of (local or global) stabilization of control systems with inputs of a norm not exceeding a prescribed value.

In this paper, we study the stabilizability of a one-dimensional linear unstable reaction-diffusion equation (also called heat equation) using a saturated control. Many results exist in the literature for the control of this class of equations, using either bounded or unbounded control operators, with or without input delays. More specifically, in [19] a backstepping approach is applied to design a boundary delayed feedback control for a heat equation (see [20] for further results using the same design methods). In [10] a stable heat partial differential equation (PDE) is controlled by means of a delayed bounded linear control operator (see also [37] for the semilinear case). For the case of unbounded control operators, see [29], [31] for the computation of delayed control to stabilize the reaction-diffusion equation.

A. Mironchenko is with Faculty of Computer Science and Mathematics, University of Passau, 94030 Passau, Germany. Email: andrii.mironchenko@uni-passau.de. Corresponding author.

C. Prieur is with Univ. Grenoble Alpes, CNRS, Grenoble INP, GIPSA-lab, 38000 Grenoble, France. Email: Christophe.Prieur@gipsa-lab.fr

F. Wirth is with Faculty of Computer Science and Mathematics, University of Passau, 94030 Passau, Germany. Email: fabian.(lastname)@uni-passau.de.

A. Mironchenko has been supported by DFG, grant Nr. MI 1886/2-1. C. Prieur has been partially supported by MIAI @ Grenoble Alpes, (ANR-19-P3IA-0003). A preliminary version of this paper was presented at the 11-th IFAC Symposium on Nonlinear Control Systems (NOLCOS 2019), see [27].

To analyze the effect of input saturations and to design saturated controllers for infinite-dimensional systems, different results and techniques are available. One of the first papers in this field is [36], where compact and bounded control operators are considered. Another notable early reference is [21] where an observability assumption is stated for the study of PDEs with constrained controllers. In these results, the control inputs are functions from a certain input space, and the saturation map has to be understood as a limitation on the norm of the input in this space. A more physically relevant type of saturation is given by pointwise saturations, which limit the values of the control input at each point by a prescribed value.

Pointwise saturations are more complex and require further developments, some of which we investigate in the present paper. Other types of saturation functions can also be useful in practice, see [25] for a discussion.

The definition of the saturation map used in the present paper is aligned to that used in [30] where Lyapunov methods are shown to be useful for the stability analysis of wave equations subject to saturated inputs. See also [25], [23], where systems in Hilbert spaces with applications to the Korteweg–de Vries equation have been addressed.

Our main result is the derivation of a saturated feedback control, which locally stabilizes the unstable heat equation, and for which estimates of the region of attraction for the closed-loop system can be provided. These estimates are formulated in terms of matrix inequalities which can be efficiently solved numerically. We emphasize that, in contrast to the works [36], [21], [30], where the stress is on the global stabilization of marginally stable hyperbolic systems (which have infinitely many eigenvalues on the imaginary axis), in our paper we consider parabolic systems, which have finitely many exponentially unstable modes. Therefore the control of the system with a saturating controller only provides local asymptotic stability, paralleling what is known for finite-dimensional systems (see in particular [42]).

We would like to mention the related research on model predictive control of parabolic PDEs under constraints on the state and the input [6], and on Lyapunov-based control of parabolic systems by controls of a bounded magnitude [7]. Note however, that in these papers the problem of estimating the region of attraction has not been studied, which is the key objective of this work. In [16] stabilizing backstepping-based boundary controllers for coupled heat-ODE systems with time-varying state delays in the presence of actuator saturation are developed, and in [17] boundary stabilization of a nonlinear Schrödinger equation with state delay and bounded internal disturbance is performed. Both in [16], [17] estimates of the

region of attraction of the closed-loop system are provided.

Our approach is based on the spectral decomposition of the open-loop dynamics of a heat equation into the finite-dimensional unstable part and infinite-dimensional stable dynamics, which is a classical tool used in PDE control [32], [2], [1]. To stabilize the unstable finite-dimensional part using a saturated control we apply techniques, which are well-known for ODE systems (see e.g. [41], [45], [33] for systems with non-delayed input, and [22] for systems with a delayed input).

Using Lyapunov functions, we derive in Section III linear matrix inequalities (LMIs) whose solutions provide estimates of the region of attraction of the closed-loop finite-dimensional system, see [4], [35], [43] for an introduction to matrix inequalities. We note, that a different Lyapunov function has been used for local ISS stabilization of the diffusion equation by saturated controls in the paper [39]. Then we show how asymptotic stability and estimates for the region of attraction can be obtained for the reaction-diffusion equation in closed-loop with the nonlinear saturated control.

We show that our results can be extended in a variety of different directions. In Section III-E we demonstrate, how dynamic controllers can be used to enlarge the region of attraction. In Section V-A we show that more complex types of saturation functions can be tackled, such as pointwise (L_∞) saturation functions. *Although our main results concern the stabilization via distributed controllers, in Section V-B we show how the methods can be applied to unbounded scalar control operators*, designed as the output of a finite-dimensional boundary control plant, which is fed with the saturated input.

In this light, our approach seems to be useful for any infinite-dimensional systems for which there exist only finitely many unstable modes and which is to be controlled through a bounded input operator (as is the case for systems with input delays considered in e.g. [9]), see also Remark 2.

The remaining part of this paper is organized as follows. We first introduce the reaction-diffusion equation with bounded input operator in Section II. The saturation function is defined and the spectral decomposition is provided. A saturated feedback is designed in Section III and an estimate for the region of attraction is provided, first for the open-loop unstable part, and then for both stable and unstable part. Numerical experiments conducted in Section IV illustrate our design method and the obtained estimates of the region of attraction depending on the saturation level. Other saturations are considered in Section V-A. In Section V-B, we show how our results can be applied in the case of unbounded input operators, that is to saturating boundary controllers resulting of a finite-dimensional dynamical system. Concluding remarks and possible future lines of research are collected in Section VI.

Notation: The set of nonnegative reals we denote by $\mathbb{R}_+$. The Euclidean norm on $\mathbb{R}^n$ is denoted by $|\cdot|$, the operator norm induced by this norm on spaces of matrices is denoted by $\|\cdot\|$. The interior of a set S in a topological space is denoted $\operatorname{int} S$ and $\overline{S}$ denotes its closure. In any normed vector space, the ball of radius r around 0 is denoted by $B_r(0)$. By $\mathbb{N}$ we denote the nonnegative integers. For convenience, $\mathbb{N}^* := \mathbb{N} \setminus \{0\}$. For $k \in \mathbb{N}, L > 0$, $H^k(0,L)$ denotes the Sobolev space of functions from the space $L_2(0,L)$, which have weak derivatives of order $\leqslant k$, all of which belong to $L_2(0,L)$. $H_0^k(0,L)$ is the closure of $C_0^k(0,L)$ (the k-times continuously differentiable functions with compact support in $(0,L)$) in the norm of $H^k(0,L)$.

II. Problem formulation

We consider the stabilization problem of a one-dimensional linear reaction-diffusion equation by means of a distributed control $u : \mathbb{R}_+ \to \mathbb{R}^m$. Let $L > 0$. We are given m functions $b_k : [0,L] \to \mathbb{R}$, $k = 1, \dots, m$ which describe at which places the control input $u_k \in \mathbb{R}$ is acting. The function c models the place-dependent reaction rate. The system model is then

$$\begin{aligned} w_t(t,x) &= w_{xx}(t,x) + c(x)w(t,x) \\ &\qquad + \sum_{k=1}^{m} b_k(x)\mathrm{sat}(u_k(t)),\ t > 0,\ x \in (0,L), \\ w(t,0) &= w(t,L) = 0, \quad t > 0, \\ w(0,x) &= w^0(x), \quad x \in (0,L). \end{aligned} \tag{1}$$

We assume that the state space of this system is $X := L_2(0,L)$ and that $c, b_k \in X$, $k = 1, \dots, m$.

Here sat is a component-wise saturation function, that is, for all $k = 1, \dots, m$ and for any $v \in \mathbb{R}^m$,

$$\mathrm{sat}(v)_k := \begin{cases} v_k & \text{if } |v_k| \leqslant \ell, \\ \frac{\ell}{|v_k|} v_k & \text{if } |v_k| \geqslant \ell, \end{cases} \tag{2}$$

where $\ell > 0$ is the given level of the saturation, which is assumed to be uniform with respect to the index k.

Remark 1: Systems of the form (1) also occur in the problem of stabilizing the linear heat equation by means of boundary control subject to delays or saturation, see e.g. [31] as well as Section V-B below. ∘

For the design of a stabilizing feedback, we use the well-known eigenfunction decomposition method. Define

$$A = \partial_{xx} + c(\cdot)\mathrm{id} : X \to X \tag{3}$$

with domain $D(A) = H^2(0,L) \cap H_0^1(0,L)$. Then the control system (1) takes the form

$$w_t(t,\cdot) = Aw(t,\cdot) + \sum_{k=1}^{m} b_k \mathrm{sat}(u_k(t)). \tag{4}$$

We note that A is selfadjoint and has compact resolvent, see Appendix B. Hence, the spectrum of A consists of only isolated eigenvalues with finite multiplicity, see [18, Theorem III.6.29]. Furthermore, there exists a Hilbert basis $(e_j)_{j \geqslant 1}$ of X consisting of eigenfunctions of A, associated with the sequence of eigenvalues $(\lambda_j)_{j \geqslant 1}$. Note that

$$-\infty < \cdots < \lambda_j < \cdots < \lambda_1 \quad \text{and} \quad \lambda_j \underset{j \to +\infty}{\longrightarrow} -\infty$$

and that $e_j(\cdot) \in D(A)$ for every $j \geqslant 1$.

We consider (mild) solutions of the system (1) (see [5, Section 3.1]), which exist and are unique for any initial condition in X and for any $u_k \in L_{1,loc}([0,\infty))$, for $k = 1, \dots, m$.

Every solution $w(t,\cdot) \in D(A)$ of (4) can be expanded as a series in the eigenfunctions $e_j(\cdot)$, convergent in $H_0^1(0,L)$,

$$\begin{aligned} w(t,\cdot) &= \sum_{j=1}^{\infty} w_j(t) e_j(\cdot), \\ w_j(t) &:= \langle w(t,\cdot), e_j(\cdot)\rangle_{L_2(0,L)},\ j \in \mathbb{N}^*. \end{aligned} \tag{5}$$

Analogously, we can expand the coefficients b_k in the series

$$b_k(\cdot) = \sum_{j=1}^{\infty} b_{jk} e_j(\cdot), \quad b_{jk} = \langle b_k(\cdot), e_j(\cdot)\rangle_{L_2(0,L)},\ j \in \mathbb{N}^*.$$

As discussed in Appendix A, (4) is equivalent to the infinite-dimensional control system

$$\begin{aligned} \dot{w}_j(t) &= \lambda_j w_j(t) + \sum_{k=1}^{m} b_{jk} \mathrm{sat}(u_k(t)) \\ &= \lambda_j w_j(t) + \mathbf{b}_j \cdot \mathrm{sat}(u(t)), \qquad j \in \mathbb{N}^*, \end{aligned} \tag{6}$$

where "$\cdot$" is the scalar product in $\mathbb{R}^m$, $\mathrm{sat}(u(t)) \in \mathbb{R}^m$ is the vector with entries $\mathrm{sat}(u_k(t))$ and $\mathbf{b}_j$ is the row vector with entries b_{jk}, $k = 1, \ldots, m$.

Let $n \in \mathbb{N}^*$ be the number of nonnegative eigenvalues of A and let $\eta > 0$ be such that

$$\forall j > n: \quad \lambda_j < -\eta < 0. \tag{7}$$

With the matrix notations

$$z := \begin{pmatrix} w_1 \\ \vdots \\ w_n \end{pmatrix}, \mathbf{A} := \begin{pmatrix} \lambda_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_n \end{pmatrix}, \mathbf{B} := \begin{pmatrix} b_{11} & \cdots & b_{1m} \\ \vdots & & \vdots \\ b_{n1} & \cdots & b_{nm} \end{pmatrix} \tag{8}$$

the n first equations of (6) form the unstable finite-dimensional control system

$$\dot{z}(t) = \mathbf{A} z(t) + \mathbf{B}\mathrm{sat}(u(t)). \tag{9}$$

Remark 2: (Possible extensions of our results).

(i) The same method can be used for stabilization with a prescribed decay rate $-\beta$. Just define for a suitable $\eta > \beta$ the minimal n such that (7) holds, and then use the method described here.

(ii) Since the input operator is bounded and the input space is finite-dimensional, the system can only be stabilized if the unstable spectrum consists of finitely many eigenvalues (counting multiplicities), see [14, Theorem 1.1] and the discussion in that paper relating to the history of the result. If this restriction is satisfied (e.g., this is the case for systems with input delays considered in [9]), the methods of this paper apply. ∘

III. Estimation of the region of attraction for saturated inputs

A. Decomposition of the system into stable and unstable part

We now introduce a decomposition of the state space into a finite-dimensional space on which the stabilization problem has to be solved and its orthogonal complement, which is invariant under the free dynamics.

Let X_n be the subspace of $L_2(0,L)$ spanned by $(e_i(\cdot))_{i=1}^n$ and π_n be the orthogonal projection onto X_n, that is

$$\pi_n w(t,\cdot) := \sum_{j=1}^{n} w_j(t) e_j(\cdot). \tag{10}$$

We define also $X_n^\perp$ as the orthogonal complement of X_n in X. Let $\iota : \mathbb{R}^n \to X_n$ be the isomorphism defined by $\iota(e^j) = e_j(\cdot)$, where $(e^j)_{j=1,\ldots,n}$ is the canonical basis of $\mathbb{R}^n$. We will use the isometric representation of $L_2(0,L)$ as $\ell_2(\mathbb{N}^*, \mathbb{R})$ obtained by the isomorphism induced by $e_j(\cdot) \mapsto e^j$, where

$$\ell_2(\mathbb{N}^*, \mathbb{R}) := \Big\{ (x_k)_{k \in \mathbb{N}^*} \in \mathbb{R}^{\mathbb{N}^*} \ : \ \sum_{k=1}^{\infty} |x_k|^2 < \infty \Big\},$$

where $e^j, j \in \mathbb{N}^*$ are the standard basis vectors in $\ell_2(\mathbb{N}^*, \mathbb{R})$ and we use the standard norm on that space. Corresponding to the decomposition $L_2(0,L) = X_n \bigoplus X_n^\perp$, where "$\bigoplus$" is the orthogonal sum of subspaces, we denote $\ell_2(\mathbb{N}^*, \mathbb{R}) = \mathbb{R}^n \bigoplus \ell_{2,j>n}$, where we identify $\mathbb{R}^n$ with the sequences with support in $\{1, \ldots, n\}$ and $\ell_{2,j>n}$ is the set of sequences in $\ell_2(\mathbb{N}^*, \mathbb{R})$ which are 0 in the first n entries.

Given a linear map $K : X_n \to \mathbb{R}^m$, consider the feedback

$$\begin{aligned} u =& K\pi_n w(\cdot) = K\Big(\sum_{j=1}^{n} w_j e_j(\cdot)\Big) = \sum_{j=1}^{n} w_j K e_j(\cdot) \\ =& \sum_{j=1}^{n} w_j \mathbf{K}_j = \mathbf{K} z, \end{aligned} \tag{11}$$

where $\mathbf{K}_j := K e_j(\cdot) \in \mathbb{R}^m$, $j = 1, \ldots, n$, and where we use the notation from (8) in the final step and set

$$\mathbf{K} := (\mathbf{K}_1, \ldots, \mathbf{K}_n) \in \mathbb{R}^{m \times n}.$$

Hence the system (6) with the feedback (11) is equivalent to the following set of differential equations:

$$\dot{w}_j(t) = \lambda_j w_j(t) + \mathbf{b}_j \cdot \mathrm{sat}(\mathbf{K} z(t)), \quad j \in \mathbb{N}^*. \tag{12}$$

Using (8), we rewrite the first n equations of (12) as

$$\dot{z}(t) = \mathbf{A} z(t) + \mathbf{B}\mathrm{sat}(\mathbf{K} z(t)). \tag{13}$$

Now, (12) can be considered as a cascade interconnection of an n-dimensional part, described by the equations (13) and of an infinite-dimensional part described by the equations

$$\dot{w}_j(t) \ = \ \lambda_j w_j(t) + \mathbf{b}_j \cdot \mathrm{sat}(\mathbf{K} z(t)), \quad j \geqslant n+1. \tag{14}$$

Our general assumption is:

Assumption 1: *The pair $(\mathbf{A}, \mathbf{B})$ is stabilizable, i.e. there is $\mathbf{K} \in \mathbb{R}^{m \times n}$: $\mathbf{A} + \mathbf{B}\mathbf{K}$ is Hurwitz.*

Remark 3: Clearly, the closed-loop system (13) with a matrix $\mathbf{K}$ as in Assumption 1 is locally asymptotically stable.

Conversely, if a feedback $\mathbf{K}$ renders the closed-loop system (13) locally asymptotically stable, then also

$$\dot{z} = \mathbf{A} z + \mathbf{B} u \tag{15}$$

is locally and hence globally asymptotically stabilized by means of the feedback $u(t) := \mathbf{K} z(t)$. Thus, local asymptotic stability of (13) implies that the pair $(\mathbf{A}, \mathbf{B})$ is stabilizable.

We note that in the case $m = 1$ the situation simplifies further as then (15) is a linear diagonal system with scalar control input. The criterion for stabilizability is then that $b_{j1} \neq 0$ for all $j = 1, \dots, n$ and $\lambda_k \neq \lambda_j$ for all $k, j = 1, \dots, n$, $k \neq j$ (which is an easy exercise). That is, the localization function b_1 should not be orthogonal to an unstable eigenfunction and that all unstable eigenvalues need to be simple. ◦

Next we show that the problem of exponential stabilization of the overall system (6) boils down to the exponential stabilization of the finite-dimensional unstable system (9). This latter problem will be elaborated in Section III-B.

Definition 1: *Assume that* $\mathbf{K}$ *is chosen so that* 0 *is a locally asymptotically stable fixed point of* (13)*. We say that* S *is a* region of attraction *of* 0 *if*

(i) $0 \in \operatorname{int} S$*;*
(ii) *for any* $z_0 \in S$ *the corresponding solution of* (13) *satisfies* $z(t; z_0) \to 0$ *as* $t \to \infty$*;*
(iii) S *is forward invariant, i.e. for any* $z_0 \in S$ *it holds that* $z(t; z_0) \in S$ *for all* $t \geqslant 0$*.*

The largest set (with respect to set inclusion) with the properties (i)-(iii) is called the maximal region of attraction*.*

As unions of regions of attraction are again a region of attraction, it is immediate that the maximal region of attraction is uniquely defined in this way and coincides with what is called domain of attraction in [12].

Definition 2: (13) *is called* locally exponentially stable in 0 with region of attraction S*, if the following conditions hold:*

(i) *there exist* $\varepsilon, M, a > 0$ *such that for any initial condition satisfying* $|z_0| \leqslant \varepsilon$*, it holds*

$$|z(t; z_0)| \leqslant M e^{-at} |z_0| \quad \forall t \geqslant 0 \tag{16}$$

(ii) $\overline{B_\varepsilon(0)} \subset S$ *and* S *is a region of attraction of* (13)*.*

We call (13) globally exponentially stable in 0 *if* (16) *holds for some* $a, M > 0$ *and all* $z_0 \in \mathbb{R}^n$*.*

Definitions 1 and 2 can be stated analogously for system (12).

We note that if the maximal region of attraction is not $\mathbb{R}^n$, then the system cannot be exponentially stable on the maximal region of attraction (for Lipschitz continuous systems), see [12]. Thus by analyzing regions of attraction with exponential stability we necessarily restrict the region of attraction.

Proposition 1: *Assume* $\mathbf{K}$ *is chosen such that the subsystem* (13) *is locally exponentially stable in* 0 *with region of attraction* $S \subset \mathbb{R}^n$*. Then:*

(i) *system* (12) *is locally exponentially stable in* 0 *with region of attraction* $S \times \ell_{2, j>n}$*.*
(ii) *system* (1) *with the feedback* (11) *is locally exponentially stable in* 0 *with region of attraction* $\iota(S) \times X_n^\perp$*.*

In addition, for any closed and bounded set $G \subset \operatorname{int}(\iota(S) \times X_n^\perp)$*, there exist two positive values* M *and* a *such that for any initial condition* $w(0, \cdot)$ *in* G*, the solution* $w(\cdot)$ *to* (1) *with the controller* (11) *satisfies*

$$\|w(t, \cdot)\|_X \leqslant M e^{-at} \|w(0, \cdot)\|_X \quad \forall t \geqslant 0. \tag{17}$$

Proof. Pick a compact subset G' of $\operatorname{int} S$. Since we assume that (13) is locally exponentially stable in 0 with region of attraction $S \subset \mathbb{R}^n$, it follows from a standard compactness argument that there exist $M, a > 0$ so that for any $z_0 \in G'$ the solution $z(\cdot; z_0)$ to (13) satisfies

$$|z(t; z_0)| \leqslant M e^{-at} |z_0|, \quad t \geqslant 0.$$

From equations (6) and (11), we derive that for $j = n+1, \dots, \infty$, for any $t \geqslant 0$ and for any $(w_{n+1}(0), w_{n+2}(0), \dots) \in \ell_{2, j>n}$ it holds that

$$w_j(t) = e^{\lambda_j t} w_j(0) + \mathbf{b}_j \cdot \int_0^t e^{\lambda_j (t-s)} \operatorname{sat}(\mathbf{K} z(s)) ds.$$

From (2) it follows that for all $z \in \mathbb{R}^n$ we have

$$|\operatorname{sat}(\mathbf{K} z)| \leqslant |\mathbf{K} z| \leqslant \|\mathbf{K}\| |z|.$$

Also due to the Cauchy-Bunyakovsky-Schwarz inequality we have that, for all $j \in \mathbb{N}$,

$$|b_{jk}| = \left| \langle b_k(\cdot), e_j(\cdot) \rangle_X \right| \leqslant \|b_k\|_X \|e_j\|_X = \|b_k\|_X. \tag{18}$$

Thus, for all $j \geqslant n+1$, we obtain exploiting (18) that

$$\begin{aligned} |w_j(t)| \leqslant & e^{-\eta t} |w_j(0)| + |\mathbf{b}_j| \int_0^t e^{-\eta(t-s)} |\mathbf{K} z(s)| ds \\ \leqslant & e^{-\eta t} |w_j(0)| + |\mathbf{b}_j| \|\mathbf{K}\| \int_0^t e^{-\eta(t-s)} M e^{-as} |z(0)| ds \\ = & e^{-\eta t} |w_j(0)| + |\mathbf{b}_j| \frac{M \|\mathbf{K}\|}{\eta - a} (e^{-at} - e^{-\eta t}) |z(0)|. \end{aligned}$$

The above computations have been performed for the case when $\eta \neq a$. If $a = \eta$, then it holds that

$$|w_j(t)| \leqslant e^{-\eta t} |w_j(0)| + M |\mathbf{b}_j| \|\mathbf{K}\| t e^{-\eta t} |z(0)|.$$

Now, using the inequality $(a+b)^2 \leqslant 2(a^2 + b^2)$ for any $(a, b) \in \mathbb{R}^2$, and the square summability of $|w_j(0)|$ and $|b_{jk}|$, $k = 1, \dots, m$, it follows that $\sum_{j=n+1}^\infty |w_j(t)|^2$ decays exponentially as well.

We now obtain local exponential stability of (12) by choosing G' such that $0 \in \mathbb{R}^n$ is in the interior of G' and noting that then $0 \in X$ is in the interior of $\iota(G') \times X_n^\perp$.

For the final statement of the proposition, pick a closed and bounded set $G \subset \operatorname{int}(\iota(S) \times X_n^\perp)$. Select $G' = \iota^{-1} \circ \pi_n(G)$, then G' is a compact subset of $\operatorname{int} S$, the previous computations yield (17) for suitable constants M and a and for the superset $\iota(G') \times X_n^\perp$ which contains G. □

Remark 4: It is not hard to see that (14) is input-to-state stable (ISS) with respect to the input z. Hence (12) is asymptotically stable as a cascade interconnection of a locally asymptotically stable system and an ISS system. However, since the general theorem on cascade interconnections does not guarantee exponential convergence, we needed an extra argument for Proposition 1. ◦

Remark 5: Our feedback controller is robust w.r.t. additive actuator disturbances. Let the control input to system (1) be $u(t) := \mathbf{K} z(t) + d(t)$, where $d \in PC(\mathbb{R}_+, \mathbb{R}^m)$ is a piecewise continuous actuator disturbance. Then there are

$r > 0$, $M_1, a_1, \gamma_1 > 0$ so that for all $w(0,\cdot) \in B_r$ and all $d \in PC(\mathbb{R}_+, \mathbb{R}^m)$ with $\sup_{s\geqslant 0} |d(s)| < r$ the solution of

$$\begin{aligned} &w_t(t,x) = w_{xx}(t,x) + c(x)w(t,x) \\ &+ \sum_{k=1}^{m} b_k(x)\mathrm{sat}\big((\mathbf{K}z(t))_k + d_k(t)\big),\ t>0,\ x\in(0,L), \\ &w(t,0) = w(t,L) = 0, \quad t>0, \\ &w(0,x) = w^0(x), \quad x\in(0,L). \end{aligned} \tag{19}$$

satisfies the local input-to-state stability (LISS) estimate

$$\|w(t,\cdot)\|_X \leqslant M_2 e^{-a_2 t}\|w(0,\cdot)\|_X + \gamma_2 \sup_{s\geqslant 0}|d(s)|. \tag{20}$$

This property can be obtained quite easily, since for small enough w and d we have $\mathrm{sat}(\mathbf{K}z(t) + d(t)) = \mathbf{K}z(t) + d(t)$, and thus (19) is a linear system with a bounded disturbance operator $d \mapsto \sum_{k=1}^m b_k(\cdot)d_k$, acting on $PC(\mathbb{R}_+, \mathbb{R}^m)$, and (20) follows as exponential stability of (19) without disturbances implies LISS for a system (19) with disturbances. It is, however, much harder to estimate a region of attraction of system (19), as in addition to the complexities arising in the undisturbed case the interplay between the size of a region of attraction and the maximal norm of a disturbance has to be analyzed. This can be an interesting topic for a future research. For more on ISS theory of infinite-dimensional systems the reader may consult [26], [28], [40] and references therein. ∘

B. Estimate of the region of attraction for the finite-dimensional part

In view of Proposition 1, it is important to study the local exponential stability and to estimate the region of attraction of the finite-dimensional system (13). We perform this task in this section.

Defining the *deadzone nonlinearity* $\phi : \mathbb{R}^m \to \mathbb{R}^m$ by

$$\phi(u) = \mathrm{sat}(u) - u, \quad u \in \mathbb{R}^m,$$

where sat is defined in (2), the following *generalized sector condition* holds (see [41, Lemma 1.6, Page 45] for a proof):

Lemma 1: *If for some $\mathbf{C} \in \mathbb{R}^{m\times n}$, $z \in \mathbb{R}^n$ and $j \in \{1,\dots,m\}$ it holds that $|((\mathbf{K}-\mathbf{C})z)_j| \leqslant \ell$, then*

$$\phi_j(\mathbf{K}z)(\phi_j(\mathbf{K}z) + (\mathbf{C}z)_j) \leqslant 0.$$

Remark 6: Note that other properties exist for saturation yielding to other representations of the saturation maps (see, e.g., [41, Section 1.7]). In the present work, we insist on using generalized sector conditions because they provide constructive methods to design Lyapunov functions. ∘

As a consequence of Lemma 1, for all diagonal positive definite matrices $\mathbf{D} \in \mathbb{R}^{m\times m}$, for all $\mathbf{C} \in \mathbb{R}^{m\times n}$, and all $z \in \mathbb{R}^n$ such that $|((\mathbf{K}-\mathbf{C})z)_j| \leqslant \ell$, $j = 1\dots,m$, we have:

$$\phi(\mathbf{K}z)^\top \mathbf{D}(\phi(\mathbf{K}z) + \mathbf{C}z) \leqslant 0. \tag{21}$$

We recall the following well-known Schur complement lemma (see e.g. [4, Chapter 2, p. 7]):

Lemma 2: *Let $A \in \mathbb{R}^{n\times n}$, $B \in \mathbb{R}^{m\times n}$, $C \in \mathbb{R}^{m\times m}$ and let $M := \begin{pmatrix} A & B^\top \\ B & C \end{pmatrix}$. If C is positive definite, then M is positive semidefinite if and only if its Schur complement $M/C := A - B^\top C^{-1} B$ is positive semidefinite.*

The following result is only a small variation of [11, Theorem 1]. Let us give a full proof using the notation used in this paper, for the sake of completeness.

Proposition 2: *Under Assumption 1, let $\mathbf{K} \in \mathbb{R}^{m\times n}$ be such that $\mathbf{A}+\mathbf{B}\mathbf{K}$ is Hurwitz. Let $P \in \mathbb{R}^{n\times n}$ be symmetric positive definite, $\mathbf{D} \in \mathbb{R}^{m\times m}$ diagonal positive definite and $\mathbf{C} \in \mathbb{R}^{m\times n}$ such that*

$$M_1 := \begin{bmatrix} (\mathbf{A}+\mathbf{B}\mathbf{K})^\top P + P(\mathbf{A}+\mathbf{B}\mathbf{K}) & P\mathbf{B} - (\mathbf{D}\mathbf{C})^\top \\ (P\mathbf{B})^\top - \mathbf{D}\mathbf{C} & -2\mathbf{D} \end{bmatrix} < 0 \tag{22}$$

and

$$M_2 := \begin{bmatrix} P & (\mathbf{K}-\mathbf{C})^\top \\ \mathbf{K}-\mathbf{C} & \ell^2 I_m \end{bmatrix} \geqslant 0. \tag{23}$$

Then the finite-dimensional system (13) *is locally exponentially stable in* 0 *with a region of attraction given by*

$$\mathcal{A} := \{z :\ z^\top P z \leqslant 1\}. \tag{24}$$

Moreover, in $\mathcal{A}$, the function V_1 defined by $V_1(z) := z^\top P z$, $z \in \mathbb{R}^n$, decreases exponentially fast to 0 *along the solutions to* (13)*, i.e. there is a constant $\alpha > 0$ so that*

$$\dot V_1(z) \leqslant -\alpha|z|^2, \quad z \in \mathcal{A}. \tag{25}$$

Proof. Let us show that (22), (23) are feasible. As $\mathbf{A}+\mathbf{B}\mathbf{K}$ is Hurwitz, there is a symmetric positive definite matrix $P \in \mathbb{R}^{n\times n}$ such that $(\mathbf{A}+\mathbf{B}\mathbf{K})^\top P + P(\mathbf{A}+\mathbf{B}\mathbf{K}) < 0$ and, multiplying if needed P by a positive constant, we may additionally assume that $\begin{bmatrix} P & \mathbf{K}^\top \\ \mathbf{K} & \ell^2 I_m \end{bmatrix} \geqslant 0$. Then, letting $\mathbf{C} = 0 \in \mathbb{R}^{m\times n}$, and picking a diagonal positive definite matrix $\mathbf{D} \in \mathbb{R}^{m\times m}$ with a norm sufficiently large, we get diagonal positive definite matrix $\mathbf{D} \in \mathbb{R}^{m\times m}$ and a matrix $\mathbf{C} \in \mathbb{R}^{m\times n}$ such that (22) and (23) hold.

Now pick $P, \mathbf{D}, \mathbf{C}$ such that 22) and (23) hold and consider the Lyapunov function candidate

$$V_1(z) = z^\top P z,$$

where $P \in \mathbb{R}^{n\times n}$ is the symmetric positive definite matrix given by the assumptions. The time-derivative of V_1 along the solutions to (13) for $z \in \mathbb{R}^n$ is given by

$$\dot V_1(z) = z^\top[(\mathbf{A}+\mathbf{B}\mathbf{K})^\top P + P(\mathbf{A}+\mathbf{B}\mathbf{K})]z + 2z^\top P\mathbf{B}\phi(\mathbf{K}z).$$

Now assuming $|((\mathbf{K}-\mathbf{C})z)_j| \leqslant \ell$ for all $j = 1,\dots,m$, it follows from (21) that

$$\begin{aligned} \dot V_1(z) &\leqslant z^\top[(\mathbf{A}+\mathbf{B}\mathbf{K})^\top P + P(\mathbf{A}+\mathbf{B}\mathbf{K})]z \\ &\quad + 2z^\top P\mathbf{B}\phi(\mathbf{K}z) - 2\phi(\mathbf{K}z)^\top \mathbf{D}(\phi(\mathbf{K}z) + \mathbf{C}z) \\ &= \begin{bmatrix} z \\ \phi(\mathbf{K}z) \end{bmatrix}^\top M_1 \begin{bmatrix} z \\ \phi(\mathbf{K}z) \end{bmatrix}. \end{aligned}$$

In view of (22), this implies the estimate (25) selecting $-\alpha$ as the maximal eigenvalue of M_1.

It remains to ensure that $|((\mathbf{K}-\mathbf{C})z)_j| \leqslant \ell$ is satisfied for all $j = 1,\dots,m$. To do this, we use the Lyapunov function V_1 and we impose that the ellipsoid $\{z \in \mathbb{R}^n :\ z^\top P z \leqslant 1\}$ is included in the ellipsoid $\{z \in \mathbb{R}^n :\ |(\mathbf{K}-\mathbf{C})z| \leqslant$

$\ell\}$ (this implies that $|((\mathbf{K}-\mathbf{C})z)_j| \leqslant \ell$ holds for all $j=1,\dots,m$). This inclusion is equivalent to the inclusion of $\{z \in \mathbb{R}^n : \ z^\top P z \leqslant 1\}$ in the set $\{z \in \mathbb{R}^n : z^\top(\mathbf{K}-\mathbf{C})^\top(\mathbf{K}-\mathbf{C})z \leqslant \ell^2\}$ which is again equivalent to the matrix inequality $P-(\mathbf{K}-\mathbf{C})^\top \frac{1}{\ell^2}(\mathbf{K}-\mathbf{C}) \geqslant 0$. As $\ell^2 I_m$ is positive definite, Lemma 2 ensures that this latter matrix inequality is equivalent to (23). □

Remark 7: We note that the formulation of Proposition 2 immediately gives room for a larger estimate for the domain of attraction so that the estimate given by $\{z : \ z^\top P z \leqslant 1\}$ can never be optimal. Indeed, in this region we have $\dot{V}_1(z) \leqslant -\alpha|z|^2$, so that by a continuity and compactness argument it follows that $\dot{V}_1(z) < 0$ on an enlarged region of the form $\{z : \ z^\top P z \leqslant 1+\varepsilon\}$, for a suitable $\varepsilon > 0$. ∘

With the previous result, it is also possible to analyze global stability. The region of attraction is global as soon as for all $z \in \mathbb{R}^n$, it holds that $|((\mathbf{K}-\mathbf{C})z)_j| \leqslant \ell$, $j=1,\dots,m$. This is equivalent to $\mathbf{K}=\mathbf{C}$. We thus obtain the following corollary:

Corollary 1: *If there exist a symmetric positive definite matrix $P \in \mathbb{R}^{n\times n}$ and a diagonal positive matrix $\mathbf{D} \in \mathbb{R}^{m\times m}$ such that* (22) *holds with $\mathbf{C} := \mathbf{K}$, then the finite-dimensional system* (13) *is globally exponentially stable in* 0*. Moreover the Lyapunov function V_1 decreases exponentially fast to* 0 *along the solutions to* (13)*.*

In our context, the main interest of Proposition 2 lies in the following consequence for system (1).

Theorem 1: *Consider system* (1) *with $(\mathbf{A},\mathbf{B})$ stabilizable. Let $\mathbf{K}$ be such that $\mathbf{A}+\mathbf{BK}$ is Hurwitz and K be as in* (11)*. Then the closed-loop system with the feedback K*

$$\begin{aligned} w_t(t,x) &= w_{xx}(t,x) + c(x)w(t,x) \\ &\quad + \sum_{j=1}^{m} b_k(x)\mathrm{sat}\big((K\pi_n w(t,\cdot))_k\big), \ t>0, \ x\in(0,L), \\ w(t,0) &= w(t,L) = 0, \quad t>0, \\ w(0,x) &= w^0(x), \quad x\in(0,L). \end{aligned} \tag{26}$$

is locally exponentially stable in 0 *with region of attraction $\imath(\mathcal{A})\times X_n^\perp$. In addition, the constants of decay can be chosen uniformly on $\imath(\mathcal{A})\times X_n^\perp$.*

Proof. Exponential stabilization and the region of attraction are immediate consequences of Proposition 2 in combination with Proposition 1. As the exponential estimate can be chosen uniformly on the set $\mathcal{A}$, the proof of Proposition 1 shows that the uniformity holds on all of $\imath(\mathcal{A})\times X_n^\perp$. □

C. Numerical implementation of Proposition 2

The inequalities (22) and (23) are *bilinear* matrix inequalities, since they have the bilinear cross term $\mathbf{DC}$ in the unknowns $\mathbf{D}$ and $\mathbf{C}$. The nonlinearity complicates the numerical analysis of (22) and (23). However, they can be reformulated into equivalent *linear matrix inequalities (LMIs)*, which are easier to solve numerically.

Let $S=P^{-1}$ and $\mathbf{E}=\mathbf{D}^{-1}$. Since $M_1<0$ if and only if

$$\begin{bmatrix} S & 0 \\ 0 & \mathbf{E} \end{bmatrix}^\top M_1 \begin{bmatrix} S & 0 \\ 0 & \mathbf{E} \end{bmatrix} < 0,$$

we obtain that (22) is equivalent to

$$\begin{bmatrix} S(\mathbf{A}+\mathbf{BK})^\top + (\mathbf{A}+\mathbf{BK})S & \mathbf{BE}-S\mathbf{C}^\top \\ (\mathbf{BE}-S\mathbf{C}^\top)^\top & -2\mathbf{E} \end{bmatrix} < 0. \tag{27}$$

Now pre- and post-multiplying (23) by $\begin{bmatrix} S & 0 \\ 0 & I_m \end{bmatrix}$, we get that (23) is equivalent to

$$\begin{bmatrix} S & S\mathbf{K}^\top - S\mathbf{C}^\top \\ (S\mathbf{K}^\top - S\mathbf{C}^\top)^\top & \ell^2 I_m \end{bmatrix} \geqslant 0. \tag{28}$$

Letting $\mathbf{Y}=S\mathbf{C}^\top$, we obtain (see also [11, Theorem 1] in a different context):

Proposition 3: *Existence of a symmetric positive definite matrix $P \in \mathbb{R}^{n\times n}$, a diagonal positive definite matrix $\mathbf{D} \in \mathbb{R}^{m\times m}$ and a matrix $\mathbf{C} \in \mathbb{R}^{m\times n}$ such that (22) and (23) hold is equivalent to the existence of a symmetric positive definite matrix $S \in \mathbb{R}^{n\times n}$, a diagonal positive definite matrix $\mathbf{E} \in \mathbb{R}^{m\times m}$ and $\mathbf{Y} \in \mathbb{R}^{m\times n}$ for which it holds that*

$$\begin{bmatrix} S(\mathbf{A}+\mathbf{BK})^\top + (\mathbf{A}+\mathbf{BK})S & \mathbf{BE}-\mathbf{Y} \\ (\mathbf{BE}-\mathbf{Y}^\top)^\top & -2\mathbf{E} \end{bmatrix} < 0 \tag{29}$$

and

$$\begin{bmatrix} S & S\mathbf{K}^\top - \mathbf{Y}^\top \\ (S\mathbf{K}^\top - \mathbf{Y}^\top)^\top & \ell^2 I_m \end{bmatrix} \geqslant 0. \tag{30}$$

This reformulation is important as the constraints (29) and (30) are linear in the variables S, $\mathbf{E}$ and $\mathbf{Y}$, in contrast to bilinear inequalities (22) and (23).

Finally, no structure is imposed on both the matrix $\mathbf{C} \in \mathbb{R}^{m\times n}$ and on the matrix $\mathbf{Y} \in \mathbb{R}^{m\times n}$, and we let $\mathbf{C} = (S^{-1}\mathbf{Y})^\top$ from the knowledge of S and $\mathbf{Y}$ (and respectively we may let $\mathbf{Y}=P^{-1}\mathbf{C}^\top$ from the knowledge of P and $\mathbf{C}$).

D. Scalar control inputs

In this section we specialize our results for scalar control inputs, i.e. $m=1$, which allows for a numerically more efficient method. We need the following lemma.

Lemma 3: *Consider a symmetric matrix $P = \begin{pmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{pmatrix} \in \mathbb{R}^{(n_1+n_2)\times(n_1+n_2)}$, where P_{ij} are matrices of appropriate dimension and $P_{11}>0$, $P_{22}>0$. Define $P_a := \begin{pmatrix} aP_{11} & P_{12} \\ P_{21} & P_{22} \end{pmatrix}$ for $a \geqslant 0$. Then there exists $a^* \geqslant 0$:*

(i) *$P_{a^*} \geqslant 0$, and P_{a^*} is not positive definite,*
(ii) *$P_a > 0$ for all $a > a^*$,*
(iii) *for all $a \in [0,a^*)$ the matrix P_a is not positive semidefinite.*

Proof. As $P_{22}>0$, the matrix P_a is positive definite if and only if the Schur complement

$$P_a/P_{22} = aP_{11} - P_{12}P_{22}^{-1}P_{21}$$

is positive definite, see Lemma 2. Note that as $(P_{12})^\top = P_{21}$ and $(P_{22}^{-1})^\top = P_{22}^{-1}$, we have that

$$(P_{12}P_{22}^{-1}P_{21})^\top = P_{12}^\top (P_{22}^{-1})^\top (P_{21})^\top = P_{12}P_{22}^{-1}P_{21},$$

and as P_{11} is positive definite, P_a/P_{22} is a symmetric matrix.

Weyl's inequality (see [38, Chapter IV, Corollary 4.9, p. 203]) implies that all eigenvalues of P_a/P_{22} are strictly

increasing functions of a with a slope bounded below everywhere by $\lambda_{\min}(P_{11}) > 0$. The claim is now immediate. Clearly, P_0 is not positive definite. As all eigenvalues are strictly increasing with an affine lower bound, eventually at some $a^* \geqslant 0$ the smallest eigenvalue is equal to 0 by continuity of eigenvalues. For all $a > a^*$ all eigenvalues are positive. □

The main result of this section is

Proposition 4: *Under Assumption 1, let $\mathbf{K} \in \mathbb{R}^{m\times n}$ be such that $\mathbf{A}+\mathbf{BK}$ is Hurwitz. Let $\tilde{P} \in \mathbb{R}^{n\times n}$ be symmetric positive definite and $\mathbf{C} \in \mathbb{R}^{1\times n}$ such that*

$$\tilde{M}_1 := \begin{bmatrix} (\mathbf{A}+\mathbf{BK})^\top \tilde{P} + \tilde{P}(\mathbf{A}+\mathbf{BK}) & \tilde{P}\mathbf{B} - \mathbf{C}^\top \\ (\tilde{P}\mathbf{B})^\top - \mathbf{C} & -2 \end{bmatrix} < 0. \quad (31)$$

Then the finite-dimensional system (13) *is locally exponentially stable in* 0 *with a region of attraction given by*

$$\{z : \ z^\top \tilde{P} z \leqslant \mathbf{D}^{-1}\}, \quad (32)$$

where $\mathbf{D}$ is the minimal real number such that:

$$\tilde{M}_2 := \begin{bmatrix} \mathbf{D}\tilde{P} & (\mathbf{K}-\mathbf{C})^\top \\ \mathbf{K}-\mathbf{C} & \ell^2 \end{bmatrix} \geqslant 0. \quad (33)$$

Moreover, in this region of attraction, the function V_1 defined by $V_1(z) = z^\top P z := \mathbf{D} z^\top \tilde{P} z$, for all $z \in \mathbb{R}^n$, decreases exponentially fast to 0 *along the solutions to* (13)*, i.e. there is a constant $\alpha > 0$ so that*

$$\dot{V}_1(z) \leqslant -\alpha |z|^2. \quad (34)$$

Proof. Define $P := \mathbf{D}\tilde{P}$. Substituting this expression into (22) and (23) and multiplying it by $\mathbf{D}^{-1}$ to obtain the inequalities (31) and (33). If there exist $\tilde{P}$ and $\mathbf{C}$ as in Proposition 4 so that the LMI (31) holds, then according to Lemma 3, one can find $\mathbf{D} > 0$ so that (33) holds as well.

The estimate for the region of attraction follows from Proposition 2. We choose the minimal $\mathbf{D}$ to obtain the largest estimate of a region of attraction (for given $\tilde{P}$ and $\mathbf{C}$). □

Remark 8: Lemma 3 can be helpful in finding the optimal $\mathbf{D}$ as it shows that the problem is to find the unique root of a piecewise analytic function. ◦

E. Enlarging the region of attraction using a dynamic controller

Consider a dynamic controller for system (13), instead of a static controller. By doing so, we add some degrees of freedom and thus we may enlarge the region of attraction. Such approach is useful for many control systems (see e.g. [41, Example 8.1] for a simple 2×2 example). To be more specific, we consider an additional finite-dimensional state z_c in $\mathbb{R}^{n_c}$ (for a given integer n_c) and design matrices $\mathbf{K}_1$, $\mathbf{K}_2$, $\mathbf{A}_c$ and $\mathbf{B}_c$ of appropriate dimensions so that the (estimation of the) region of attraction of

$$\begin{aligned} \dot{z} &= \mathbf{A}z + \mathbf{B}\mathrm{sat}(\mathbf{K}_1 z + \mathbf{K}_2 z_c) \\ \dot{z}_c &= \mathbf{A}_1 z_c + \mathbf{A}_2 z \end{aligned} \quad (35)$$

is larger (in the z-direction) than the estimate provided by Proposition 2 for (13). We rewrite the dynamics (35) by

$$\dot{Z} = \bar{\mathbf{A}} Z + \bar{\mathbf{B}}\mathrm{sat}(\bar{\mathbf{K}} Z) \quad (36)$$

where

$$\bar{\mathbf{A}} = \begin{pmatrix} \mathbf{A} & 0 \\ \mathbf{A}_2 & \mathbf{A}_1 \end{pmatrix}, \ \bar{\mathbf{B}} = \begin{pmatrix} \mathbf{B} \\ 0 \end{pmatrix}, \ \bar{\mathbf{K}} = \begin{pmatrix} \mathbf{K}_1 & \mathbf{K}_2 \end{pmatrix}, \quad (37)$$

with $\bar{\mathbf{A}} \in \mathbb{R}^{(n+n_c)\times(n+n_c)}$, $\bar{\mathbf{B}} \in \mathbb{R}^{(n+n_c)\times m}$ and $\bar{\mathbf{K}} \in \mathbb{R}^{m\times(n+n_c)}$.

Remark 9: For given symmetric positive definite matrices $P \in \mathbb{R}^{n\times n}$ and $\bar{P} \in \mathbb{R}^{(n+n_c)\times(n+n_c)}$, the inclusion of the ellipsoid $\{z \in \mathbb{R}^n : \ z^\top P z \leqslant 1\}$ in the projection (onto $\mathbb{R}^n$) of the ellipsoid $\{Z \in \mathbb{R}^{n+n_c} : \ Z^\top \bar{P} Z \leqslant 1\}$ is equivalent to

$$\begin{pmatrix} I_n & 0 \end{pmatrix} \bar{P} \begin{pmatrix} I_n \\ 0 \end{pmatrix} - P \leqslant 0 \ . \quad (38)$$

◦

Using this remark and applying Proposition 2 to system (35), we have the following:

Proposition 5: *Consider system* (13) *with a stabilizable pair $(\mathbf{A},\mathbf{B}) \in \mathbb{R}^{n\times n} \times \mathbb{R}^{n\times m}$, and matrices $\bar{\mathbf{A}}$, $\bar{\mathbf{B}}$ and $\bar{\mathbf{K}}$ defined in (37). Choose a symmetric positive definite matrix $\bar{P} \in \mathbb{R}^{(n+n_c)\times(n+n_c)}$, a diagonal positive matrix $\bar{\mathbf{D}} \in \mathbb{R}^{m\times m}$ and $\bar{\mathbf{C}} \in \mathbb{R}^{m\times(n+n_c)}$ such that* (38) *holds and*

$$\begin{bmatrix} (\bar{\mathbf{A}}+\bar{\mathbf{B}}\bar{\mathbf{K}})^\top \bar{P} + \bar{P}(\bar{\mathbf{A}}+\bar{\mathbf{B}}\bar{\mathbf{K}}) & \bar{P}\bar{\mathbf{B}} - (\bar{\mathbf{D}}\bar{\mathbf{C}})^\top \\ (\bar{P}\bar{\mathbf{B}})^\top - \bar{\mathbf{D}}\bar{\mathbf{C}} & -2\bar{\mathbf{D}} \end{bmatrix} < 0, \quad (39)$$

$$\begin{bmatrix} \bar{P} & (\bar{\mathbf{K}}-\bar{\mathbf{C}})^\top \\ \bar{\mathbf{K}}-\bar{\mathbf{C}} & \ell^2 I_m \end{bmatrix} \geqslant 0. \quad (40)$$

Then the finite-dimensional system (35) is locally exponentially stable in 0 *with a region of attraction given by*

$$\mathcal{A}_{\bar{P}} := \{Z \in \mathbb{R}^{n+n_c} : \ Z^\top \bar{P} Z \leqslant 1\}.$$

Moreover, the projection of the ellipsoid $\{Z : \ Z^\top \bar{P} Z \leqslant 1\}$ onto $\mathbb{R}^n$ is larger than the region of attraction $\mathcal{A}$ given by Proposition 2 in (24) *for system* (13)*.*

Finally, in $\mathcal{A}_{\bar{P}}$, the function V_1 defined by

$$\overline{V}_1(Z) = Z^\top \bar{P} Z, \quad Z \in \mathbb{R}^{n+n_c},$$

decreases exponentially fast to 0 *along the solutions to (35), i.e. there is a constant $\alpha > 0$ so that for all $Z \in \mathcal{A}_{\bar{P}}$*

$$\dot{\overline{V}_1}(Z) \leqslant -\alpha |Z|^2. \quad (41)$$

As in Section III-C, the inequalities (39) and (40) can be reformulated as linear matrix inequalities.

F. Lyapunov analysis of the closed-loop system under saturation control

Now under the assumptions of Proposition 2, we provide a Lyapunov function $V : \ell_2(\mathbb{N}^*,\mathbb{R}) \to \mathbb{R}_+$ for the system (12). To this end, we use the Lyapunov function P provided by Proposition 2 for the finite-dimensional subsystem (13).

Proposition 6: *Consider system* (12) *and assume $\mathbf{K}$ is chosen such that the subsystem* (13) *is locally exponentially stable in* 0*. Assume further that $P \in \mathbb{R}^{n\times n}$ is symmetric positive definite and such that for $\tilde{V}(z) = z^\top P z, z \in \mathbb{R}^n$ we have $\dot{\tilde{V}}(z) \leqslant -\alpha|z|^2$ on the set $\mathcal{A} := \{z \in \mathbb{R}^n : \tilde{V}(z) \leqslant 1\}$ along the solutions of* (13)*. Then there exist $\gamma, C > 0$ such that for $Q : \ell_2(\mathbb{N}^*,\mathbb{R}) \to \mathbb{R}$ defined by $Q(w) := \sum_{j=n+1}^{\infty} w_j^2$ the function*

$$V(w) := \tilde{V}(\pi_n w) + \gamma Q(w) = z^\top P z + \gamma Q(w) \quad (42)$$

(with the identification $\pi_n w = z$) satisfies for all $w \in \mathcal{A} \times \ell_{2,j>n}$ that along the solutions of (12)

$$\dot{V}(w) \leqslant -C\|w\|_{\ell_2}^2.$$

In particular, it follows that $S \times \ell_{2,j>n}$ is a region of attraction of 0 *for system* (12).

Proof. We write $w \in \ell_2(\mathbb{N}^*, \mathbb{R})$ as $w = z + z^\perp$, where $z \in \mathbb{R}^n, z^\perp \in \ell_{2,j>n}$. Here we identify $\mathbb{R}^n$ with the sequences with support in $\{1, \dots, n\}$ and $\ell_{2,j>n}$ is the set of sequences in $\ell_2(\mathbb{N}^*, \mathbb{R})$ which are 0 in the first n entries.

This decomposition is unique. Due to (7), we have along the solutions to (12)

$$\begin{aligned}\dot{Q}(w) =& 2\sum_{j=n+1}^{\infty} w_j(t)\dot{w}_j(t)\\ =& 2\sum_{j=n+1}^{\infty} w_j(t)\Big(\lambda_j w_j(t) + \mathbf{b}_j \cdot \mathrm{sat}(\mathbf{K}z(t))\Big)\\ \leqslant& -2\eta\sum_{j=n+1}^{\infty} w_j^2(t) + 2\sum_{j=n+1}^{\infty} |w_j(t)||\mathbf{b}_j||\mathrm{sat}(\mathbf{K}z(t))|\\ \leqslant& -2\eta\sum_{j=n+1}^{\infty} w_j^2(t) + 2\sum_{j=n+1}^{\infty} |w_j(t)||\mathbf{b}_j|\|\mathbf{K}\||z(t)|.\end{aligned}$$

Cauchy-Bunyakovsky-Schwarz inequality implies that

$$\dot{Q}(w) \leqslant -2\eta\|z^\perp(t)\|_{\ell_2}^2 + 2\|z^\perp(t)\|_{\ell_2}\|b^\perp(t)\|_{\ell_2}\|\mathbf{K}\||z(t)|.$$

Using Young's inequality we have for all $\kappa > 0$ that $2\|z^\perp(t)\|_{\ell_2}|z(t)| \leqslant \kappa\|z^\perp(t)\|_{\ell_2}^2 + \frac{1}{\kappa}|z(t)|^2$. We proceed to:

$$\dot{Q}(w) \leqslant -\big(2\eta - \kappa\|b^\perp\|_{\ell_2}\|\mathbf{K}\|\big)\|z^\perp(t)\|_{\ell_2}^2 + \frac{1}{\kappa}\|b\|_{\ell_2}\|\mathbf{K}\||z(t)|^2.$$

Therefore for any choice of γ in (42), we have by an application of Proposition 2 along the solutions to (12) that

$$\begin{aligned}\dot{V}(w) \leqslant& -\big(\alpha - \frac{\gamma}{\kappa}\|b^\perp\|_{\ell_2}\|\mathbf{K}\|\big)|z(t)|^2\\ &- \gamma\big(2\eta - \kappa\|b^\perp\|_{\ell_2}\|\mathbf{K}\|\big)\|z^\perp(t)\|_{\ell_2}^2.\end{aligned}$$

Therefore selecting first $\kappa > 0$ such that $2\eta - \kappa\|b^\perp\|_{\ell_2}\|\mathbf{K}\| > 0$ and then selecting $\gamma > 0$ such that $\alpha - \frac{\gamma}{\kappa}\|b^\perp\|_{\ell_2}\|\mathbf{K}\| > 0$, we get the existence of $C > 0$ such that

$$\dot{V}(w) \quad\leqslant\quad -C|z(t)|^2 - C\|z^\perp(t)\|_{\ell_2}^2 \tag{43}$$

provided that $z(t)$ lies in the ellipsoid $\mathcal{A}$, whatever the value of $z^\perp(t)$. As $\mathcal{A} \times \ell_{2,j>n}$ is an invariant subset of ℓ_2 for a system (12), we obtain that V is a Lyapunov function for (12), with a guaranteed region of attraction containing $\mathcal{A} \times \ell_{2,j>n}$. □

Note that this provides an alternative proof of Proposition 1. We can also use Proposition 5 and obtain a similar result for a dynamical controller.

IV. Numerical experiments

In this section we use Proposition 4 to obtain estimates for a region of attraction for the unstable heat equation (1) subject to a saturated scalar feedback controller. Let $c(\cdot)$ in equation (1) be a constant function. With slight abuse of notation we will write $c(\cdot) = c = const$.

According to [13, pp. 16-17] the eigenvalues of the operator

$$A := \partial_{xx} + c\,\mathrm{id} : X \to X \tag{44}$$

on the domain $D(A) = H^2(0, L) \cap H_0^1(0, L)$ are given by

$$\lambda_j := -\frac{\pi^2}{L^2}j^2 + c, \quad j \in \mathbb{N}^*, \tag{45}$$

and the eigenfunctions e_j, $j \in \mathbb{N}^*$ of $(A, D(A))$, which form a basis of $L_2(0, 1)$ are given by

$$e_j(x) := \Big(\frac{2}{L}\Big)^{1/2}\sin\frac{j\pi x}{L}, \quad j \in \mathbb{N}^*, \quad x \in (0, L). \tag{46}$$

Next we estimate the region of attraction of system (1) for the following choice of parameters:

$$c(x) \equiv 10, \quad L = 2, \quad \ell = 2, \quad b = e_1 + e_2,$$

where the e_j are defined in (46). For these parameters there are only 2 unstable eigenvalues. Using the eigenfunction decomposition of solutions of (1) as in Section II, we see that the matrices $\mathbf{A}, \mathbf{B}$ defined in (8) have the form

$$\mathbf{A} \approx \begin{pmatrix} 7.5325989 & 0 \\ 0 & 0.1303956 \end{pmatrix}, \qquad \mathbf{B} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

As the diagonal entries in the matrix $\mathbf{A}$ are distinct, and all the components of $\mathbf{B}$ are nonzero, the pair $(\mathbf{A}, \mathbf{B})$ is stabilizable. Different choices of the matrix $\mathbf{K}$ for the stabilizing feedback $u(t) = \mathbf{K}z(t)$ lead to different attraction rates and different regions of attraction. We demonstrate this with two examples.

A. Choice 1: Placing the poles at $(-1, -1)$

First we choose the matrix $\mathbf{K}$ so that $\sigma(\mathbf{A} + \mathbf{B}\mathbf{K}) = \{-1\}$, which results in

$$\mathbf{K} \approx \begin{pmatrix} -9.835618 & 0.1726235 \end{pmatrix}. \tag{47}$$

To estimate a region of attraction of (13) by Proposition 4, we proceed in two steps.

(i) First we solve the inequality (31) together with the additional constraints: $\tilde{P} - \tilde{P}^\top = 0$ and $\tilde{P} > 0$. Additionally, we impose an optimality condition for $\mathbf{C}$:

$$(\mathbf{K} - \mathbf{C}) \cdot (\mathbf{K} - \mathbf{C})^\top \to \min, \tag{48}$$

where $\cdot$ is the scalar product of vectors.

(ii) The idea behind (48) is to minimize the off-diagonal elements of the matrix $\tilde{M}_2$, which gives us at the second step the possibility to find large $\mathbf{D}$ (optimal for the given $\tilde{P}$, $\mathbf{C}$) satisfying the bilinear matrix inequality (33).

This algorithm is implemented in Scilab. For solution of the LMI (31) the LMITOOL package has been used. The resulting matrices $\tilde{P}, \mathbf{C}$ and the real number $\mathbf{D}$ are (approximately):

$$\tilde{P} = \begin{pmatrix} 2.1277468 & -0.0655569 \\ -0.0655569 & 0.0243008 \end{pmatrix}, \tag{49a}$$

$$\mathbf{C} = (-2.0635579, 0.0844904), \quad \mathbf{D} = 7.359375. \tag{49b}$$

Figure 1 shows an elliptic region of attraction (24), subject to $\tilde{P}, \mathbf{C}, \mathbf{D}$ given by (49) (in blue). Furthermore, in the same figure some trajectories are depicted obtained through

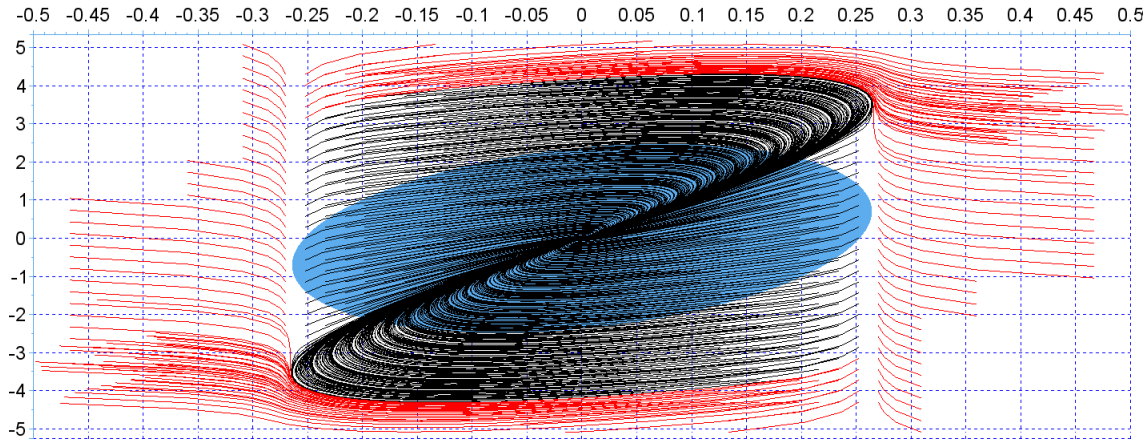

Fig. 1. Region of attraction of (13) (in blue) for the choice (47), (49), computed via the LMI technique. Trajectories of (13) are computed by direct solution of the ODE, trajectories attracted to the origin are in black, diverging trajectories are in red.

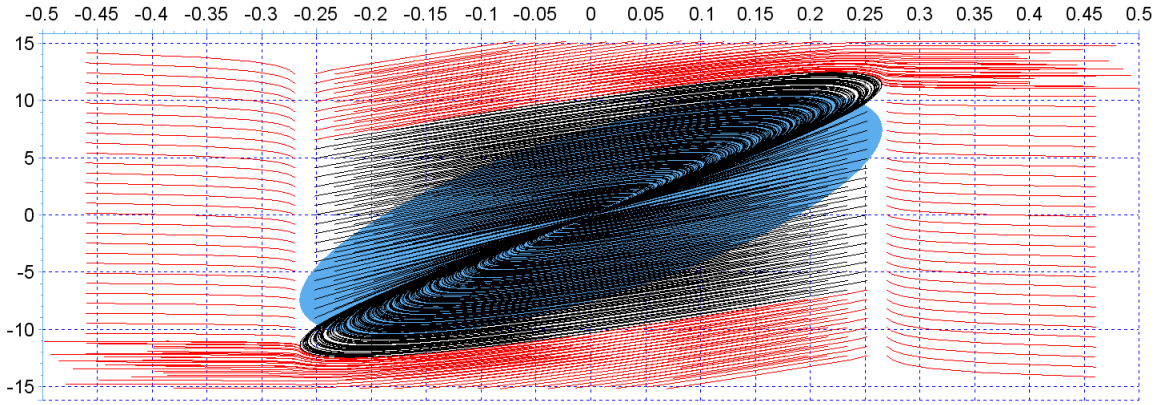

Fig. 2. Region of attraction of (13) (in blue) for the choice (50), (51), computed via the LMI technique. Trajectories of (13) are computed by direct solution of the ODE, trajectories attracted to the origin are in black, diverging trajectories are in red.

direct simulation of (13). The trajectories in black tend to the origin while those in red are diverging. This provides an approximation of the maximal region of attraction of (13). It can be seen that in one direction the ellipsoid obtained by our method approximates the actual region of attraction very well, but the results are not tight in the orthogonal direction.

Using Theorem 1, we obtain an approximation of the region of attraction of the PDE (1) subject to the feedback (11).

Remark 10: (Importance of the optimality conditions) Different solutions $\tilde{P}$, $\mathbf{C}$, $\mathbf{D}$ of the matrix inequalities (31), (33) lead to very different estimates of a region of attraction (24) of the model (13). Thus, it is important to pick solutions resulting in as large regions of attraction as possible. Here we chose a solution satisfying the optimality condition (48). Enforcing further optimality conditions may provide different estimates. The union of regions of attractions is again a region of attraction and so the maximal region can be explored further by solving (31), (33) for different optimality conditions.

In the two dimensional case, one option is to fix the eigendirections of P and to optimize the eigenvalues to get an idea of the extension of the maximal domain of attraction in particular directions. Other size criteria exist for optimization of domains of attraction, as volume maximization and trace minimization. See [41, Section 2.2.5] for more details and a complete introduction of such optimization problems. ○

Remark 11: (Computational cost) The time needed to solve the problem was (on a system with the specs: Intel(R) Core(TM) i5-3317U 1.70GHz, 16 GB RAM, Windows 10):

- Finding $\tilde{P}$, $\mathbf{C}$, $\mathbf{D}$ via LMIs: 0.0166561 seconds.
- Plotting the obtained region: 0.0071209 seconds.
- Time for solving the ODE (13) for $31^2 = 961$ distinct initial conditions on the time-interval $[0, 10]$ on a grid consisting of 100 points and for plotting the resulting trajectories: 43.359664 seconds.

This shows the computational efficiency of our method. ○

B. Choice 2: Putting the poles to $\{-0.1, -0.2\}$

Now let us choose the matrix $\mathbf{K}$ so that $\sigma(\mathbf{A} + \mathbf{BK}) = \{-0.1, -0.2\}$. This choice makes the attraction rate of the closed-loop system much slower, than in the previous simulation. This has however, some advantages, as we will see next. The resulting matrix $\mathbf{K}$ is:

$$\mathbf{K} = \begin{pmatrix} -7.9732782 & 0.0102837 \end{pmatrix}. \tag{50}$$

As in the previous simulation, we solve the LMI (31) subject to the additional optimality condition (48). The corresponding matrices $\tilde{P}, \mathbf{C}, \mathbf{D}$ are:

$$\tilde{P} = \begin{pmatrix} 0.3108695 & -0.0054849 \\ -0.0054849 & 0.000195 \end{pmatrix}, \tag{51a}$$

$$\mathbf{C} = (-0.3053879, 0.0054754), \quad \mathbf{D} = 90.625. \tag{51b}$$

As we see, with the choice of the stabilizing feedback (50), the region of attraction becomes significantly larger, although at the cost of reducing the rate of convergence of the trajectories to the origin. Furthermore, the choice of the matrices (51) leads to a better estimate of the region of attraction, in comparison to the situation in Section IV-A. This can be seen by comparing the Figures 1 and 2.

Remark 12: (Computational costs) For this problem the elapsed time is (on a system with the specs: Intel(R) Core(TM) i5-3317U 1.70GHz, 16 GB RAM, Windows 10)

- Finding $\tilde{P}$, $\mathbf{C}$, $\mathbf{D}$ via LMIs: 0.018131 seconds.
- Plotting the obtained region: 0.0129341 seconds.
- Time for solving the ODE (13) for $31^2 = 961$ distinct initial conditions on the time-interval $[0, 60]$ on a grid consisting of 600 points and for the plotting of the resulting trajectories: 91.802833 seconds.

We have chosen a longer time-span for solution of the ODE (13), since in this simulation the attraction rate of the closed loop system is much slower than in the previous simulation. Again, we obtain a considerable approximation of the region of attraction in a computationally efficient way. ∘

V. Extensions

A. Pointwise saturations

The type of the saturation which we have considered until now, i.e. component-wise saturation of finite-dimensional vectors, is not the only type of saturation functions, which appears in engineering practice. A general class of saturation functions has been considered in [24].

1) Saturation functions: Various definitions for the saturation map exist. Normwise saturations limit the norm of the input u, i.e. for a given norm $\|\cdot\|$ on $\mathbb{R}^n$ we may consider

$$\mathrm{sat}_{\|\cdot\|}(u) := \begin{cases} u & , \text{ if } \|u\| \leqslant \ell \\ \ell \frac{u}{\|u\|} & , \text{ if } \|u\| \geqslant \ell \end{cases} = \ell \min\left\{\frac{1}{\ell}, \frac{1}{\|u\|}\right\} u. \tag{52}$$

Another physically motivated saturation map is the following (pointwise) L_∞ saturation map, defined for all $x \in (0, L)$ by

$$\begin{aligned} \mathrm{sat}_\infty(u)(x) := & \begin{cases} u(x) & , \text{ if } |u(x)| \leqslant \ell \\ \ell \frac{u(x)}{|u(x)|} & , \text{ if } |u(x)| \geqslant \ell. \end{cases} \\ = & \ell \min\left\{\frac{1}{\ell}, \frac{1}{|u(x)|}\right\} u(x). \end{aligned} \tag{53}$$

In this section, we depart from the saturation model in (1) and study instead a heat equation with pointwise saturation in each input channel:

$$w_t(t,x) = w_{xx}(t,x) + c(x)w(t,x) + \sum_{k=1}^{m} \mathrm{sat}(b_k(x)u_k(t)). \tag{54}$$

In a certain sense, in the equation (54) the whole terms $u_k b_k$ are considered as the input which saturates pointwise. The assumptions on the domain of the problem and the functions c, b_k are the same as in Section II.

In order to stabilize (54), we are going to use the same stabilizing control (11). We now aim to provide an estimate for a region of attraction for (54). So we assume that $\mathbf{K}$ as in (11) is given and we consider the closed-loop system

$$w_t(t,x) = w_{xx}(t,x) + c(x)w(t,x) + \sum_{k=1}^{m} \mathrm{sat}\big(b_k(x)\left(\mathbf{K}z(t)\right)_k\big).$$

Representing this equation in the basis e_j, $j \in \mathbb{N}^*$ as in Section II, we obtain the equations for the coordinates

$$\dot{w}_j(t) = \lambda_j w_j(t) + \sum_{k=1}^{m} \Big\langle e_j, \mathrm{sat}_\infty\big(b_k(\cdot)\left(\mathbf{K}z(t)\right)_k\big)\Big\rangle, \quad j \in \mathbb{N}^*. \tag{55}$$

Let us state the following simple result that will be instrumental to bound the differences

$$\Delta\big(\left(\mathbf{K}z(t)\right)_k, b_k(\cdot)\big) := b_k(\cdot)\mathrm{sat}\big(\mathbf{K}z(t)_k\big) - \mathrm{sat}_\infty\big(b_k(\cdot)\mathbf{K}z(t)_k\big).$$

Lemma 4: *Let $r, k \in \mathbb{R}$ and denote $\Delta(r,k) = r\mathrm{sat}(k) - \mathrm{sat}(rk)$. Then*

$$|k| \leqslant \ell \text{ and } |rk| \leqslant \ell \quad \Rightarrow \quad \Delta(r,k) = 0. \tag{56}$$

$$|\Delta(r,k)| \leqslant \ell(1 + |r|). \tag{57}$$

Let $k \in \mathbb{R}$ and $b \in L_2(0, L)$ be given. Then

$$|k| \leqslant \ell \text{ and } \|b(\cdot)k\|_\infty \leqslant \ell \quad \Rightarrow \quad \Delta(b(\cdot), k) \equiv 0 \text{ a.e.} \tag{58}$$

Moreover

$$\|b(\cdot)\mathrm{sat}(k) - \mathrm{sat}_\infty(b(\cdot)k)\|_2 \leqslant \ell(\|1 + |b|\|_2). \tag{59}$$

If $|k| \leqslant \ell$ and χ is the characteristic function of the set $U := \{x \in [0, L] : |kb(x)| > \ell\}$ then

$$\|b(\cdot)\mathrm{sat}(k) - \mathrm{sat}_\infty(b(\cdot)k)\|_2 \leqslant \ell(\|\chi + |b\chi|\|_2). \tag{60}$$

If $b(\cdot)$ is essentially bounded, then

$$\|b(\cdot)\mathrm{sat}(k) - \mathrm{sat}_\infty(b(\cdot)k)\|_\infty \leqslant \ell(1 + \|b\|_\infty), \tag{61}$$

where $\|b\|_\infty$ denotes the L_∞ norm of the function b.

Proof. The first claim follows as the assumption guarantee that $r\mathrm{sat}(k) = \mathrm{sat}(rk) = rk$. The claim in (57) is a direct consequence of the triangle inequality as

$$|r\mathrm{sat}(k) - \mathrm{sat}(rk)| \leqslant |r||\mathrm{sat}(k)| + |\mathrm{sat}(rk)|.$$

The remaining claims follow immediately by applying the pointwise estimates (56) and (57). The claim (60) follows by applying (59) to $b\chi$, after noting that the complement of U does not contribute to the norm of the left hand side. □

Proposition 7: *Consider system* (54) *and assume all functions $b_k(\cdot) \in L_\infty([0, L])$, $k = 1, \ldots, m$. Consider also the associated system* (13) *with a stabilizable pair $(\mathbf{A}, \mathbf{B}) \in \mathbb{R}^{n\times n} \times \mathbb{R}^{n\times m}$ and let $\mathbf{K} \in \mathbb{R}^{m\times n}$ be such that $\mathbf{A} + \mathbf{B}\mathbf{K}$ is Hurwitz. Choose a symmetric positive definite matrix $P \in \mathbb{R}^{n\times n}$ as in Proposition 6. Consider the Lyapunov function candidate V defined in (42).*

Then there exist $\alpha, \beta > 0$ such that if for $w = z + z^\perp \in \ell_2(\mathbb{N}^*, \mathbb{R})$ we have $z \in \mathcal{A}_\beta := \{z : \ z^\top P z \leqslant \beta\}$ then

$$\dot{V}(w) \leqslant -\frac{\alpha}{2}|w|^2. \tag{62}$$

In particular, $\{z : \ z^\top P z \leqslant \beta\} \times X_n^\perp$ is a region of attraction for system (55)*.*

Proof. First we rewrite (54) by adding and subtracting the term $\sum b_k(x)\text{sat}(u_k(t))$ to obtain

$$\begin{aligned} w_t(t,x) = w_{xx}(t,x) + c(x)w(t,x) + \sum_{k=1}^m b_k(x)\text{sat}(u_k(t)) \\ + \sum_{k=1}^m \Big(\text{sat}\big(b_k(x)u_k(t)\big) - b_k(x)\text{sat}\big(u_k(t)\big)\Big). \end{aligned} \tag{63}$$

Define

$$\Delta(t) := \sum_{k=1}^m \text{sat}_\infty\big(b_k(\cdot)\left(\mathbf{K}z(t)\right)_k\big) - b_k(\cdot)\text{sat}(\mathbf{K}z(t)_k). \tag{64}$$

For $j \in \mathbb{N}^*$ we define $y_j(t) := \langle e_j, \Delta(t)\rangle$, and let $y(t) \in \mathbb{R}^n$ be the vector with components $y_j(t)$, for $j = 1, \dots, n$.

Considering the Lyapunov function V defined in (42), we compute its time-derivative along the solutions to (55). Using (43), we get for all $w = z + z^\perp$ with $z^\top P z \leqslant 1$ that

$$\begin{aligned} \dot{V}(w) \leqslant -C|z(t)|^2 - C\|z^\perp(t)\|_{\ell_2}^2 + 2z(t)^\top P y(t) \\ + 2\gamma \sum_{j=n+1}^\infty w_j(t)y_j(t). \end{aligned} \tag{65}$$

Using (59) from Lemma 4, we obtain along the solutions to (54) and as long as $z^\top P z \leqslant 1$,

$$\begin{aligned} \dot{V}(w) \leqslant & -C|z(t)|^2 - C\|z^\perp(t)\|_{\ell_2}^2 + 2\lambda_{\max}|z(t)|\|\Delta(t)\|_\infty \\ & + 2\gamma\|z^\perp(t)\|_{\ell_2}\|\Delta(t)\|_\infty, \end{aligned}$$

where $\lambda_{\max}$ denotes the maximal eigenvalue of the matrix P. Therefore for any positive values κ and κ',

$$\begin{aligned} \dot{V}(w) \leqslant & -\big(C - \frac{\lambda_{\max}}{\kappa}\big)|z(t)|^2 - (C - \frac{\gamma}{\kappa'})\|z^\perp(t)\|_{\ell_2}^2 \\ & + (\lambda_{\max}\kappa + \gamma\kappa')\|\Delta(t)\|_\infty^2. \end{aligned}$$

Pick $\kappa > 0$ and $\kappa' > 0$ such that $C - \frac{\lambda_{\max}}{\kappa} > \frac{3C}{4}$ and $C - \frac{\gamma}{\kappa'} > \frac{C}{2}$. Due to (58), there exists $\beta > 0$, such that for all z in $\{z : \ z^\top P z \leqslant \beta\}$, $\|\Delta(t)\|_\infty^2 \leqslant \frac{C}{4(\lambda_{\max}\kappa + \gamma\kappa')}|z(t)|^2$. We get, for all solutions to (54), as long as $z(t)$ is in $\{z : \ z^\top P z \leqslant \beta\}$,

$$\dot{V}(w) \leqslant -\frac{C}{2}|z(t)|^2 - \frac{C}{2}\|z^\perp(t)\|_{\ell_2}^2 \tag{66}$$

Moreover, we have, along the solutions to (54),

$$\dot{w}_j(t) = \lambda_j w_j(t) + \big\langle \text{sat}(bKz(t)), e_j\big\rangle, \qquad j = 1, 2, \dots \tag{67}$$

therefore, considering V_1 as previously defined, we may check that (62) holds following the same computation as for V along the solutions to (54), and using the fact the dynamics (67) in X_n does not depend on the component in $X_n^\perp$. Therefore the set $\{z : \ z^\top P z \leqslant \beta\} \times X_n^\perp$ is invariant along the dynamics to (54). With (66), we get that $\{z : \ z^\top P z \leqslant \beta\} \times X_n^\perp$ is a region of attraction and V is a Lyapunov function. □

B. Applications to boundary control of heat equation subject to control saturations

Let us now start from a heat equation with a dynamical boundary condition

$$y_t(t,x) = y_{xx}(t,x) + c(x)y(t,x), \ t \geqslant 0, \ x \in (0,L), \tag{68a}$$

$$y(t,0) = 0, \ y(t,L) = y_d, \ t \geqslant 0, \tag{68b}$$

where y_d is the (scalar) output of finite-dimensional dynamical system given by

$$\dot{x}_d = A_d x_d + B_d \text{sat}(u(t)), \tag{69a}$$

$$y_d = C_d x_d. \tag{69b}$$

Here x_d in $\mathbb{R}^{n_d}$ is the finite-dimensional state and the dynamics are subject to a saturating control, A_d, B_d and C_d are three matrices of appropriate dimension, and u is the scalar control input for the PDE (68) and the ODE (69) that is subject to a saturation map. Inspired by [31], we introduce the following change of variable:

$$w(t,x) = y(t,x) - \frac{x}{L}y_d(t), \quad t \geqslant 0, \ x \in (0,L).$$

The PDE for w then reads as:

$$\begin{aligned} w_t(t,x) &= y_t(t,x) - \frac{x}{L}\dot{y}_d(t) \\ &= y_{xx}(t,x) + c(x)y(t,x) - \frac{x}{L}C_d\dot{x}_d(t) \\ &= w_{xx}(t,x) + c(x)\big(w(t,x) + \frac{x}{L}y_d(t)\big) \\ &\qquad -\frac{x}{L}C_d\big(A_d x_d(t) + B_d\text{sat}(u(t))\big) \\ &= w_{xx}(t,x) + c(x)\big(w(t,x) + \frac{x}{L}C_d x_d(t)\big) \\ &\qquad -\frac{x}{L}C_d\big(A_d x_d(t) + B_d\text{sat}(u(t))\big) \\ &= w_{xx}(t,x) + c(x)w(t,x) \\ &\qquad + \big(\underbrace{c(x)\frac{x}{L}C_d - \frac{x}{L}C_d A_d}_{=:d(x)}\big)x_d(t) \\ &\qquad + \big(\underbrace{-\frac{x}{L}C_d B_d}_{=:b(x)}\big)\text{sat}(u(t)). \end{aligned} \tag{70}$$

Please note that b is a scalar function, and d is a row vector function with $d(x) \in \mathbb{R}^{1\times n_d}$, $x \in [0,L]$.

The boundary conditions for the variable w take the form:

$$w(t,0) = w(t,L) = 0, \ t \geqslant 0. \tag{71}$$

The heat equation (70), (71) has to be analyzed along with the ODE (69a).

Performing similar computations as in Section II for the PDE (1) and, using the same notation for w_j and λ_j, we get

$$\dot{w}_j(t) = \lambda_j w_j(t) + b_j\text{sat}(u(t)) + d_j x_d(t), \ j = 1, 2, \dots,$$

where b and d are defined for x in $[0,L]$ in (70) and $b_j = \langle b(\cdot), e_j(\cdot)\rangle_{L_2(0,L)}$, $d_j = \langle d(\cdot), e_j(\cdot)\rangle_{L_2(0,L)}$, for $j = 1, 2, \dots$.

Let us consider the first n equations with the ODE (69) and rewrite this finite-dimensional system as follows:

$$\begin{aligned} z'(t) &= \mathbf{A}z(t) + \mathbf{B}\mathbf{K}z(t) + \mathbf{B}\phi(\mathbf{K}z(t)) \\ &= \mathbf{A}z(t) + \mathbf{B}\text{sat}(\mathbf{K}z(t)), \end{aligned} \tag{72}$$

where $\mathbf{K}$ in $\mathbb{R}^{1\times(n+n_d)}$ is a row vector to be designed,

$$\begin{aligned} z(t) &:= (x_d^\top(t), \omega_1(t), \dots, \omega_n(t))^\top, \quad t \geqslant 0 \\ \mathbf{B} &:= (B_d^\top, b_1, \dots, b_n)^\top \in \mathbb{R}^{(n+n_d)\times 1}, \end{aligned}$$

and $\mathbf{A}:= \begin{pmatrix} A_d & 0 \\ D & \Lambda \end{pmatrix} \in \mathbb{R}^{(n+n_d)\times(n+n_d)}$, where

$$D:= \begin{pmatrix} d_{11} & d_{12} & \cdots & d_{1n_d} \\ \vdots & \vdots & \vdots & \vdots \\ d_{n1} & d_{n2} & \cdots & d_{nn_d} \end{pmatrix}, \quad \Lambda:= \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix}.$$

As the control input is scalar, we can apply Proposition 4 to system (72) instead of system (13), and get sufficient conditions for the estimation of the region of attraction of (72). Coming back to the infinite-dimensional systems (70) and (68), and, applying Proposition 1, we get sufficient conditions for an estimation of attraction region of (68):

Corollary 2: *Consider system* (13) *with a stabilizable pair* $(\mathbf{A}, \mathbf{B}) \in \mathbb{R}^{n\times n} \times \mathbb{R}^{n\times 1}$ *and let* $\mathbf{K} \in \mathbb{R}^{1\times n}$ *be such that* $\mathbf{A}+\mathbf{BK}$ *is Hurwitz. Pick a symmetric positive definite* $\tilde{P} \in \mathbb{R}^{(n+n_d)\times(n+n_d)}$ *and a* $\mathbf{C} \in \mathbb{R}^{1\times(n+n_d)}$ *such that* (31) *holds.*

Then the finite-dimensional system (72) is locally exponentially stable in 0 *with a region of attraction given by*

$$\mathcal{A} := \{z : \ z^\top \tilde{P} z \leqslant \mathbf{D}^{-1}\}, \tag{73}$$

where $\mathbf{D}$ *is the minimal real number such that* (33) *holds.*

Moreover,

(i) (70) *is locally exponentially stable with a region of attraction* $\imath(\mathcal{A}) \times X_n^\perp$,

(ii) (68) *is locally exponentially stable.*

VI. Conclusion

A linear unstable reaction-diffusion equation has been considered in this paper. Both boundary control and in-domain control cases have been considered. For this control problem, saturated feedback control laws have been designed so that the origin is a locally asymptotically stable equilibrium. The region of attraction has been estimated by an appropriate Lyapunov function and LMI technique. The interest and the efficiency of our approach have been illustrated by means of numerical simulations.

This work leaves several questions open. In particular it could be useful to consider other classes of Lyapunov functions than the ones considered in this work, and to compare the associated estimations of region of attraction. Moreover, it could be interesting to use this work for the estimation of the region of attraction in presence of disturbance and to study local input-to-state stability (as presented in Remark 5). Finally, let us note that extension of our method to nonlinear systems, other boundary conditions and control input may be considered as a future work.

References

[1] V. Barbu, I. Lasiecka, and R. Triggiani. *Tangential boundary stabilization of Navier-Stokes equations*. American Mathematical Soc., 2006.

[2] V. Barbu and R. Triggiani. Internal stabilization of Navier-Stokes equations with finite-dimensional controllers. *Indiana University Mathematics Journal*, 53(5):1443–1494, 2004.

[3] D. S. Bernstein and A. N. Michel. A chronological bibliography on saturating actuators. *International Journal of Robust and Nonlinear Control*, 5(5):375–380, 1995.

[4] S. Boyd, L. El Ghaoui, E. Feron, and V. Balakrishnan. *Linear Matrix Inequalities in System and Control Theory*. SIAM, 1994.

[5] R. F. Curtain and H. Zwart. *An Introduction to Infinite-Dimensional Linear Systems Theory*. Springer-Verlag, New York, 1995.

[6] S. Dubljevic, N. H. El-Farra, P. Mhaskar, and P. D. Christofides. Predictive control of parabolic pdes with state and control constraints. *International Journal of Robust and Nonlinear Control: IFAC-Affiliated Journal*, 16(16):749–772, 2006.

[7] N. H. El-Farra, A. Armaou, and P. D. Christofides. Analysis and control of parabolic PDE systems with input constraints. *Automatica*, 39(4):715–725, Apr. 2003.

[8] L. C. Evans. *Partial Differential Equations*. Graduate studies in mathematics. American Mathematical Society, 2010.

[9] Y. Fiagbedzi and A. Pearson. Feedback stabilization of linear autonomous time lag systems. *IEEE Transactions on Automatic Control*, 31(9):847–855, 1986.

[10] E. Fridman and Y. Orlov. Exponential stability of linear distributed parameter systems with time-varying delays. *Automatica*, 45(1):194–201, 2009.

[11] J. M. Gomes da Silva Jr and S. Tarbouriech. Antiwindup design with guaranteed regions of stability: an LMI-based approach. *IEEE Transactions on Automatic Control*, 50(1):106–111, 2005.

[12] W. Hahn. *Stability of Motion*. Springer-Verlag, New York, 1967.

[13] D. Henry. *Geometric Theory of Semilinear Parabolic Equations*. Springer-Verlag, Berlin, 1981.

[14] B. Jacob and H. Zwart. Equivalent Conditions for Stabilizability of Infinite-Dimensional Systems with Admissible Control Operators. *SIAM J. Control Optim.*, 37(5):1419–1455, 1999.

[15] B. Jacob and H. J. Zwart. *Linear Port-Hamiltonian Systems on Infinite-Dimensional Spaces*. Springer, Basel, 2012.

[16] W. Kang and E. Fridman. Boundary control of delayed ODE–heat cascade under actuator saturation. *Automatica*, 83:252–261, 2017.

[17] W. Kang and E. Fridman. Boundary constrained control of delayed nonlinear Schrödinger equation. *IEEE Transactions on Automatic Control*, 63(11):3873–3880, 2018.

[18] T. Kato. *Perturbation Theory for Linear Operators*. Classics in Mathematics. Springer-Verlag, Berlin, 1995.

[19] M. Krstic. Control of an unstable reaction-diffusion PDE with long input delay. *Systems & Control Letters*, 58(10-11):773–782, 2009.

[20] M. Krstic and A. Smyshlyaev. *Boundary Control of PDEs: A Course on Backstepping Designs*. SIAM, Philadelphia, PA, USA, 2008.

[21] I. Lasiecka and T. I. Seidman. Strong stability of elastic control systems with dissipative saturating feedback. *Systems & Control Letters*, 48(3):243–252, 2003.

[22] K. Liu and E. Fridman. Delay-dependent methods and the first delay interval. *Systems & Control Letters*, 64:57–63, 2014.

[23] S. Marx, V. Andrieu, and C. Prieur. Cone-bounded feedback laws for m-dissipative operators on Hilbert spaces. *Mathematics of Control, Signals, and Systems*, 29(4):18, 2017.

[24] S. Marx, E. Cerpa, C. Prieur, and V. Andrieu. Global stabilization of a Korteweg–De Vries equation with saturating distributed control. *SIAM Journal on Control and Optimization*, 55(3):1452–1480, 2017.

[25] S. Marx, Y. Chitour, and C. Prieur. Stability results for infinite-dimensional linear control systems subject to saturations. In *Proceedings of the European Control Conference (ECC'18)*, Limassol, Cyprus, 2018.

[26] A. Mironchenko and C. Prieur. Input-to-state stability of infinite-dimensional systems: recent results and open questions. *Accepted to SIAM Review, preprint: http://arxiv.org/abs/1910.01714*, 2020.

[27] A. Mironchenko, C. Prieur, and F. Wirth. Design of saturated controls for an unstable parabolic PDE. In *Proc. of the 11th IFAC Symposium on Nonlinear Control Systems (NOLCOS 2019)*, pages 937–942, 2019.

[28] A. Mironchenko and F. Wirth. Characterizations of input-to-state stability for infinite-dimensional systems. *IEEE Transactions on Automatic Control*, 63(6):1602–1617, 2018.

[29] S. Nicaise, J. Valein, and E. Fridman. Stability of the heat and wave equations with boundary time-varying delays. *Discrete and Continuous Dynamical Systems*, 2:559–581, 2009.

[30] C. Prieur, S. Tarbouriech, and J. M. Gomes da Silva Jr. Wave equation with cone-bounded control laws. *IEEE Transactions on Automatic Control*, 61(11):3452–3463, 2016.

[31] C. Prieur and E. Trélat. Feedback stabilization of a 1-D linear reaction–diffusion equation with delay boundary control. *IEEE Transactions on Automatic Control*, 64(4):1415–1425, 2018.

[32] D. L. Russell. Controllability and Stabilizability Theory for Linear Partial Differential Equations: Recent Progress and Open Questions. *SIAM Review*, 20(4):639–739, 1978.

[33] A. Saberi, A. A. Stoorvogel, and P. Sannuti. *Control of Linear Systems with Regulation and Input Constraints*. Springer Science & Business Media, 2012.

[34] A. M. Savchuk and A. A. Shkalikov. Sturm-Liouville operators with distribution potentials (in Russian). *Tr. Mosk. Mat. Obs.*, 64:159–212, 2003.

[35] C. Scherer and S. Weiland. Linear matrix inequalities in control. *Lecture Notes, Dutch Institute for Systems and Control, Delft, The Netherlands*, 3:2, 2000.

[36] M. Slemrod. Feedback stabilization of a linear control system in Hilbert space with an a priori bounded control. *Mathematics of Control, Signals, and Systems*, 2(3):265–285, 1989.

[37] O. Solomon and E. Fridman. Stability and passivity analysis of semilinear diffusion PDEs with time-delays. *International Journal of Control*, 88(1):180–192, 2015.

[38] G. W. Stewart and J.-G. Sun. *Matrix Perturbation Theory*. Computer Science and Scientific Computing. Academic Press, 1990.

[39] A. Tanwani, S. Marx, and C. Prieur. Local input-to-state stabilization of 1-D linear reaction-diffusion equation with bounded feedback. In *23rd International Symposium on Mathematical Theory of Networks and Systems (MTNS2018)*, pages 576–581, Hong Kong, 2018.

[40] A. Tanwani, C. Prieur, and S. Tarbouriech. Disturbance-to-state stabilization and quantized control for linear hyperbolic systems. *arXiv preprint arXiv:1703.00302*, 2017.

[41] S. Tarbouriech, G. Garcia, J. M. Gomes da Silva Jr, and I. Queinnec. *Stability and Stabilization of Linear Systems with Saturating Actuators*. Springer Science & Business Media, 2011.

[42] A. Teel. Global stabilization and restricted tracking for multiple integrators with bounded controls. *Systems & Control Letters*, 18:165–171, 1992.

[43] J. G. Van Antwerp and R. D. Braatz. A tutorial on linear and bilinear matrix inequalities. *Journal of Process Control*, 10(4):363–385, 2000.

[44] K. Yosida. *Functional Analysis*. Springer-Verlag, Berlin-Heidelberg, 1980.

[45] L. Zaccarian and A. Teel. *Modern Anti-windup Synthesis: Control Augmentation for Actuator Saturation*. Princeton Series in Applied Mathematics. Princeton University Press, 2011.

Andrii Mironchenko received his MSc at the I.I. Mechnikov Odessa National University in 2008 and his PhD at the University of Bremen in 2012. He has held a research position at the University of Würzburg and was a Postdoctoral JSPS fellow at the Kyushu Institute of Technology (2013–2014). In 2014 he joined the Faculty of Mathematics and Computer Science at the University of Passau. His research interests include infinite-dimensional systems, stability theory, hybrid systems and applications of control theory to biological systems.

Christophe Prieur was born in 1974. He is currently a senior researcher of the CNRS at the Gipsa-lab, Grenoble, France. He is currently a member of the EUCA-CEB, an associate editor for the AIMS Evolution Equations and Control Theory and IEEE Trans. on Control Systems Technology, a senior editor for the IEEE Control Systems Letters, and an editor for the IMA Journal of Mathematical Control and Information. He was the Program Chair of the 9th IFAC Symposium on Nonlinear Control Systems (NOLCOS 2013) and of the 14th European Control Conference (ECC 2015). His current research interests include nonlinear control theory, hybrid systems, and control of partial differential equations.

Fabian Wirth received his PhD from the Institute of Dynamical Systems at the University of Bremen in 1995. He has since held positions at the Centre Automatique et Systèmes of Ecole des Mines, the Hamilton Institute at NUI Maynooth, Ireland, the University of Würzburg and IBM Research Ireland. He now holds the chair for Dynamical Systems at the University of Passau. His current interests include stability theory, switched systems and large scale networks with applications to networked systems and in the domain of smart cities.

APPENDIX

A. Series expansion of solutions

Denoting the saturated nonlinearity in (1) by $f : \mathbb{R}_+ \to X$, we have for the mild solution of (4) that

$$w(t) = T(t)w(0) + \int_0^t T(t-s)f(s)ds\,,$$

where $T(t)$ is the strongly continuous semigroup generated by A. Here the integral on the right-hand side is well-defined as the Bochner integral, see [15, Example A.1.13].

Let $(e_j)_{j\geqslant 1}$ be the Hilbert basis of X given by the eigenfunctions of A. Then we can define

$$\begin{aligned} w_j(t) &:= \langle w(t), e_j\rangle \\ &= \langle T(t)w(0), e_j\rangle + \left\langle \int_0^t T(t-s)f(s)ds, e_j \right\rangle \end{aligned}$$

and by [44, Corollary V.5.2] we may interchange the linear map $\langle \cdot, e_j\rangle$ with the integral to obtain

$$\begin{aligned} w_j(t) &= \langle T(t)w(0), e_j\rangle + \int_0^t \langle T(t-s)f(s), e_j\rangle\, ds \\ &= \langle w(0), T(t)^* e_j\rangle + \int_0^t \langle f(s), T(t-s)^* e_j\rangle\, ds \\ &= e^{\lambda_j t}\langle w(0), e_j\rangle + \int_0^t e^{\lambda_j (t-s)}\langle f(s), e_j\rangle\, ds\,. \end{aligned}$$

Here the integral on the right is a standard Lebesque integral and so w_j solves an integral equation. Thus w_j is absolutely continuous and satisfies, for almost all t, the Carathéodory equation

$$\begin{aligned} \dot{w}_j(t) &= \lambda_j w_j(t) + \left\langle \sum_{k=1}^m b_k \mathrm{sat}\left(u_k(t)\right), e_j \right\rangle \\ &= \lambda_j w_j(t) + \sum_{k=1}^m b_{kj}\mathrm{sat}\left(u_k(t)\right)\,. \end{aligned}$$

This justifies the consideration of (6) as an equivalent system for (4).

B. Compactness of the resolvent for Sturm-Liouville operators

The following discussion summarizes some results from [34], which provide the necessary arguments to show the compactness of the resolvent of the operator A introduced in Section II.

Here we use the following notation. All function spaces are considered on the interval $[0, L]$. The Sobolev space W_p^k is the space of L_p-functions ($1 \leqslant p < \infty$) such that the function is k-times weakly differentiable and the corresponding derivatives

are again in L_p. Note in particular that $W_2^k = H^k$. For negative indices, we set $W_2^{-1} := \left(W_{0,2}^1\right)^*$ (the dual to $W_{0,2}^1 = H_0^1$).

Recall that an operator $A \in L(X)$ is said to be compact, if A maps bounded sets into precompact sets. For a densely defined linear operator $(A, D(A)) : X \to X$ with a nonempty resolvent set $\rho(A)$, it is an easy consequence of the resolvent identity that the resolvent $R_\lambda(A)$ is compact for some λ in the resolvent set, if and only if it is compact on the entire resolvent set, see [18, Theorem III.6.29].

Definition 3: *We say that a closed densely defined linear operator $(A, D(A)) : X \to X$ has a compact resolvent, if there exists a $\lambda \in \rho(A)$ so that $R_\lambda(A)$ is compact.*

1) Some results from [34]: We note that in [34] the case $L = \pi$ is considered, which requires some rescaling to use their results.

Let $X := L_2(0, L)$, $q \in W_2^{-1}(0, L)$, and define (for $y \in W_1^1(0, L)$) the quasiderivative

$$y^{[1]}(x) := \frac{dy}{dx} - Q(x)y(x), \tag{74}$$

where

$$Q(x) := \int_0^x q(s)ds.$$

For $q \in W_2^{-1}(0, L)$ it holds that $Q \in X$, and for $q \in X$ it holds that $Q \in W_2^1(0, L)$.

Let $q \in L_1(0, L)$ be given. Consider the formal Sturm-Liouville operator $SL : X \to X$ defined by

$$SL(y) := -\frac{d^2y}{dx^2} + q(x)y(x), \quad x \in (0, L),$$

where $(0, L) \subset \mathbb{R}$. In order to fully define the operator SL, we have to introduce its domain of definition. Following [34, Section 1.1] we define the maximal operator L_M, defined by

$$L_M y = SLy, \tag{75a}$$

$$D(L_M) := \{y : y, y^{[1]} \in W_1^1(0, L), SL(y) \in X\}. \tag{75b}$$

The following result has been shown in [34, Theorem 1.5]:

Theorem 2: *Let the operator A be the restriction of L_M to the domain*

$$D(A) := \{y \in D(L_M) : U_1(y) = U_2(y) = 0\}, \tag{76}$$

where for $j = 1, 2$ it holds that

$$U_j(y) = a_{j1}y(0) + a_{j2}y^{[1]}(0) + b_{j1}y(L) + a_{j2}y^{[1]}(L), \tag{77}$$

where $a_{j1}, a_{j2}, b_{j1}, a_{j2}$ are real numbers, for $j = 1, 2$.

Let $J_{\alpha\beta}$ be the determinant of the α-th and β-th column of a matrix

$$\begin{pmatrix} a_{11} & a_{12} & b_{11} & b_{12} \\ a_{21} & a_{22} & b_{21} & b_{22} \end{pmatrix}. \tag{78}$$

Then the operator A has a nonempty resolvent set $\rho(A)$, has a compact resolvent and discrete spectrum, if one of the following conditions holds:

(i) $J_{42} \neq 0$,
(ii) $J_{42} = 0$, $J_{14} + J_{32} \neq 0$,
(iii) $J_{42} = J_{14} = J_{32} = 0$, $J_{12} + J_{34} = 0$, $J_{13} \neq 0$.

Remark 13: Compactness of the resolvent of A is not mentioned in the formulation of [34, Theorem 1.5], but its compactness was shown in the proof. ∘

Remark 14: If one of conditions (i)–(iii) holds, then the boundary conditions (77) are called *Birkhoff-regular*. ∘

2) Application to the system in Section II: Consider the operator A as in Section II, i.e.

$$A := \partial_{xx} + c(\cdot)\mathrm{id} : X \to X, \tag{79a}$$

$$D(A) = H^2(0, L) \cap H_0^1(0, L) \tag{79b}$$

Let us show that Theorem 2 can be used for our operator A. We need the following lemma, see [8, Exercise 4 at p. 306].

Lemma 5: *Let $p \in [1, +\infty)$. Then $f \in W_p^1(0, L)$ if and only if f is equal a.e. to an absolutely continuous function, the derivative f' exists a.e. and is an element of $L_p(0, L)$.*

First we show

Lemma 6: *The domain $D(A)$ of the operator* (79) *has the form* (76) *for $U_1(y) = y(0)$ and $U_2(y) = y(L)$.*

Proof. Let $AC((0, L), \mathbb{R})$ be the space of absolutely continuous functions from $(0, L)$ to $\mathbb{R}$. In view of Lemma 5 the set $D(A)$ can be rewritten as

$$D(A) = \{f \in AC((0, L), \mathbb{R}) : \frac{df}{dx} \in AC((0, L), \mathbb{R}), \frac{d^2f}{dx^2} \in L_2(0, L),\ f(0) = f(L) = 0\}. \tag{80}$$

Now consider the domain defined in (76).

As we restrict our attention to the case when $q \in X$, for any $y \in W_1^1(0, L)$ it holds that $Q(\cdot)y(\cdot) \in W_1^1(0, L)$, and thus $y^{[1]} \in W_1^1(0, L)$ if and only if $\frac{dy}{dx} \in W_1^1(0, L)$.

Hence the domain $D(L_M)$ can be equivalently written as

$$D(L_M) = \{f : f, \frac{df}{dx} \in W_1^1(0, L), SL(f) \in X\}.$$

Using again Lemma 5, we see that

$$D(L_M) = \{f \in AC((0, L), \mathbb{R}) : \frac{df}{dx} \in AC((0, L), \mathbb{R}), \frac{d^2f}{dx^2} \in L_1(0, L),\ SL(f) \in X\}. \tag{81}$$

Furthermore, as $f \in D(L_M)$ is absolutely continuous on $(0, L)$, and $q \in X$, it holds that $qf \in X$ and it holds that $SL(f) \in X$ if and only if $\frac{d^2f}{dx^2} \in X$. As $X \subset L_1(0, L)$, we can finally restate $D(L_M)$ as

$$D(L_M) = \{f \in AC((0, L), \mathbb{R}) : \frac{df}{dx} \in AC((0, L), \mathbb{R}), \frac{d^2f}{dx^2} \in L_2(0, L)\}. \tag{82}$$

Using the boundary conditions $U_1(y) = y(0)$ and $U_2(y) = y(L)$, we see that the set (76) is precisely (80). □

In view of Lemma 6, we can use Theorem 2 to study the spectral properties of the Sturm-Liouville operator A.

Proposition 8: *The operator* (79) *has a compact resolvent.*

Proof. For the boundary conditions $U_1(y) = y(0)$ and $U_2(y) = y(L)$ the coefficients a_{ij} and b_{ij} from the formulation of Theorem 2 have the form: $a_{11} = 1$, $b_{21} = 1$, and all other entries of a matrix in (78) are zeros. By item (iii) of Theorem 2 the boundary conditions are Birkhoff-regular and the operator A has a compact resolvent. □