A continuous-time fundamental lemma and its application in data-driven optimal control*

Philipp Schmitz[a], Timm Faulwasser[b], Paolo Rapisarda[c], Karl Worthmann[a]

[a] *Technische Universität Ilmenau, Institute of Mathematics, Optimization-based Control Group, Ilmenau, Germany*
[b] *Hamburg University of Technology, Institute of Control Systems, Hamburg, Germany*
[c] *University of Southampton, Southampton, United Kingdom*

Abstract

Data-driven control of discrete-time and continuous-time systems is of tremendous research interest. In this paper, we explore data-driven optimal control of continuous-time linear systems using input-output data. Based on a density result, we rigorously derive error bounds for finite-order polynomial approximations of elements of the system behavior. To this end, we leverage a link between latent variables and flat outputs of controllable systems. Combined with a continuous-time counterpart of Willems et al.'s fundamental lemma, we characterize the suboptimality resulting from polynomial approximations in data-driven linear-quadratic optimal control. Finally, we draw upon a numerical example to illustrate our results.

Keywords: continuous time, data-driven control, differential flatness, identifiable, persistency of excitation, polynomial approximation

1. Introduction

Data-driven control, i.e., the design of controller and feedback laws directly from measured data, is a topic of ongoing research interest, see [1, 2] and the range of articles in these special issues. A pivotal result at the core of many developments in linear discrete-time systems is Willems et al.'s fundamental lemma [3].

Lemma 1 (Discrete-time fundamental lemma)**.** *Consider the discrete-time controllable linear time-invariant system*

$$x(k+1) = Ax(k) + Bu(k), \quad y(k) = Cx(k) + Du(k)$$

and let $\hat{w} = \operatorname{col}(\hat{u}, \hat{y})$ be a length-N input-output trajectory of the system such that $\hat{u}$ is persistently exciting (cf. [3]) of oder $n + L$, where $L \in \mathbb{N} \setminus \{0\}$ and n is the systems' state dimension. Then $w = \operatorname{col}(u, y)$ is a length-L input-output trajectory of the system if and only if

$$\begin{bmatrix} w(k) \\ \vdots \\ w(L-1) \end{bmatrix} \in \operatorname{im} \begin{bmatrix} \hat{w}(0) & \dots & \hat{w}(N-L) \\ \vdots & & \vdots \\ \hat{w}(L-1) & \dots & \hat{w}(N-1) \end{bmatrix}.$$

This lemma enables the parameterization of the external (input-output) finite-horizon behavior of controllable systems using sufficiently informative data arranged in a Hankel matrix. For more details and recent advancements, see the survey [4] and the references therein. The interest in this result has been catalyzed by recent works such as [5, 6, 7]; applications of data-driven control concepts are discussed in the literature [8, 9, 10, 11]. In [12, 13] and [14] continuous-time extensions of the fundamental lemma have been proposed. While the former works require to solve a scalar ODE to compute future trajectories, in the latter paper an approach to compute a generating representation based on polynomial series expansions of input-output trajectories is proposed. However, the latter result has not yet been used to design data-driven controllers. In the present paper, we address this gap: We extend the results from [14] by deriving errors bounds for polynomial series expansions on elements of the behavior. We also prove that the set of polynomial system trajectories is dense in the set of all system trajectories. We use the approximation bounds to derive bounds on the optimality gap resulting from using finite-order polynomials to solve linear–quadratic regulator (LQR) problem formulated in terms of the behavior. Finally, we establish a continuous-time fundamental lemma involving the Gramians of trajectories and, based on this, present a data-driven approximation for the LQR problem.

The remainder of this paper is structured as follows: In Section 2 we revisit foundational concepts in behavioral systems theory, with most results rederived to suit our specific setting. Utilizing the connection between flat out-

*Funding: P. Schmitz is grateful for the support from the Carl Zeiss Foundation (VerneDCt – Project No. 2011640173). K. Worthmann gratefully acknowledges funding by the German Research Foundation (DFG; project number 507037103).

*Corresponding Author: Philipp Schmitz

Email addresses: philipp.schmitz@tu-ilmenau.de (Philipp Schmitz), timm.faulwasser@ieee.org (Timm Faulwasser), pr3@ecs.soton.ac.uk (Paolo Rapisarda), karl.worthmann@tu-ilmenau.de (Karl Worthmann)

puts and latent variables, in Lemma 6, we derive a specific behavioral representation, which is beneficial for subsequent approximation results. Moreover, we recap properties of Legendre polynomials (Subsections 2.2 and 2.3) and analyze their advantage in the approximation of behavioral elements (Subsection 2.4) as shown in Proposition 11. In Section 3, we introduce a version of the finite-horizon linear-quadratic optimal control problem and its finite-dimensional approximation in the space of linear combinations of Legendre polynomials including a convergence analysis. In Section 4, we define the concept of persistency of excitation (Subsection 4.1) and we state a continuous-time fundamental lemma (Theorem 22). Subsection 4.3 demonstrates how the fundamental lemma can be applied in system identification. In Section 5, we discuss the application of our results to the data-driven solution of the finite-horizon optimal control problem before we conclude the paper in Section 6 pointing out directions of current and future research.

Notation: Given two sets X and Ω, the set of functions $f : \Omega \to X$ is denoted by X^Ω. Let $\mathcal{I}$ be a real interval; we denote by $\overline{\mathcal{I}}$ the closure of $\mathcal{I}$. Let $d, k \in \mathbb{N}$; then $L^2(\mathcal{I}, \mathbb{R}^d)$ denotes the space of equivalence classes of square integrable functions $f \in (\mathbb{R}^d)^\mathcal{I}$ and $H^k(\mathcal{I}, \mathbb{R}^d)$ is the kth order Sobolev space associated with $L^2(\mathcal{I}, \mathbb{R}^d)$. The scalar product in $L^2(\mathcal{I}, \mathbb{R}^d)$ and its induced norm are given by $\langle f, g\rangle = \int_\mathcal{I} f(\tau)^\top g(\tau)\,\mathrm{d}\tau$ and $\|f\| = \sqrt{\langle f, f\rangle}$. The usual norm in $H^k(\mathcal{I}, \mathbb{R}^d)$ is denoted by $\|\cdot\|_{H^k}$. For $k \in \mathbb{N}\setminus\{0\}$ and $f \in H^{k-1}(\mathcal{I}, \mathbb{R}^d)$ we set

$$\Lambda_k(f) = \begin{bmatrix} f \\ \vdots \\ f^{(k-1)} \end{bmatrix} \in L^2(\mathcal{I}, \mathbb{R}^{k\cdot d}). \tag{1}$$

In particular, $\|f\|_{H^{k-1}} = \|\Lambda_k(f)\|$. $\mathcal{C}^\infty(\mathcal{I}, \mathbb{R}^d)$ is the space of infinitely differentiable functions from $\mathcal{I}$ to $\mathbb{R}^d$ and $\mathcal{C}^\infty_\mathrm{c}(\mathcal{I}, \mathbb{R}^d)$ consists of those functions of $\mathcal{C}^\infty(\mathcal{I}, \mathbb{R}^d)$ with compact support. Given a Hilbert space X, $\ell^2(\mathbb{N}, X)$ is the space of square summable sequences in $X^\mathbb{N}$.

The identity operator from a vector space X onto itself is denoted by I_X or simply I when clear from the context. In the case of a finite dimensional space $X = \mathbb{R}^d$ we also write I_d. The Euclidean norm in $\mathbb{R}^d$ is denoted by $\|\cdot\|_2$. If $A_0, \ldots, A_k$ are matrices with the same number of columns, we define $\mathrm{col}(A_0, \ldots, A_k) := \begin{bmatrix} A_0^\top & \ldots & A_k^\top \end{bmatrix}^\top$. We always identify $\mathbb{R}^d$ with $\mathbb{R}^{d\times 1}$. Given a matrix M, we denote by $\mathrm{im}\, M$ and $\ker M$ its image and kernel. Further, $M^\top$ and $M^\dagger$ denote the transpose and Moore–Penrose inverse.

Finally, we denote by $\mathbb{R}[s]$ the ring of polynomials with real coefficients in the indeterminate s, and by $\mathbb{R}^{g\times q}[s]$ the ring of $g \times q$ polynomial matrices with real coefficients.

2. Linear differential systems

We first recapitulate behavioral concepts for linear time-invariant systems. We then connect this paradigm to polynomial series expansions of trajectories and explore how polynomial trajectories can approximate system behaviors.

2.1. Behaviors

In the following, we deal with dynamical systems given by the time interval $\mathcal{I} = (-1, 1)$[1], the signal space $\mathbb{R}^q$ and a *behavior* $\mathcal{B} \subset (\mathbb{R}^q)^\mathcal{I}$. We focus on *linear differential behaviors*, i.e. the set of solutions to a system of linear, constant-coefficient differential equations

$$R\left(\frac{\mathrm{d}}{\mathrm{d}t}\right) w = 0, \tag{2}$$

where $R(s) = R_0 s^0 + \cdots + R_r s^r$ is a polynomial matrix in $\mathbb{R}^{g\times q}[s]$. In this way (2) is to be understood as $\sum_{i=0}^r R_i \frac{\mathrm{d}^i w}{\mathrm{d}t^i} = 0$. Here, the solution w of (2) is meant in the sense of *weak solutions*, i.e., $w \in L^2(\mathcal{I}, \mathbb{R}^q)$ and it satisfies

$$0 = \sum_{i=0}^r (-1)^i \int_\mathcal{I} w^\top R_i^\top \tfrac{\mathrm{d}^i \phi}{\mathrm{d}t^i}\,\mathrm{d}t \tag{3}$$

for all test functions $\phi \in \mathcal{C}^\infty_\mathrm{c}(\mathcal{I}, \mathbb{R}^g)$. Given $w \in \mathcal{C}^\infty(\overline{\mathcal{I}}, \mathbb{R}^q)$ integration by parts shows that (3) is valid if and only (2) holds pointwise.

Given our choice of solution set, we need to slightly generalize and prove some well-known results from [15], where the solutions of (2) are assumed to be infinitely differentiable. In particular, the equivalence of the different representations of behaviors, established essentially for smooth functions in [15], requires verification in the context of weak L^2-solutions.

Lemma 2. *The behavior*

$$\mathcal{B} := \{w \in L^2(\mathcal{I}, \mathbb{R}^q) \mid R(\tfrac{\mathrm{d}}{\mathrm{d}t})w = 0\} \tag{4}$$

is closed in $L^2(\mathcal{I}, \mathbb{R}^q)$.

Proof. Consider a sequence $(w_n)_{n\in\mathbb{N}}$ in $\mathcal{B}$ which converges to some $w \in L^2(\mathcal{I}, \mathbb{R}^q)$. Note that for each $\phi \in \mathcal{C}^\infty_\mathrm{c}(\mathcal{I}, \mathbb{R}^g)$ the right hand side in (3) defines a linear, continuous functional $f_\phi : L^2(\mathcal{I}, \mathbb{R}^q) \to \mathbb{R}$. By continuity $0 = \lim_{n\to\infty} f_\phi(w_n) = f_\phi(w)$ for every $\phi \in \mathcal{C}^\infty_\mathrm{c}(\mathcal{I}, \mathbb{R}^g)$, that is w solves (2) and $w \in \mathcal{B}$. □

We recall the notion of behavioral controllability, cf. Definition 5.2.2 in [15]. The behavior $\mathcal{B}$ is *controllable* if for each two trajectories w_0, $w_1 \in \mathcal{B}$ there is $t_1 \in (0, 1)$ and $w \in \mathcal{B}$ such that

$$w(t) = \begin{cases} w_0(t) & \text{if } t \in (-1, 0], \\ w_1(t - t_1) & \text{if } t \in [0, 1). \end{cases} \tag{5}$$

[1] We choose such interval purely for simplicity of notation; with straightforward modifications, *any* other bounded open interval can be used.

Partitioning R compatibly with a known selection of inputs and outputs $w = \mathrm{col}(u, y) \in \mathcal{B}$, cf. Definition 3.3.1 in [15], one obtains the *input-output representation*

$$P\left(\frac{\mathrm{d}}{\mathrm{d}t}\right) y = Q\left(\frac{\mathrm{d}}{\mathrm{d}t}\right) u, \tag{6}$$

where $P \in \mathbb{R}^{p\times p}[s]$ can be assumed to be nonsingular and $Q \in \mathbb{R}^{p\times m}[s]$ with $q = m + p$.

A linear differential behavior $\mathcal{B}$ also admits an *input-state-output representation* (see [16]), i.e., there are matrices $A \in \mathbb{R}^{n\times n}$, $B \in \mathbb{R}^{n\times n}$, $C \in \mathbb{R}^{p\times n}$, $D \in \mathbb{R}^{p\times m}$ such that for all $w = \mathrm{col}(u, y) \in \mathcal{B} \cap \mathcal{C}^\infty(\overline{\mathcal{I}}, \mathbb{R}^q)$ there exists $x \in \mathcal{C}^\infty(\overline{\mathcal{I}}, \mathbb{R}^n)$ with

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t} x &= Ax + Bu \\ y &= Cx + Du. \end{aligned} \tag{7}$$

The following result relates the *external behavior* described by (6), consisting of all input-output trajectories, and its input-state-output counterpart in the terms of closed L^2-subspaces.

Lemma 3. *Given an input-state-output representation (7) of $\mathcal{B}$, one has*

$$\mathcal{B} = \left\{ \mathrm{col}(u, y) \in L^2(\mathcal{I}, \mathbb{R}^{m+p}) \;\middle|\; \begin{array}{l} \exists\, x \in H^1(\mathcal{I}, \mathbb{R}^n) \\ \textit{s.t. } (7) \textit{ holds} \end{array} \right\}. \tag{8}$$

Proof. Denote the set on the right hand side of (8) by $\widetilde{\mathcal{B}}$. We show that $\widetilde{\mathcal{B}}$ is a closed subspace of $L^2(\mathcal{I}, \mathbb{R}^{m+p})$. To this end consider the solution operator $S : L^2(\mathcal{I}, \mathbb{R}^m) \to H^1(\mathcal{I}, \mathbb{R}^n)$ defined by

$$(Su)(t) := \int_{-1}^{t} \exp(A(t-\tau))Bu(\tau)\,\mathrm{d}\tau. \tag{9}$$

From the definition of $\widetilde{\mathcal{B}}$ it follows that $\mathrm{col}(u, y) \in \widetilde{\mathcal{B}}$ if and only if there exists $x^0 \in \mathbb{R}^d$ such that $x = \exp(A(\cdot + 1))x^0 + Su$ with $y = Cx + Du$. Therefore, $\widetilde{\mathcal{B}}$ is the direct sum of the finite-dimensional space

$$\widetilde{\mathcal{B}}_0 := \{\mathrm{col}(0, C\exp(A(\cdot + 1))x^0) \,|\, x^0 \in \mathbb{R}^n\}$$

and

$$\widetilde{\mathcal{B}}_1 := \{\mathrm{col}(u, (CS + D)u) \,|\, u \in L^2(\mathcal{I}, \mathbb{R}^m)\}. \tag{10}$$

The space $\widetilde{\mathcal{B}}_1$ is closed in $L^2(\mathcal{I}, \mathbb{R}^{m+p})$ as it is the graph of the bounded linear operator $(CS + D) : L^2(\mathcal{I}, \mathbb{R}^m) \to L^2(\mathcal{I}, \mathbb{R}^p)$. This shows the closedness.

The assertion follows with $\mathcal{B} \cap \mathcal{C}^\infty(\mathcal{I}, \mathbb{R}^{m+p}) = \widetilde{\mathcal{B}} \cap \mathcal{C}^\infty(\mathcal{I}, \mathbb{R}^{m+p})$ and a density argument. □

Remark 4. Obtaining a kernel representation (2) from an input-state-output one can be achieved by *elimination* of the state variable x, see Section 6.2.2 of [15].

In this paper we use a couple of integer system invariants. The *McMillan degree* of $\mathcal{B}$, denoted $\mathfrak{n}(\mathcal{B})$, is the minimal dimension of the state space among all possible input-state-output representations (7) of $\mathcal{B}$. If the state space dimension equals $\mathfrak{n}(\mathcal{B})$, this particular input-state-output representation is said to be *minimal*. Define

$$\mathcal{O}_k := \begin{cases} C & \text{if } k = 0 \\ \begin{bmatrix} \mathcal{O}_{k-1} \\ CA^k \end{bmatrix} & \text{if } k \geq 1 \end{cases}; \tag{11}$$

the *system lag*, denoted $\mathfrak{l}(\mathcal{B})$, is defined by

$$\mathfrak{l}(\mathcal{B}) := \min\{k \in \mathbb{N} \mid \mathrm{rank}\ \mathcal{O}_k = \mathrm{rank}\ \mathcal{O}_{k-1}\}.$$

Evidently, $\mathfrak{l}(\mathcal{B}) \leq \mathfrak{n}(\mathcal{B})$. Further, $\mathfrak{l}(\mathcal{B})$ is the highest order of differentiation in a "shortest lag" description of $\mathcal{B}$, see pp. 569-570 of [17].

Given an input-state-output representation of $\mathcal{B}$ the state variable x is called *observable*, if it can be recovered from the input-output trajectory, i.e., there is a polynomial matrix $F \in \mathbb{R}^{n\times q}[s]$ such that for all $w = \mathrm{col}(u, y) \in \mathcal{B} \cap \mathcal{C}^\infty(\overline{\mathcal{I}}, \mathbb{R}^q)$

$$x = F\left(\frac{\mathrm{d}}{\mathrm{d}t}\right) w. \tag{12}$$

Observability of the state variable is equivalent to (A, C) being observable in the usual sense (see e.g. [18]), which is satisfied if $\mathcal{O}_{(n-1)} = n$. For an observable pair (A, C), the lag $\mathfrak{l}(\mathcal{B})$ is the observability index of the pair.

In a manner akin to observability, a connection between behavioral controllability and controllability of the pair (A, B) can be established in terms of input-state-output representations, see, e.g., [18].

Lemma 5. *Suppose $\mathcal{B}$ is controllable and consider a minimal input-state-output representation (7) of $\mathcal{B}$. Then (A, B) is controllable and (A, C) is observable.*

Proof. An input-state-output representation of $\mathcal{B}$ is minimal if and only if (A, C) is observable the input-state-output representation is *state trim*, i.e., for all $x^0 \in \mathbb{R}^n$ there is $\mathrm{col}(u, y) \in \mathcal{B} \cap \mathcal{C}^\infty(\overline{\mathcal{I}}, \mathbb{R}^q)$ and $x \in \mathcal{C}^\infty(\overline{\mathcal{I}}, \mathbb{R}^n)$ such that (7) and $x(0) = x^0$ hold, cf. [16]. We only need to show the controllability of (A, B), that is for arbitrary initial value $x^0 \in \mathbb{R}^n$ and terminal value $x^1 \in \mathbb{R}^n$ there is a control input u and a time instance $t_1 \in (0, 1)$ such that the state solution x of $\frac{\mathrm{d}}{\mathrm{d}t}x = Ax + Bu$ satisfies $x(0) = x^0$ and $x(t_1) = x^1$. By state trimness we find $w_0 = \mathrm{col}(u_0, y_0)$, $w_1 = \mathrm{col}(u_1, y_1) \in \mathcal{B} \cap \mathcal{C}^\infty(\overline{\mathcal{I}}, \mathbb{R}^q)$ and corresponding states $x_0, x_1 \in \mathcal{C}^\infty(\overline{\mathcal{I}}, \mathbb{R}^n)$ with $x_0(0) = x^0$ and $x_1(0) = x^1$. Since $\mathcal{B}$ is controllable, there further exists $w = \mathrm{col}(u, y) \in \mathcal{B} \cap \mathcal{C}^\infty(\overline{\mathcal{I}}, \mathbb{R}^n)$ and $t_1 \in (0, 1)$ satisfying (5), cf. Theorem 5.2.9 in [15]. With the observability of the state (12) we see

$$\begin{aligned} x(0) &= \left(F(\tfrac{\mathrm{d}}{\mathrm{d}t})w\right)(0) = \left(F(\tfrac{\mathrm{d}}{\mathrm{d}t})w_0\right)(0) = x(0) = x^0, \\ x(t_1) &= \left(F(\tfrac{\mathrm{d}}{\mathrm{d}t})w\right)(t_1) = \left(F(\tfrac{\mathrm{d}}{\mathrm{d}t})w_1\right)(0) = x_1(0) = x^1. \end{aligned}$$ □

For a linear differential behavior controllability is equivalent to the existence of an *image representation* (see Theorem 6.6.1 p. 229 in [15], i.e., there exists a polynomial matrix $M \in \mathbb{R}^{(m+p)\times m}[s]$ such that $w \in \mathcal{B} \cap \mathcal{C}^\infty(\overline{\mathcal{I}}, \mathbb{R}^{m+p})$ if and only if there exists a *latent variable* trajectory $\ell \in \mathcal{C}^\infty(\overline{\mathcal{I}}, \mathbb{R}^m)$ such that

$$w = M\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)\ell. \tag{13}$$

A flat output of a differentially flat system leads to a specific image representation endowed with helpful characteristics.

Lemma 6. *Suppose that $\mathcal{B}$ is controllable. Then there exists $M \in \mathbb{R}^{q\times m}[s]$ with $\deg(M) \le \mathfrak{n}(\mathcal{B}) + 1$ such that*

$$\mathcal{B} = \left\{ \mathrm{col}(u, y) \in L^2(\mathcal{I}, \mathbb{R}^{m+p}) \;\middle|\; \begin{array}{l} \exists\, \ell \in L^2(\mathcal{I}, \mathbb{R}^m) \\ s.t.\ (13)\ holds \end{array} \right\}. \tag{14}$$

Moreover, given $w \in H^k(\mathcal{I}, \mathbb{R}^q)$ for some $k \in \mathbb{N}$ the latent variable in (13) satisfies $\ell \in H^{k+1}(\mathcal{I}, \mathbb{R}^m)$.

Proof. We consider a minimal input-state-output representation (7) of $\mathcal{B}$, that is $n = \mathfrak{n}(\mathcal{B})$ and (A, B) is controllable. Then there is a *flat output* defined by

$$\ell = \widetilde{C}x \tag{15}$$

with output matrix $\widetilde{C} \in \mathbb{R}^{m\times n}$, see pp. 84–ff. in [19] and Remark 2 p. 72 of [20]. In more detail, there are polynomial matrices $X \in \mathbb{R}^{n\times m}[s]$, $U \in \mathbb{R}^{m\times m}[s]$ with $\deg(X) \le n$ and $\deg(U) \le n + 1$ such that, given $w = \mathrm{col}(u, y) \in \mathcal{B} \cap \mathcal{C}^\infty(\overline{\mathcal{I}}, \mathbb{R}^q)$ with corresponding state $x \in \mathcal{C}^\infty(\overline{\mathcal{I}}, \mathbb{R}^n)$,

$$x = X(\tfrac{\mathrm{d}}{\mathrm{d}t})\ell, \quad u = U(\tfrac{\mathrm{d}}{\mathrm{d}t})\ell, \quad y = (CX(\tfrac{\mathrm{d}}{\mathrm{d}t}) + DU(\tfrac{\mathrm{d}}{\mathrm{d}t}))\ell.$$

The associated image representation (13) is established through

$$M = \begin{bmatrix} U \\ CX + DU \end{bmatrix} \tag{16}$$

and the flat output ℓ serves as latent variable.

We derive (14) for this particular M. Denote the set on the right side of (14) by $\widetilde{\mathcal{B}}$. We show that $\widetilde{\mathcal{B}}$ is closed in $L^2(\mathcal{I}, \mathbb{R}^q)$. Let $(\mathrm{col}(u_k, y_k))_{k\in\mathbb{N}}$ be a sequence in $\widetilde{\mathcal{B}}$ which converges in $L^2(\mathcal{I}, \mathbb{R}^q)$ to some $\mathrm{col}(u, y)$. The state and latent variable corresponding to $\mathrm{col}(u_k, y_k)$ are denoted by x_k and ℓ_k, respectively. As the latent variable is given via a flat output (see (15)) we find

$$\ell_k = \widetilde{C}x_k = \widetilde{C}Su_k$$

where S is the solution operator defined in the proof of Lemma 3. The convergence of $(u_k)_{k\in\mathbb{N}}$ and the boundedness of S imply that $(\ell_k)_{k\in\mathbb{N}}$ converges to some $\ell \in L^2(\mathcal{I}, \mathbb{R}^m)$. With Lemma 2 we know that

$$\mathcal{B}_\ell = \{\mathrm{col}(u, y, \ell) \in L^2(\mathcal{I}, \mathbb{R}^{q+m}) \,|\, (13) \text{ holds}\}$$

is closed in $L^2(\mathcal{I}, \mathbb{R}^{q+m})$. Therefore, $\mathrm{col}(u, y, \ell) \in \mathcal{B}_\ell$ and $\mathrm{col}(u, y) \in \widetilde{\mathcal{B}}$. Since $\widetilde{\mathcal{B}} \cap \mathcal{C}^\infty(\overline{\mathcal{I}}, \mathbb{R}^q)$ and $\mathcal{B} \cap \mathcal{C}^\infty(\overline{\mathcal{I}}, \mathbb{R}^q)$ coincide, a density argument yields (14).

Moreover, if $\mathrm{col}(u, y) \in H^k(\mathcal{I}, \mathbb{R}^q)$, then the corresponding state satisfies $x = Su \in H^{k+1}(\mathcal{I}, \mathbb{R}^n)$ and, thus, $\ell = \widetilde{C}x \in H^{k+1}(\mathcal{I}, \mathbb{R}^m)$. □

The flatness-based image representation (16) provides key insights into the smoothness of external trajectories, as determined by the differentiability of latent variables. This understanding is crucial for assessing how effectively polynomial approximations can capture the system behavior, as explored in the following subsections.

2.2. Polynomial lift

Let $(\pi_i)_{i\in\mathbb{R}}$ be the sequence of Legendre polynomials with normalization $\pi_i(1) = 1$, $i \in \mathbb{N}$, which forms an orthogonal basis of $L^2(\mathcal{I}, \mathbb{R})$. Recall that $\pi_0(t) = 1$, $\pi_1(t) = t$ and for $i = 1, \ldots$

$$\pi_{i+1}(t) = \frac{2i+1}{i+1} t\pi_i(t) - i\pi_{i-1}(t)\ .$$

Given $f \in L^2(\mathcal{I}, \mathbb{R})$ there is a unique series expansion

$$f = \sum_{i\in\mathbb{N}} \hat{f}_i \pi_i, \tag{17}$$

where $\hat{f}_i := \langle f, \pi_i\rangle \|\pi_i\|^{-2}$ and $\hat{f} := (\hat{f}_i)_{i\in\mathbb{N}}$. From Bessel's theorem it follows that $\hat{f} \in \ell^2(\mathbb{N}, \mathbb{R})$.

Any function $f \in L^2(\mathcal{I}, \mathbb{R}^d)$ with dimension $d \ge 1$ can be represented with coefficients defined by

$$\hat{f}_i := \sum_{k=0}^{d-1} \frac{\langle f, e_k\pi_i\rangle}{\|\pi_i\|^2} e_k, \tag{18}$$

where $\{e_0, \ldots e_{d-1}\}$ is the canonical basis of $\mathbb{R}^d$ and $\hat{f} = (f_i)_{i\in\mathbb{N}} \in \ell^2(\mathbb{N}, \mathbb{R}^d)$. We define

$$\Pi : L^2(\mathcal{I}, \mathbb{R}^d) \to \ell^2(\mathbb{N}, \mathbb{R}^d), \quad f \mapsto \hat{f}, \tag{19}$$

which is an isometric isomorphism.

The differential operator $\frac{\mathrm{d}}{\mathrm{d}t}$ on $H^1(\mathcal{I}, \mathbb{R}^d)$ can be represented as an operator $\mathcal{D}$ acting in $\ell^2(\mathbb{N}, \mathbb{R}^d)$, [21, Equation (2.3.18)]. For functions $f \in H^1(\mathcal{I}, \mathbb{R}^d)$ with $\hat{f} = \Pi f$ and $\widehat{f^{(1)}} := \Pi\frac{\mathrm{d}f}{\mathrm{d}t}$ one has

$$(\widehat{f^{(1)}})_i = (\mathcal{D}\hat{f})_i := (2i+1) \sum_{\substack{j=i+1 \\ i+j \text{ odd}}}^{\infty} \hat{f}_j, \quad i \in \mathbb{N}, \tag{20}$$

or equivalently written by means of an infinite matrix

$$\begin{bmatrix} (\widehat{f^{(1)}})_0 \\ (\widehat{f^{(1)}})_1 \\ (\widehat{f^{(1)}})_2 \\ \vdots \end{bmatrix} = \begin{bmatrix} 0 & I & 0 & I & 0 & I & \dots \\ & 0 & 3I & 0 & 3I & 0 & \dots \\ & & 0 & 5I & 0 & 5I & \dots \\ & & & 0 & 7I & 0 & \dots \\ & & & & \ddots & \ddots & \ddots \end{bmatrix} \begin{bmatrix} \hat{f}_0 \\ \hat{f}_1 \\ \hat{f}_2 \\ \vdots \end{bmatrix}. \tag{21}$$

Similarly to the differential operator $\frac{\mathrm{d}}{\mathrm{d}t}$ one can define powers and polynomials of $\mathcal{D}$.

Employing the kernel representation (2) we find the following characterization of the behavior.

Lemma 7 (Behavioral lift)**.** *Let $w \in \mathcal{C}^\infty(\overline{\mathcal{I}}, \mathbb{R}^{m+p})$. Then $w \in \mathcal{B}$ if and only if*

$$R(\mathcal{D})\Pi w = 0. \tag{22}$$

Example 8 (No finite expansion)**.** We show for the linear time-invariant system described by (7) with $A = C = I$ and $B = D = 0$ that its trajectories with nontrivial output have no series expansion involving only finitely many polynomials π_i. Let $\operatorname{col}(u, y) \in \mathcal{B}$ and assume that y has a finite expansion, i.e. there is some $N \in \mathbb{N}$ such that $\hat{y}_i = 0$ for $i \geq N$. It is no restriction to assume $\operatorname{col}(u, y) \in \mathcal{B} \cap \mathcal{C}^\infty(\overline{\mathcal{I}}, \mathbb{R}^q)$. Note that the kernel representation (2) of $\mathcal{B}$ is given via $R(s) = \begin{bmatrix} 0 & (s-1)I \end{bmatrix}$. With Lemma 7 and the definition of $\mathcal{D}$ in (20) we see that

$$\hat{y}_i = (\mathcal{D}\hat{y})_i = (2i+1) \sum_{\substack{j=i+1 \\ i+j \text{ odd}}}^{\infty} \hat{y}_j, \quad i \in \mathbb{N}. \tag{23}$$

The finiteness of the expansion yields that for $i = N-1$ all summands on the right hand side in (23) vanish and, hence, $\hat{y}_{N-1} = 0$. It is not difficult to see that this successively implies $\hat{y}_i = 0$ for all $i \in \mathbb{N}$. Therefore y is trivial. □

In Example 8 we illustrated the case when the dynamics of a particular linear time-invariant system (7) cannot be described by a finite representation in terms of Legendre polynomials. This is a generic situation: the Legendre series representation of exponential functions $e^{\lambda t}$, $\lambda \neq 0$ (generically present in the solution of (7), e.g. in the free response), involve an infinite number of terms (see e.g. p. 39 of [22]). Such considerations lead naturally to working with truncated Legendre expansions of solutions of (7).

2.3. Truncated expansion

In the light of Example 8, we study approximation bounds when considering truncated series of Legendre polynomials. To this end we introduce the orthogonal projection $P_N : L^2(\mathcal{I}, \mathbb{R}^d) \to L^2(\mathcal{I}, \mathbb{R}^d)$ defined by

$$P_N f := \sum_{i<N} \hat{f}_i \pi_i, \tag{24}$$

Note that $\operatorname{im} P_N$ coincides with the N-dimensional space of $\mathbb{R}^d[s]$-polynomials of degree up to $N-1$, which is spanned by $\pi_0, \ldots, \pi_{N-1}$. Since $(\pi_i)_{i\in\mathbb{N}}$ is an orthogonal basis, $P_N f$ converges to f as $N \to \infty$ with respect to the L^2-norm. The speed of convergence is related to the smoothness of f, as we discuss in the following.

Recall that π_i is an eigenfunction corresponding to the ith eigenvalue $\lambda_i := i(i+1)$ of the Sturm–Liouville operator

$$\mathcal{L}f = \ell f := -\frac{\mathrm{d}}{\mathrm{d}t}\left(p\frac{\mathrm{d}}{\mathrm{d}t}f\right), \quad p(t) = (1-t^2),$$

$$D(\mathcal{L}) := \left\{ f \in L^2(\mathcal{I}, \mathbb{C}) \;\middle|\; \begin{aligned} f, pf^{(1)} &\in \mathcal{AC}(\mathcal{I}, \mathbb{R}), \\ \ell &\in L^2(\mathcal{I}, \mathbb{C}), \\ pf^{(1)}(-1) = pf^{(1)}(1) &= 0 \end{aligned} \right\},$$

see [23, Theorem 3.6]. Here, $\mathcal{AC}(\mathcal{I}, \mathbb{C})$ denotes the space of locally absolutely continuous functions from $\mathcal{I}$ to $\mathbb{C}$. Let $\mathcal{L}^s$ for $s \in (0, \infty)$ denote the sth power the self-adjoint operator $\mathcal{L}$, defined via functional calculus, see e.g. Section 5.3 in [24].

Lemma 9. *If $f \in D(\mathcal{L}^s)$ for some $s > 0$, then*

$$\|(I - P_N)f\| = \|\mathcal{L}^s f\| \mathcal{O}(N^{-2s}) \qquad (N \to \infty). \tag{25}$$

Moreover, $H^k(\mathcal{I}, \mathbb{C}) \subset D(\mathcal{L}^{\frac{k}{2}})$ for $k \in \mathbb{N}$.

Proof. Let $\tilde{\pi}_i = \pi_i / \|\pi_i\|$, i.e. $(\tilde{\pi}_i)_{i\in\mathbb{N}}$ form an orthonormal basis in $L^2(\mathcal{I}, \mathbb{C})$. For $f \in D(\mathcal{L}^s)$ we have

$$\begin{aligned} \|(I - P_N)f\|^2 &= \left\| f - \sum_{i<N} \langle f, \tilde{\pi}_i \rangle \tilde{\pi}_i \right\|^2 \\ &= \left\| \sum_{i\geq N} \langle f, \tilde{\pi}_i \rangle \tilde{\pi}_i \right\|^2 = \left\| \sum_{i\geq N} \lambda_i^{-s} \langle f, \mathcal{L}^s \tilde{\pi}_i \rangle \tilde{\pi}_i \right\|^2 \\ &= \left\| \sum_{i\geq N} \lambda_i^{-s} \langle \mathcal{L}^s f, \tilde{\pi}_i \rangle \tilde{\pi}_i \right\|^2. \end{aligned}$$

Since $(\lambda_i)_{i\in\mathbb{N}}$ is an increasing sequence, we find

$$\begin{aligned} \|(I - P_N)f\|^2 &\leq \lambda_N^{-2s} \left\| \sum_{i\geq N} \langle \mathcal{L}^s f, \tilde{\pi}_i \rangle \tilde{\pi}_i \right\|^2 \\ &\leq \big(N(N+1)\big)^{-2s} \|\mathcal{L}^s\|^2, \end{aligned}$$

which shows the first claim.

We show the inclusion $H^k(\mathcal{I}, \mathbb{C}) \subset D(\mathcal{L}^{\frac{k}{2}})$, $k \in \mathbb{N}$. For $k = 0$ there is nothing to prove. For $k = 1$ we find that $D(\mathcal{L}^{\frac{1}{2}}) = \{f \in L^2(\mathcal{I}, \mathbb{C}) \cap \mathcal{AC}(\mathcal{I}, \mathbb{C}) \,|\, \sqrt{p} f^{(1)} \in L^2(\mathcal{I}, \mathbb{C})\}$ by [25, Theorem 6.8.5 (i)]. With the uniform boundedness of p on $[-1, 1]$ this shows $H^1(\mathcal{I}, \mathbb{C}) \subset D(\mathcal{L}^{\frac{1}{2}})$.

We continue with the case of even $k \geq 2$. Let $f \in H^k(\mathcal{I}, \mathbb{C})$. Then it is clear that f, $pf^{(1)} \in \mathcal{AC}(\mathcal{I}, \mathbb{C})$ and $\ell f \in L^2(\mathcal{I}, \mathbb{C})$. Moreover, $f^{(1)}$ is bounded on $[-1, 1]$ as $f^{(1)} \in H^1(\mathcal{I}, \mathbb{C})$. Consequently, $f \in D(\mathcal{L})$ and $\mathcal{L}f \in H^{k-2}(\mathcal{I}, \mathbb{C})$. Repeating this argument yields $f \in D(\mathcal{L}^{\frac{k}{2}})$, showing $H^k(\mathcal{I}, \mathbb{C}) \subset D(\mathcal{L}^{\frac{k}{2}})$ for even k.

Finally, we consider $f \in H^{k+1}(\mathcal{I}, \mathbb{C})$ for even $k \geq 2$. From the previous observations we know that $f \in D(\mathcal{L}^{\frac{k}{2}})$ and $\mathcal{L}^{\frac{k}{2}} f \in H^1(\mathcal{I}, \mathbb{C}) \subset D(\mathcal{L}^{\frac{1}{2}})$. Therefore, $f \in D(\mathcal{L}^{\frac{k+1}{2}})$, which shows $H^{k+1}(\mathcal{I}, \mathbb{C}) \subset D(\mathcal{L}^{\frac{k+1}{2}})$. □

Corollary 10. *If $f \in H^k(\mathcal{I}, \mathbb{R}^d)$ for some $k \in \mathbb{N}$, then*

$$\|(I - P_N)f\| = \mathcal{O}(N^{-k}) \qquad (N \to \infty). \tag{26}$$

2.4. Polynomial trajectories

Next we show that the space of polynomial trajectories $\mathcal{B} \cap \bigcup_{N\in\mathbb{N}} \operatorname{im} P_N$ is dense in $\mathcal{B} \cap H^s(\mathcal{I}, \mathbb{R}^q)$, $s \in \mathbb{N}$, and, particularly, in $\mathcal{B}$.

Proposition 11. *Suppose $\mathcal{B}$ is controllable and let $w \in \mathcal{B} \cap H^{\mathfrak{n}(\mathcal{B})+s+k}(\mathcal{I}, \mathbb{R}^{m+p})$ for some $s \in \mathbb{N} \setminus \{0\}$, $k \in \mathbb{N}$. For all $N \in \mathbb{N}$, $N \geq \mathfrak{n}(\mathcal{B}) + s + 1$, there is $w^N \in \mathcal{B} \cap \operatorname{im} P_N$ such that*

$$\Lambda_s(w^N - w)(-1) = 0$$

satisfying

$$\|w - w^N\|_{H^{s-1}} = \mathcal{O}(N^{-k}) \qquad (N \to \infty). \tag{27}$$

The integers $k \in \mathbb{N}$ and $s \in \mathbb{N} \setminus \{0\}$ in Proposition 11 determine the convergence order and the highest derivative up to which the asymptotic behavior is valid, cf. (27). These can be considered as user specifiable, provided w is sufficiently smooth.

Proof of Proposition 11. Controllability of $\mathcal{B}$ implies the existence of an image representation (13). Here we consider a particular image representation given by a flat output, see Lemma 6. Let $w \in \mathcal{B} \cap H^{\mathfrak{n}(\mathcal{B})+s+k}(\mathcal{I}, \mathbb{R}^{m+p})$. Then there is $\ell \in H^{\mathfrak{n}(\mathcal{B})+s+k+1}(\mathcal{I}, \mathbb{R}^m)$ such that

$$w = M\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)\ell \tag{28}$$

holds. We construct a polynomial which approximates ℓ and its derivatives up to order $\gamma := \mathfrak{n}(\mathcal{B}) + s + 1$, while matching the initial values. Let $v_\gamma^N := P_{N-\gamma}\ell^{(\gamma)}$. Then by Corollary 10 as $N \to \infty$ one has

$$\|v_\gamma^N - \ell^{(\gamma)}\| = \|(I - P_{N-\gamma})\ell^{(\gamma)}\| = \mathcal{O}(N^{-k}). \tag{29}$$

Define

$$v_i^N(t) := \ell^{(i)}(-1) + \int_{-1}^{t} v_{i+1}^N(\tau)\,\mathrm{d}\tau, \quad i = 0, \ldots, \gamma-1. \tag{30}$$

By construction $(v_0^N)^{(i)} = v_i^N$, $(v_0^N)^{(i)}(-1) = v_i^N(-1) = \ell^{(i)}(-1)$, i.e.

$$\Lambda_\gamma(v_0^N - \ell)(-1) = 0. \tag{31}$$

Moreover, one sees with the Cauchy-Schwarz inequality

$$\begin{aligned}\|\ell^{(i)} - (v_0^N)^{(i)}\|^2 &= \int_{\mathcal{I}} \left| \int_{-1}^{\tau} \ell^{(i+1)}(t) - (v_0^N)^{(i+1)}(t)\,\mathrm{d}t \right|^2 \mathrm{d}\tau \\ &\leq 4\|\ell^{(i+1)} - (v_0^N)^{(i+1)}\|^2\end{aligned}$$

and, thus,

$$\|\ell - v_0^N\|_{H^\gamma} = \mathcal{O}(N^{-k}). \tag{32}$$

By construction $v_0^N \in \operatorname{im} P_N$. Recall that the polynomial matrix M in (28) satisfies $\deg(M) \leq \mathfrak{n}(\mathcal{B}) + 1$. Let $w^N = M(\frac{\mathrm{d}}{\mathrm{d}t})v_0^N$, which is an element of $\mathcal{B} \cap \operatorname{im} P_N$. Now, (31) and (32) yield

$$\Lambda_s(w^N - w)(-1) = \Lambda_s\big(M(\tfrac{\mathrm{d}}{\mathrm{d}t})(v_0^N - \ell)\big)(-1) = 0$$

and

$$\|w^N - w\|_{H^{s-1}} = \|M(\tfrac{\mathrm{d}}{\mathrm{d}t})(v_0^N - \ell)\|_{H^{s-1}} = \mathcal{O}(N^{-k}). \qquad \square$$

3. The LQR problem and its approximation

Our aim is to solve the quadratic optimal control problem

$$\underset{w\in\mathcal{B}\cap H^{\mathfrak{l}(\mathcal{B})}(\mathcal{I},\mathbb{R}^q)}{\text{minimize}} \quad J(w) \quad \text{s.t.} \tag{33a}$$

$$\Lambda_{\mathfrak{l}(\mathcal{B})}(w)(-1) = \xi^0, \tag{33b}$$

with the cost function

$$J(w) := \|y\|^2 + \|u^{(\mathfrak{l}(\mathcal{B}))}\|^2, \quad w = \operatorname{col}(u, y) \in \mathcal{B}. \tag{34}$$

The initial condition (33b) uniquely determines the latent state, provided the latter is observable from the inputs and the outputs. Including the higher-order derivative term of the input into the objective function (34) ensures feasibility of the LQR problem.

Lemma 12. *Problem (33) has a unique solution $w^\star$, and $w^\star \in \mathcal{C}^\infty(\overline{\mathcal{I}}, \mathbb{R}^q)$. Moreover, every feasible trajectory w satisfies*

$$J(w - w^\star) \leq 2\big(J(w) - J(w^\star)\big). \tag{35}$$

Proof. We fix a minimal input-state-output representation (7), that is (A, C) is observable. Consider any $w = \operatorname{col}(u, y) \in \mathcal{B} \cap H^{\mathfrak{l}(\mathcal{B})}$. Then there is $x \in H^1(\mathcal{I}, \mathbb{R}^n)$ satisfying (7). With

$$\mathcal{T}_k := \begin{cases} D & \text{if } k = 0, \\ \begin{bmatrix} D & \\ \mathcal{O}_{k-1}B & \mathcal{T}_{k-1} \end{bmatrix} & \text{if } k \geq 1, \end{cases} \tag{36}$$

with $\mathcal{O}_k$ being the Kalman observability matrix defined in (38) one has

$$\Lambda_{k+1}(y) = \mathcal{O}_k x + \mathcal{T}_k \Lambda_{k+1}(u) \tag{37}$$

and by employing observability

$$x = \mathcal{O}_{\mathfrak{l}(\mathcal{B})-1}^\dagger\big(\Lambda_{\mathfrak{l}(\mathcal{B})}(y) - \mathcal{T}_{\mathfrak{l}(\mathcal{B})-1}\Lambda_{\mathfrak{l}(\mathcal{B})}(u)\big). \tag{38}$$

Inserting (38) into (37) for $k = \mathfrak{l}(\mathcal{B})$ by rearranging terms one obtains a linear auxiliary system

$$\frac{\mathrm{d}}{\mathrm{d}t}\xi = \tilde{A}\xi + \tilde{B}\nu \tag{39}$$

with state $\xi = \Lambda_{\mathfrak{l}(\mathcal{B})}(w)$, input $\nu = u^{(\mathfrak{l}(\mathcal{B}))}$. Note that (33) is equivalent to the LQR problem

$$\underset{\xi,\nu}{\text{minimize}} \int_{\mathcal{I}} \xi(t)^\top \mathcal{Q}\xi(t) + \nu(t)^\top \nu(t)\,\mathrm{d}t,$$

where $\mathcal{Q} = \operatorname{diag}(0, I_p, 0, \ldots, 0)$, subject to the dynamics (39) and the initial condition $\xi(-1) = \xi^0$. By standard LQR theory the latter problem has a solution $(\xi^\star, \nu^\star)$, which is infinitely differentiable as $\nu^\star$ is a state feedback involving a solution of a Riccati differential equation. In particular, $w^\star := \operatorname{diag}(I_m, I_p, 0, \ldots, 0)\xi^\star \in \mathcal{C}^\infty(\overline{\mathcal{I}}, \mathbb{R}^q)$ solves (33).

Let w be any trajectory. It is not difficult to see that J satisfies the parallelogram identity

$$J\big(\tfrac{1}{2}(w^\star + w)\big) + J\big(\tfrac{1}{2}(w^\star - w)\big) = \tfrac{1}{2}J(w^\star) + \tfrac{1}{2}J(w).$$

Convexity of the feasibility region together with $J(w^\star) \leq J(\frac{1}{2}(w^\star + w))$ yield

$$J\big(\tfrac{1}{2}(w^\star - w)\big) \leq \tfrac{1}{2}\big(J(w) - J(w^\star)\big).$$

This shows (35). □

Instead of solving the OCP (33) directly, given $N \in \mathbb{N}$, we solve the problem restricted to polynomial trajectories, i.e.,

$$\underset{w \in \mathcal{B} \cap \operatorname{im} P_N}{\operatorname{minimize}} \; J(w) \quad \text{s.t.} \tag{40a}$$

$$\Lambda_{\mathfrak{l}(\mathcal{B})}(w)(-1) = \xi^0. \tag{40b}$$

Observe that the restriction $w \in \mathcal{B} \cap \operatorname{im} P_N$ enforces polynomial trajectories of degree at most $N-1$. In the following we show that solving (40) leads to an approximately optimal control and, as $N \to \infty$, the optimality gap decays at an polynomial rate of arbitrary order.

Theorem 13 (Convergence of optima)**.** *For given approximation order N, $N \in \mathbb{N}$, let $w^\star$ and w^N be the solutions to the OCPs (33) and (40), respectively. For every $k \in \mathbb{N}$ one has*

$$0 \leq J(w^N) - J(w^\star) = \mathcal{O}(N^{-k}) \tag{41}$$

and

$$\|w^\star - w^N\| = \mathcal{O}(N^{-k}) \tag{42}$$

as $N \to \infty$.

Proof. Recall that the solution $w^\star \in \mathcal{C}^\infty(\overline{\mathcal{I}}, \mathbb{R}^{m+p})$, see Lemma 12. Thus, by Proposition 11 for arbitrary $k \in \mathbb{N}$ there is $v^N \in \mathcal{B} \cap \operatorname{im} P_N$ with $\Lambda_{\mathfrak{l}(\mathcal{B})}(w^\star - v^N)(-1) = 0$ and

$$\|w^\star - v^N\|_{H^{\mathfrak{l}(\mathcal{B})}} = \mathcal{O}(N^{-k}). \tag{43}$$

As $w^\star$ and w^N are solutions of (33) and (40), respectively, one has

$$J(w^\star) \leq J(w^N) \leq J(v^N), \quad N \in \mathbb{N},$$

which shows the left-hand-side inequality in (41). Further, by the reverse triangle inequality

$$\Big|J(v^N)^{\frac{1}{2}} - J(w^\star)^{\frac{1}{2}}\Big| \leq J(v^N - w^\star)^{\frac{1}{2}} \leq \|v^N - w^\star\|_{H^{\mathfrak{l}(\mathcal{B})}},$$

which together with (43) implies

$$\begin{aligned} J(v^N) - J(w^\star) &= 2J(w^\star)^{\frac{1}{2}}\big(J(v^N)^{\frac{1}{2}} - J(w^\star)^{\frac{1}{2}}\big) \\ &\quad + \big(J(v^N)^{\frac{1}{2}} - J(w^\star)^{\frac{1}{2}}\big)^2 \\ &= \mathcal{O}(N^{-k}). \end{aligned}$$

Next we show (42). Let $\operatorname{col}(u^\star, y^\star) = w^\star$ and $\operatorname{col}(u^N, y^N) = w^N$. By $\Lambda_{\mathfrak{l}(\mathcal{B})}(w^\star - w^N)(-1) = 0$. We find

$$\begin{aligned} &\|(u^\star)^{(j)} - (u^N)^{(j)}\|^2 \\ &= \int_{\mathcal{I}} \left| \int_{-1}^{\tau} (u^\star)^{(j+1)}(t) - (u^N)^{(+1)}(t)\,\mathrm{d}t \right|^2 \mathrm{d}\tau \\ &\leq 4\|(u^\star)^{(j+1)} - (u^N)^{(j+1)}\|^2 \end{aligned}$$

for all $j = 0, \ldots, \mathfrak{l}(\mathcal{B}) - 1$. Thus,

$$\|w^\star - w^N\|^2 \leq 4^{\mathfrak{l}(\mathcal{B})} J(w^\star - w^N).$$

This together with (35) in Lemma 12 and (41) yields (43). □

Remark 14. Similar to the above approach, one can handle a cost function given by any *quadratic differential form*, see [26],

$$J(w) = \sum_{i,j=0}^{\mathfrak{l}(\mathcal{B})} (w^{(i)})^\top \Phi_{i,j} w^{(j)}, \tag{44}$$

with matrices $\Phi_{i,j} \in \mathbb{R}^{q \times q}$ such that $\Phi_{i,i} = \Phi_{i,i}^\top$ and

$$\Phi_{\mathfrak{l}(\mathcal{B}),\mathfrak{l}(\mathcal{B})} = \begin{bmatrix} \widetilde{\Phi} & 0 \\ 0 & 0 \end{bmatrix}, \tag{45}$$

where $\widetilde{\Phi} \in \mathbb{R}^{m \times m}$ corresponding to $u^{(\mathfrak{l}(\mathcal{B}))}$ is invertible.

Remark 15. In the case where the state is directly observable at the output, i.e. $C = I_n$ and $D = 0$ in the input-state-output representation (7), Lemma 12 and Theorem 13 likewise apply to the LQR problem

$$\underset{\operatorname{col}(u,x) \in \mathcal{B}}{\operatorname{minimize}} \; \|x\|^2 + \|u\|^2 \quad \text{s.t.} \tag{46a}$$

$$x(-1) = x^0 \tag{46b}$$

and its restrictions to polynomial trajectories.

4. A "fundamental lemma"

The main result of this section is a parametrization of the trajectories of a controllable linear differential system in terms of a constant matrix obtained from "sufficiently-informative" data. To this end, we first define some new concepts and notation and state some preliminary results.

4.1. Persistency of excitation

Given $L \in \mathbb{N} \setminus \{0\}$ and $f \in H^{L-1}(\mathcal{I}, \mathbb{R}^d)$, we define the Gramian

$$\Gamma_L(f) := \int_{\mathcal{I}} \Lambda_L(f)\Lambda_L(f)^\top \,\mathrm{d}t. \tag{47}$$

Definition 16. Let $L \in \mathbb{N}\setminus\{0\}$. A function $f : \mathcal{I} \to \mathbb{R}^d$ is called *persistently exciting of order* L, if $f \in H^{L-1}(\mathcal{I}, \mathbb{R}^d)$ and the Gramian $\Gamma_L(f)$ in (47) is positive definite.

This definition is reminiscent of the notion of *excitation* in [27, Definition 2]. In the following result we relate it to the concept of persistency of excitation used in [28], specifically property (iii) in the lemma below.

Lemma 17. *For $f \in H^{L-1}(\mathcal{I}, \mathbb{R}^d)$ with $L \in \mathbb{N} \setminus \{0\}$, the following statements are equivalent:*

(i) f is persistently exciting of order L;

(ii) $\ker(\Gamma_k(L)) = \{0\}$;

(iii) If $\eta \in \mathbb{R}^{Ld}$ is such that $\eta^\top \Lambda_L(f) = 0$ a.e., then $\eta = 0$;

(iv) The functions $f, f^{(1)}, \dots, f^{(L-1)}$ are linearly independent in $L^2(\mathcal{I}, \mathbb{R}^d)$.

Proof. We show only the equivalence of (ii) and (iii), as the equivalence of (i) and (ii) as well as (iii) and (iv) are straightforward. Observe that $\eta \in \ker(\Gamma_L(f))$ if and only if

$$0 = \eta^\top \Gamma_L(f)\eta = \int_{\mathcal{I}} \|\Lambda_L(f)(t)^\top \eta\|_2^2 \,\mathrm{d}t = \|\Lambda_L(f)^\top \eta\|^2,$$

which shows the equivalence of (ii) and (iii). □

Example 18. Let $f : \mathcal{I} \to \mathbb{R}$ be a monic polynomial of degree $L-1$ with $L \in \mathbb{N}\setminus\{0\}$. Then $f^{(k)}$ for $k = 0, \dots, L-1$ is a polynomial of degree $L - 1 - k$. Therefore, it is clear that $f, f^{(1)}, \dots, f^{(L-1)}$ are linearly independent functions in $L^2(\mathcal{I}, \mathbb{R})$, and, thus, f is persistently exciting of order L by Lemma 17.

Remark 19 (Discrete-time excitation analogue)**.** The above concept of persistency of excitation in continuous time aligns seamlessly with its discrete-time counterpart (see [3]). In discrete time, the time-shift operator serves as the analogue to differentiation, whereas summation corresponds to integration. With this in mind, given a discrete-time signal $f : \{0, \dots, N - 1\} \to \mathbb{R}^d$, we define

$$\Lambda_L(f)(t) := \begin{bmatrix} f^\top(t) & \dots & f^\top(t + L - 1) \end{bmatrix}^\top$$

and consider the Hankel matrix

$$H_L(f) := \begin{bmatrix} \Lambda_L(f)(0) & \dots & \Lambda_L(f)(N - L) \end{bmatrix}.$$

Now, the corresponding Gramian

$$\Gamma_L(f) := \sum_{j=0}^{N-L} \Lambda_L(u)(j)\Lambda_L(u)(j)^\top = H_L(u)H_L(u)^\top$$

is positive definite if and only if $H_L(u)$ has full row rank, i.e. f is persistently exciting of order L, cf. [3].

4.2. Fundamental lemma

In the following, we also need Gramians constructed from input-state trajectories of a system (7). Given $u \in H^{L-1}(\mathcal{I}, \mathbb{R}^m)$ and $x \in H^{K-1}(\mathcal{I}, \mathbb{R}^n)$ for some $L, K \in \mathbb{N} \setminus \{0\}$, we extend the notation of stacked derivatives (1) to

$$\Lambda_{L,K}(u, x) := \begin{bmatrix} \Lambda_L(u) \\ \Lambda_K(x) \end{bmatrix} \tag{48}$$

We define the Gramian $\Gamma_{L,K}(x, u)$ by

$$\Gamma_{L,K}(u, x) := \int_{\mathcal{I}} \Lambda_{L,K}(u, x)\Lambda_{L,K}(u, x)^\top \,\mathrm{d}t\,. \tag{49}$$

We now state some results instrumental to establishing a continuous-time fundamental lemma. To this end, we consider an input-state-output representation (7). Then, for fixed $L \in \mathbb{N} \setminus \{0\}$, we define for $i \in \mathbb{N}$

$$\mathcal{C}_i(A, B) := \begin{cases} I_{n+Lm} & \text{if } i = 0 \\ \begin{bmatrix} A^i & A^{i-1}B & \dots & B & 0 \\ 0 & 0 & \dots & 0 & I_{Lm} \end{bmatrix} & \text{if } i \geq 1 \end{cases}. \tag{50}$$

Lemma 20. *Consider an input-state-output representation (7) of $\mathcal{B}$ and let $\operatorname{col}(u, x)$ be an input-state trajectory with $u \in H^{L+n}(\mathcal{I}, \mathbb{R}^m)$. For $i = 0, \dots, n-1$, the following equalities hold:*

$$\begin{bmatrix} \Lambda_1(x^{(i)}) \\ \Lambda_L(u^{(i)}) \end{bmatrix} = \mathcal{C}_i(A, B) \begin{bmatrix} \Lambda_1(x) \\ \Lambda_{L+i}(u) \end{bmatrix}. \tag{51}$$

Proof. The case $i = 0$ is trivial. To prove (51) in the case $i \geq 1$, use the equation

$$x^{(i)} = A^i x + \sum_{j=0}^{i-1} A^{i-1-j} B u^{(j)}.$$ □

The next result is analogous to [28, Proposition 1]; since the proof needs to be adapted to the language and notation of this paper, we provide it in full detail.

Proposition 21. *Suppose that $\mathcal{B}$ is controllable and consider a minimal input-state-output representation (7) of $\mathcal{B}$ such that (A, B) is controllable. Let $\operatorname{col}(x, u)$ be a input-state trajectory. Assume that u is persistently exciting of order at least $n + L$, with $L \in \mathbb{N} \setminus \{0\}$. Then*

(i) If $\xi \in \mathbb{R}^{Lm+n}$ satisfies

$$\xi^\top \Lambda_{L,1}(u, x) = 0 \tag{52}$$

almost everywhere on $\mathcal{I}$, then $\xi = 0$;

(ii) $\Gamma_{L,1}(u, x)$ is positive definite.

Proof. The second statement follows in a straightforward way from the first one, cf. proof of Lemma 17. We show (i). Let $\xi^\top = \begin{bmatrix} \eta & \zeta \end{bmatrix}$, $\eta = \begin{bmatrix} \eta_0 & \dots & \eta_{L-1} \end{bmatrix}$, with $\eta_k \in \mathbb{R}^{1\times m}$,

$j = 0, \ldots, L-1$, and $\zeta \in \mathbb{R}^{1\times n}$. Differentiating (52) i times, $i = 0, \ldots, n$, we conclude that

$$\begin{bmatrix}\zeta & \eta\end{bmatrix}\begin{bmatrix}\Lambda_1(x^{(i)})\\ \Lambda_L(u^{(i)})\end{bmatrix} = 0\ ,$$

almost everywhere on $\mathcal{I}$ for $i = 0, \ldots, n$.

Using equation (51) established in the proof of Lemma 20, we conclude that for $i = 0, \ldots, n$ it that

$$0 = \begin{bmatrix}\zeta A^i & \ldots & \zeta B & \eta_0 & \ldots & \eta_{L-1}\end{bmatrix}\begin{bmatrix}\Lambda_1(x)\\ \Lambda_{L+i}(u)\end{bmatrix}\ , \tag{53}$$

holds almost everywhere on $\mathcal{I}$. Now define

$$\begin{aligned} w_0 &:= \begin{bmatrix}\zeta & \eta_0 & \ldots & \eta_{L-1} & 0_{nm}\end{bmatrix}\\ w_1 &:= \begin{bmatrix}\zeta A & \zeta B & \eta_0 & \ldots & \eta_{L-1} & 0_{(n-1)m}\end{bmatrix}\\ &\ \vdots\\ w_n &:= \begin{bmatrix}\zeta A^n \ldots & \zeta B & \eta_0 & \ldots & \eta_{L-1}\end{bmatrix}\ .\end{aligned}$$

From (53) we have that the following equations hold true almost everywhere on $\mathcal{I}$:

$$w_i\begin{bmatrix}\Lambda_1(x)\\ \Lambda_{L+n}(u)\end{bmatrix} = 0\ ,\ i = 0, \ldots, n\ . \tag{54}$$

Since u is persistently exciting of order at least $L+n$, using statement 3 of Lemma 17 we conclude that the vector-valued function $\begin{bmatrix}\Lambda_1(x)^\top & \Lambda_{L+n}(u)^\top\end{bmatrix}^\top$ has at most n "almost everywhere annihilators" on $\mathcal{I}$: it follows that the $n+1$ vectors w_i, $i = 0, \ldots, n$ are linearly dependent.

Since the last components of the w_i's are zero, $i = 0, \ldots, n$, we conclude that $\eta_{L-1} = 0_{1\times m}$, then $\eta_{L-2} = 0$, and so on until $\eta_0 = 0$. Consequently

$$\begin{aligned} w_0 &= \begin{bmatrix}\zeta & 0_{1\times(n+L)m}\end{bmatrix}\\ w_1 &= \begin{bmatrix}\zeta A & \zeta B & 0_{1\times(n+L-1)m}\end{bmatrix}\\ w_2 &= \begin{bmatrix}\zeta A^2 & \zeta AB & \zeta B & 0_{(n+L-2)m}\end{bmatrix}\\ &\ \vdots\\ w_n &= \begin{bmatrix}\zeta A^n & \zeta A^{n-1}B & \ldots & \zeta B & 0_{1\times Lm}\end{bmatrix}\ .\end{aligned}$$

Denote by α_i, $i = 0\ldots, n$ the coefficients of the characteristic polynomial of A, and using $\sum_{i=0}^n A^i\alpha_i = 0$ conclude that $\sum_{i=0}^n w_i\alpha_i$ equals

$$\begin{aligned}&\begin{bmatrix}\sum_{i=0}^n \zeta A^i\alpha_i & \sum_{i=1}^n \alpha_i\zeta A^{i-1}B & \ldots & \zeta B & 0_{1\times Lm}\end{bmatrix}\\ &\quad = \begin{bmatrix}0_{1\times n} & \sum_{i=1}^n \zeta\alpha_i A^{i-1}B & \ldots & \zeta B & 0_{1\times Lm}\end{bmatrix}\ .\end{aligned}$$

By construction, almost everywhere on $\mathcal{I}$ it holds that

$$\begin{bmatrix}\sum_{i=1}^n \alpha_i\zeta A^{i-1}B & \ldots & \alpha_n\zeta B\end{bmatrix}\Lambda_n(u) = 0\ ;$$

since u is persistently exciting of order at least $L+n$, we conclude that

$$\begin{bmatrix}\sum_{i=1}^n \alpha_i\zeta A^{i-1}B & \sum_{i=2}^n \alpha_i\zeta A^{i-2}B & \ldots & \alpha_n\zeta B\end{bmatrix} = 0\ .$$

It follows from the last m equations that $\alpha_n\zeta B = 0$; since the highest coefficient α_n of the characteristic polynomial of A equals 1, we conclude that $\zeta B = 0$. The previous m-dimensional block-entry of the vector is $\alpha_{n-1}\zeta B + \alpha_n\zeta AB = 0 + \alpha_n\zeta AB = 0$. We conclude that $\zeta AB = 0$. The same argument can be used to prove $\zeta A^iB = 0$, $i = 0, \ldots, n-1$. Since the pair (A, B) is controllable we conclude that $\zeta = 0$ and consequently that statement (i) is true. □

We now have all the necessary ingredients to formulate a continuous-time "fundamental lemma".

Theorem 22 (Continuous-time "fundamental lemma")**.** *Suppose that $\mathcal{B}$ is controllable. Let $\mathrm{col}(\overline{u}, \overline{y}) \in \mathcal{B}$ be such that $\overline{u}$ is persistently exciting of order $L + \mathfrak{n}(\mathcal{B})$, with $L \geq \mathfrak{l}(\mathcal{B}) + 1$. For $\mathrm{col}(u, y) \in H^{L-1}(\mathcal{I}, \mathbb{R}^q)$ and $K \in \mathbb{N}$, $\mathfrak{l}(\mathcal{B}) + 1 \leq K \leq L$, the following statements are equivalent:*

(i) $\mathrm{col}(u, y) \in \mathcal{B}$;

(ii) There exists $g \in L^2(\mathcal{I}, \mathbb{R}^{Lm+Kp})$ such that

$$\Lambda_{L,K}(u, y) = \Gamma_{L,K}(\overline{u}, \overline{y})g. \tag{55}$$

Moreover, $\mathrm{rank}\,\Gamma_{L,K}(\overline{u}, \overline{y}) = Lm + \mathfrak{n}(\mathcal{B})$.

Proof. Fix a minimal input-state-representation (7) of $\mathcal{B}$ and let

$$\mathcal{S}_{L,K} = \begin{bmatrix} I_{mL} & 0\\ \mathcal{T}_{K-1} & \mathcal{O}_{K-1}\end{bmatrix}, \tag{56}$$

where $\mathcal{O}_K$ is the Kalman observability matrix, see (11), and $\mathcal{T}_K$ is defined as in (36). Then given $\mathrm{col}(u, y) \in \mathcal{B}$ with corresponding state x satisfies

$$\Lambda_{L,K}(u, y) = \mathcal{S}_{L,K}\Lambda_{L,1}(u, x). \tag{57}$$

Let $\overline{x}$ be the state corresponding to $\mathrm{col}(\overline{u}, \overline{y})$. In a first step we show

$$\mathrm{im}\,\mathcal{S}_{L,K} = \mathrm{im}\,\Gamma_{L,K}(\overline{u}, \overline{y}). \tag{58}$$

Note that in order to show (58) it suffices to prove

$$\ker \mathcal{S}_{L,K}^\top = \ker \Gamma_{L,K}(\overline{u}, \overline{y}). \tag{59}$$

The former equality then follows by taking the orthogonal complements of the null spaces and employing the symmetry of $\Gamma_{L,K}(\overline{u}, \overline{y})$. With (57) and (49) one has

$$\Gamma_{L,K}(\overline{u}, \overline{y}) = \mathcal{S}_{L,K}\Gamma_{L,1}(\overline{u}, \overline{x})\mathcal{S}_{L,K}^\top. \tag{60}$$

By Proposition 21 the matrix $\Gamma_{L,1}(\overline{u}, \overline{x})$ is positive definite. This shows (59), cf. Observation 7.1.8 in [29]. In particular,

$$\begin{aligned}\mathrm{rank}\,\Gamma_{L,K}(\overline{u}, \overline{y}) &= \mathrm{rank}\,\mathcal{S}_{L,K} = Lm + \mathrm{rank}\,\mathcal{O}_{K-1}\\ &= Lm + \mathfrak{n}(\mathcal{B}).\end{aligned}$$

We show the implication (i) to (ii). Let $\mathrm{col}(u, y) \in \mathcal{B} \cap H^L(\mathcal{I}, \mathbb{R}^{m+p})$ with state x. Then (57) holds. Therefore, with (58) the function $\Lambda_{L,K}(u, y)$ maps pointwise a.e.

into $\operatorname{im}\Gamma_{L,K}(\overline{u},\overline{y})$ and, thus, $g := \Gamma_{L,K}(\overline{u},\overline{y})^{\dagger}\Lambda_{L,K}(u,y)$ satisfies (55).

We show the converse implication. Assume that (55) holds. We consider a input-output representation (6) of $\mathcal{B}$ with polynomial matrices Q and P. It is no restriction to assume that the degree of Q and P is bounded by $\mathfrak{l}(\mathcal{B})$, i.e. $Q(s) = \sum_{k=0}^{\mathfrak{l}(\mathcal{B})} Q_k s^k$ and $P(s) = \sum_{k=0}^{\mathfrak{l}(\mathcal{B})} P_k s^k$. Define

$$\widetilde{Q} := \begin{bmatrix} Q_0 & \dots & Q_{\mathfrak{l}(\mathcal{B})} & 0_{p\times m(L-\mathfrak{l}(\mathcal{B})-1)} \end{bmatrix} ,$$
$$\widetilde{P} := \begin{bmatrix} P_0 & \dots & P_{\mathfrak{l}(\mathcal{B})} & 0_{p\times p(K-\mathfrak{l}(\mathcal{B})-1)} \end{bmatrix} .$$

Since $\operatorname{col}(\overline{u},\overline{y}) \in \mathcal{B}$, it holds that

$$\begin{bmatrix} \widetilde{Q} & \widetilde{P} \end{bmatrix} \Lambda_{L,K}(\overline{u},\overline{x}) = 0$$

and, consequently,

$$\begin{bmatrix} \widetilde{Q} & \widetilde{P} \end{bmatrix} \Gamma_{L,K}(\overline{u},\overline{x}) = 0 .$$

Therefore,

$$\begin{bmatrix} \widetilde{Q} & \widetilde{P} \end{bmatrix} \Lambda_{L,K}(u,y) = \begin{bmatrix} \widetilde{Q} & \widetilde{P} \end{bmatrix} \Gamma_{L,K}(\overline{u},\overline{x}) g = 0,$$

that is (6) holds and $\operatorname{col}(u,y) \in \mathcal{B}$. □

Remark 23. Instead of using the data matrix $\Gamma_{L,K}(\overline{u},\overline{y})$, any other matrix with the same image is suitable in the description of trajectories (55). One advantageous approach, especially from a numerical perspective, is to utilize the *reduced singular value decomposition* of $\Gamma_L(\overline{u},\overline{y})$, i.e.

$$\Gamma_L(\overline{u},\overline{y}) = U_1 \Sigma_1 V_1^\top ,$$

with Σ_1 nonsingular of dimension equal to the rank of $\Gamma_L(\overline{u},\overline{y})$. Observe that the columns of U_1 form an orthogonal basis for $\operatorname{im}\Gamma_L(\overline{u},\overline{y})$.

In the absence of a feedthrough term (i.e., $D = 0$ in the input-state-output representation), the constraint $K \le L$ in Theorem 22 can be relaxed to $K \le L+1$. If, in addition, the state is directly observable (i.e. $C = I_n$), we get the following statement.

Corollary 24. *Suppose*

$$\mathcal{B} = \left\{ \operatorname{col}(u,x) \in L^2(\mathcal{I},\mathbb{R}^{m+n}) \,\middle|\, \tfrac{\mathrm{d}}{\mathrm{d}} x = Ax + Bu \right\} \quad (61)$$

is controllable. Let $\operatorname{col}(\overline{u},\overline{x}) \in \mathcal{B}$ such that $\overline{u}$ is persistently exciting of order $2+n$. Consider the partition

$$\begin{bmatrix} \Gamma_u \\ \Gamma_x \\ \Gamma_{x^{(1)}} \end{bmatrix} = \Gamma_{1,2}(\overline{u},\overline{x})$$

with $\Gamma_u \in \mathbb{R}^{m\times(m+2n)}$, $\Gamma_x, \Gamma_{x^{(1)}} \in \mathbb{R}^{n\times(m+2n)}$. Then, for $u \in L^2(\mathcal{I},\mathbb{R}^m)$ and $x \in H^1(\mathcal{I},\mathbb{R}^n)$ the following statements are equivalent:

(i) $\operatorname{col}(u,x) \in \mathcal{B}$*;*

(ii) There exists $g \in L^2(\mathcal{I},\mathbb{R}^{m+2n})$ such that

$$u = \Gamma_u g, \quad x = \Gamma_x g, \quad \tfrac{\mathrm{d}}{\mathrm{d}t}(\Gamma_x g) = \Gamma_{x^{(1)}} g. \quad (62)$$

Moreover, $\operatorname{rank}\Gamma_{1,2}(\overline{u},\overline{x}) = m+n$.

Corollary 24 allows for an complete description of $\mathcal{B}$ based only on sufficiently informative data, without knowledge of the system matrices A and B. Note that, however, the verification of condition (55) involves solving a system of linear equations. The solution of a system of linear differential equations (with time-varying coefficients) arises also in the version of the fundamental lemma in [12], see Theorem 2 therein.

4.3. System identification

The data matrix, as applied in the fundamental lemma, enables the reconstruction of behavioral representations. Suppose that the assumptions of Corollary 24 hold. The representation (62) is equivalent to

$$\widetilde{R}(\tfrac{\mathrm{d}}{\mathrm{d}t}) \operatorname{col}(u,x) = \widetilde{M}(\tfrac{\mathrm{d}}{\mathrm{d}t}) g, \quad (63)$$

where $\widetilde{R}$ and $\widetilde{M}$ are polynomial matrices given by

$$\widetilde{R}(s) = \begin{bmatrix} 0 & 0 \\ I_m & 0 \\ 0 & I_n \end{bmatrix}, \; \widetilde{M}(s) = \begin{bmatrix} \Gamma_{x^{(1)}} - s\Gamma_x \\ \Gamma_u \\ \Gamma_x \end{bmatrix} . \quad (64)$$

Observe, that g serves as a latent variable in the representation (63).

We are going to eliminate the latent variable g, cf. Theorem 6.2.6. in [15] Let

$$\begin{bmatrix} \widetilde{B} & \widetilde{A} \end{bmatrix} = \Gamma_{x^{(1)}} \begin{bmatrix} \Gamma_u \\ \Gamma_x \end{bmatrix}^{\dagger} , \quad \widetilde{B} \in \mathbb{R}^{n\times m}, \; \widetilde{A} \in \mathbb{R}^{n\times n} . \quad (65)$$

Note that

$$\operatorname{rank}\Gamma_{1,2}(\overline{u},\overline{x}) = \operatorname{rank}\Gamma_{1,1}(\overline{u},\overline{x}) = m+n,$$

cf. Proposition 21 and Corollary 24, and $\Gamma_{1,1}(\overline{u},\overline{x})$ is a submatrix of $\Gamma_{1,2}(\overline{u},\overline{x})$. Therefore, the rows of $\Gamma_{x^{(1)}}$ are linearly dependent on those of $\begin{bmatrix} \Gamma_u^\top & \Gamma_x^\top \end{bmatrix}^\top$. As a consequence, multiplication with the unimodular matrix U,

$$U(s) = \begin{bmatrix} I_n & -\widetilde{B} & (sI_n - \widetilde{A}) \\ 0 & I_m & 0 \\ 0 & 0 & I_n \end{bmatrix}, \quad (66)$$

yields

$$U(s) \begin{bmatrix} \widetilde{R}(s) & \widetilde{M}(s) \end{bmatrix} = \begin{bmatrix} -\widetilde{B} & (sI_n - \widetilde{A}) & 0 \\ I_m & 0 & \Gamma_u \\ 0 & I_n & \Gamma_x \end{bmatrix} . \quad (67)$$

Finally, using the first n rows in $U(s)\widetilde{R}(s)$, a kernel representation (2) of $\mathcal{B}$ is obtained,

$$R(s) = \begin{bmatrix} -\widetilde{B} & (sI_n - \widetilde{A}) \end{bmatrix} \quad (68)$$

It is not difficult to see that $\widetilde{A}$ and $\widetilde{B}$ (together with $C = I_n$, $D = 0$) are suitable matrices for the input-state-output model of (7).

4.4. Expansion-based formulation

Employing the polynomial lift, see Subsection 2.2, we obtain the following two corollaries of Theorem 22 and Corollary 24, respectively.

Corollary 25. *Let the assumption of Theorem 22 hold. Consider the partition*

$$\begin{bmatrix} \Gamma_u \\ \Gamma_{u^{(1)}} \\ \vdots \\ \Gamma_{u^{(L-1)}} \\ \Gamma_y \\ \Gamma_{y^{(1)}} \\ \vdots \\ \Gamma_{y^{(L-1)}} \end{bmatrix} = \Gamma_{L,K}(\overline{u}, \overline{y}),$$

where $\Gamma_{u^{(j)}} \in \mathbb{R}^{m\times(Lm+Kp)}$, $j = 0,\dots,L-1$, *and* $\Gamma_{y^{(k)}} \in \mathbb{R}^{p\times(Lm+Kp)}$, $k = 0,\dots,K-1$. *For* $\mathrm{col}(u,y) \in H^{L-1}(\mathcal{I},\mathbb{R}^q)$ *with* $\hat{u} = \Pi u$, $\hat{y} = \Pi y$ *the following statements are equivalent:*

(i) $\mathrm{col}(u,y) \in \mathcal{B}$;

(ii) There exists $\hat{g} \in \ell^2(\mathbb{N},\mathbb{R}^{Lm+Kp})$ *such that*

$$\begin{aligned} \hat{u} &= \Gamma_{u^{(0)}}\hat{g}, \\ \hat{y} &= \Gamma_{y^{(0)}}\hat{g}, \\ \mathcal{D}(\Gamma_{u^{(j-1)}}\hat{g}) &= \Gamma_{u^{(j)}}\hat{g}, \quad j = 1,\dots,L-1, \\ \mathcal{D}(\Gamma_{y^{(k-1)}}\hat{g}) &= \Gamma_{y^{(k)}}\hat{g}, \quad k = 1,\dots,K-1, \end{aligned} \tag{69}$$

where $\mathcal{D}$ *is defined as in* (20).

In this case the k*-th derivative of* u *is given by* $u^{(k)} = \sum_{i\in\mathbb{N}}(\Gamma_{u^{(k)}}\hat{g}_i)\pi_i$; *similarly for derivatives of* y.

Corollary 26. *Let the assumption of Corollary 24 hold. Let For* $u \in L^2(\mathcal{I},\mathbb{R}^m)$ *and* $x \in H^1(\mathcal{I},\mathbb{R}^n)$ *with* $\hat{u} = \Pi u$, $\hat{x} = \Pi x$ *the following statements are equivalent:*

(i) $\mathrm{col}(u,y) \in \mathcal{B}$;

(ii) There exists $\hat{g} \in \ell^2(\mathbb{N},\mathbb{R}^{m+2n})$ *such that*

$$\hat{u} = \Gamma_u\hat{g}, \quad \hat{x} = \Gamma_x\hat{g}, \quad \mathcal{D}(\Gamma_x\hat{g}) = \Gamma_{x^{(1)}}\hat{g}, \tag{70}$$

where $\mathcal{D}$ *is defined as in* (20).

Conditions (69) and (70) are formulated in terms of infinite series, meaning that each coefficients $\hat{g}_i$ for $i \in \mathbb{N}$ must satisfy specific linear equations. This complicates numerical computations. Limiting considerations on polynomial trajectories, i.e. $\mathrm{col}(u,y) \in \mathcal{B} \cap \mathrm{im}\, P_N$, this infinite equation system is equivalently reduced to a finite one, assuming $\hat{g}_i = 0$ for all $i \geq N$.

5. Data-driven optimal control

Finally, utilizing the approximation result of Section 3 in conjunction with the fundamental lemma, we propose a data-driven approach for optimal control of input-output systems (see also [30] for a recent application of orthogonal bases of functions in iteratively solving finite-length continuous-time tracking problems).

5.1. Data-driven formulatation

Let the assumptions of Theorem 22 and Corollary 25 (with $K = L = \mathfrak{l}(\mathcal{B}) + 1$) hold. We consider the optimization problem

$$\underset{\hat{g}\in\ell^2(\mathbb{N},\mathbb{R}^{Lq})}{\mathrm{minimize}} \sum_{i<N} \big(\|\Gamma_{y^{(0)}}\hat{g}_i\|_2^2 + \|\Gamma_{u^{(\mathfrak{l}(\mathcal{B}))}}\hat{g}_i\|_2^2\big)\|\pi_i\|^2 \tag{71a}$$

subject to

$$\hat{g}_i = 0, \qquad i \geq N, \tag{71b}$$

$$\mathcal{D}(\Gamma_{u^{(k-1)}}\hat{g}) = \Gamma_{u^{(k)}}\hat{g}, \tag{71c}$$

$$\mathcal{D}(\Gamma_{u^{(k-1)}}\hat{g}) = \Gamma_{u^{(k)}}\hat{g}, \qquad k = 1,\dots,\mathfrak{l}(\mathcal{B}), \tag{71d}$$

$$\xi^0 = \sum_{i<N}(-1)^i \begin{bmatrix} \Gamma_{u^{(0)}} \\ \Gamma_{y^{(0)}} \\ \vdots \\ \Gamma_{u^{(\mathfrak{l}(\mathcal{B})-1)}} \\ \Gamma_{y^{(\mathfrak{l}(\mathcal{B})-1)}} \end{bmatrix} \hat{g}_i. \tag{71e}$$

Note, that the relationship between optimization problems (71) and (40) is established by

$$\begin{aligned} \Pi u = \hat{u} &= \Gamma_{u^{(0)}}\hat{g}, & \Pi y = \hat{y} &= \Gamma_{y^{(0)}}\hat{g}, \\ u^{(k)} &= \sum_{i<N}(\Gamma_{u^{(k)}}\hat{g}_i)\pi_i, & y^{(k)} &= \sum_{i<N}(\Gamma_{y^{(k)}}\hat{g}_i)\pi_i. \end{aligned} \tag{72}$$

Constraints (71c) and (71d) ensure that $w = \mathrm{col}(u,y) \in \mathcal{B}$, while constraint (71b) guarantees $w \in \mathrm{im}\, P_N$. The initial condition $\Lambda_{\mathfrak{l}(\mathcal{B})}(w)(-1) = \xi^0$ is reflected by (71e), where $\pi_i(-1) = (-1)^i$ is used.

The following proposition summarizes the relationship between the polynomially restricted LQR problem (40) and its data-driven formulation (71).

Proposition 27. *Let the assumptions of Theorem 22 and Corollary 25 hold. Then the polynomnially restricted LQR problem (40) and the data-driven LQR problem (71) are equivalent in the sense that* $w = \mathrm{col}(u,y)$ *solves (40) if and only if* $\hat{g}$ *is a solution to (71) such that (72) holds. In particular, their optimal values coincide.*

Proof. Via the relationship (72) the target function in (40) can be equivalently rewritten into that in (71a). Further, Corollary 25 directly yields the equivalence of LQR problem (40) and a modified data-driven formulation of (71), where in the latter problem the constraint (71b) is replaced by the seemingly more restrictive constraint

$$\begin{bmatrix} \Gamma_{u^0} \\ \Gamma_{y^0} \end{bmatrix} \hat{g}_i = 0, \qquad i \geq N. \tag{73}$$

Note that the modified condition (73) in combination with (72) is equivalent to $\operatorname{col}(u,y) \in \operatorname{im} P_N$. Replacing (71b) with (73), however, does not affect the feasibility or optimality of a trajectory $\operatorname{col}(u,y)$ with (72). Indeed, this follows form the fact that (73) together with (71c), (71d) implies

$$\begin{bmatrix} \Gamma_{u^k} \\ \Gamma_{y^k} \end{bmatrix} \hat{g}_i = 0, \qquad i \geq N, \quad k = 0, \ldots, \mathfrak{l}(\mathcal{B}) \tag{74}$$

and $\hat{g}_i$ only appears in the modified problem, when accompanied by Γ_{u^k} or Γ_{y^k}. □

The approximation result in Theorem 13 yields asymptotic bounds on the optimality gap between the data-driven LQR problem (71) and the original LQR problem (33). We emphasize that the allowed polynomial approximation order N does not depend in the persistency of excitation order of the data, that is the same informative data trajectory $(\overline{u}, \overline{y})$ can utilized for different N. Figure 1 illustrates how the various results in this paper integrate to derive a solution to the LQR problem (33) via the data-driven LQR formulation (71).

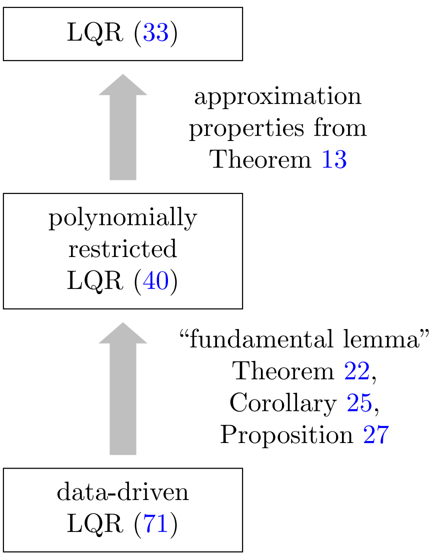

Figure 1: A schematic overview of the relations of the different LQR formulations.

Due to (71b), the optimization problem (71) can be rewritten as a finite-dimensional quadratic program. In this context, $\mathcal{D}$ in constraints (71c) and (71d) is replaced with some upper-left square submatrix of the infinite matrix representation of $\mathcal{D}$ in (21).

Note, that instead of the polynomially restricted LQR problem (40) one could likewise derive a data-driven formulation of the unrestricted LQR problem (33) using Corollary 25. Since the resulting problem does not include a condition like (71b), meaning it involves infinitely many coupled equations, finding a numerical solution, however, seems intractable.

Similarly to the previous approach, consider the scenario described in Corollary 24. The LQR problem (46) in Remark 15, constraint to polynomial trajectories, is equivalent to the data-driven optimization problem

$$\underset{\hat{g} \in \ell^2(\mathbb{N}, \mathbb{R}^{2n+m})}{\text{minimize}} \sum_{i<N} \left(\|\Gamma_x \hat{g}_i\|_2^2 + \|\Gamma_u \hat{g}_i\|_2^2 \right) \|\pi_i\|^2 \tag{75a}$$

subject to

$$\hat{g}_i = 0, \qquad i \geq N, \tag{75b}$$

$$\mathcal{D}(\Gamma_x \hat{g}) = \Gamma_{x^{(1)}} \hat{g}, \tag{75c}$$

$$x^0 = \sum_{i<N} (-1)^i \Gamma_x \hat{g}_i, \tag{75d}$$

cf. Corollary 26 and the proof of Proposition 27.

5.2. Numerical example

We illustrate the numerical feasibility of the data-driven optimal control scheme involving the fundamental lemma consider the LQR

$$\underset{\operatorname{col}(u,x)}{\text{minimize}} \int_{-1}^{1} |u(t)|^2 + |x(t)|^2 \, \mathrm{d}t \tag{76a}$$

$$\tfrac{\mathrm{d}}{\mathrm{d}t} x = -x + u, \quad x(-1) = 1\,. \tag{76b}$$

By Pontryagin's minimum principle the optimal trajectory $\operatorname{col}(u^\star, x^\star)$ to (76) together with its co-variable $\lambda^\star$ satisfies

$$\begin{aligned} \tfrac{\mathrm{d}}{\mathrm{d}t} x^\star &= -x^\star + u^\star, \quad x^\star(-1) = 1 \\ \tfrac{\mathrm{d}}{\mathrm{d}t} \lambda^\star &= \lambda^\star - x^\star, \quad \lambda^\star(1) = 0 \\ u^\star &= -\lambda^\star \end{aligned}$$

and one finds

$$x^\star(t) = \alpha \mathrm{e}^{-\sqrt{2}t} \frac{(\sqrt{2}-2)\mathrm{e}^{2\sqrt{2}t} - (\sqrt{2}+2)\mathrm{e}^{2\sqrt{2}}}{\sqrt{2}(\mathrm{e}^{2\sqrt{2}} - 1)}$$

$$u^\star(t) = -\alpha \mathrm{e}^{-\sqrt{2}t} \frac{\mathrm{e}^{2\sqrt{2}t} - \mathrm{e}^{2\sqrt{2}}}{\mathrm{e}^{2\sqrt{2}} - 1}$$

with a normalization constant α to ensure $x^\star(-1) = 1$. The optimal value is $J^\star \approx 0.4125$.

Note, that the underlying system has McMillan degree $\mathfrak{n}(\mathcal{B}) = 1$ and lag $\mathfrak{l}(\mathcal{B}) = 1$. We consider the trajectory $\operatorname{col}(\overline{u}, \overline{x})$,

$$\overline{u}(t) = t^2, \quad \overline{x}(t) = t^2 - 2t - 5\mathrm{e}^{-(t+1)} + 2, \tag{77}$$

where $\overline{u}$ is persistently exciting of order 3, see Example 18. The smallest eigenvalue of $\Gamma_3(\overline{u})$ is approximately 0.1729. We numerically solve the polynomially restricted optimal control problem, cf. (40), in its data-driven formulation (75) for different polynomial orders N. The resulting time-domain trajectories reconstructed from the expansion coefficients are illustrated in Figure 2. The deviations between the optimal value $J^\star$ and the optima of the data-driven problems are presented in Table 1. The numerical results align with the theoretical convergence order described in Theorem 13. The `Matlab` code that produced the numerical results is available.[2]

[2] https://github.com/schmitzph/contDdOC

N	$J^N - J^\star$	N	$J^N - J^\star$
1	$3.59 \cdot 10^{0}$	6	$9.58 \cdot 10^{-7}$
2	$4.11 \cdot 10^{-1}$	7	$1.25 \cdot 10^{-8}$
3	$3.36 \cdot 10^{-2}$	8	$1.30 \cdot 10^{-10}$
4	$1.70 \cdot 10^{-3}$	9	$9.72 \cdot 10^{-13}$
5	$4.79 \cdot 10^{-5}$	10	$1.73 \cdot 10^{-14}$

Table 1: The error between the optimal value $J^\star$ and the optimal value $J^N = J(\mathrm{col}(u^N, x^N))$ of the data-driven LQR problem with respect to polynomial trajectories in $\mathrm{im}\, P_N$.

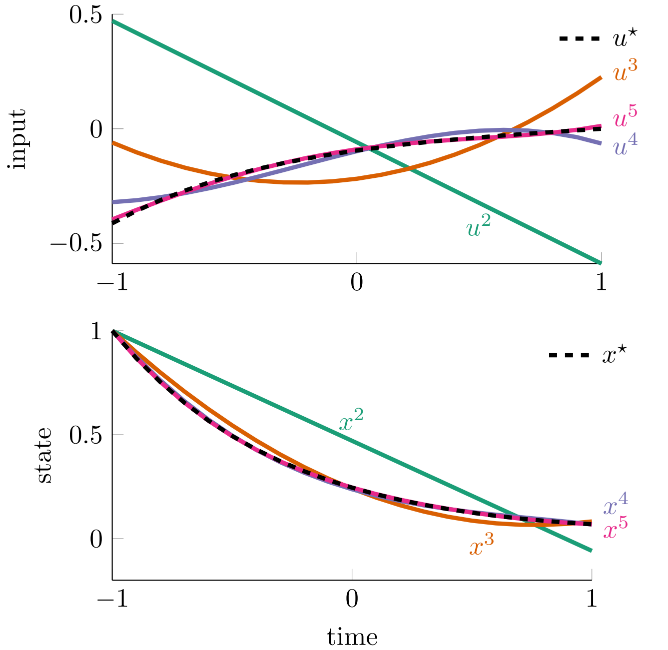

Figure 2: The optimal trajectory $w^\star = \mathrm{col}(u^\star, x^\star)$ (dashed, black) and approximate optimal trajectories $w^N = \mathrm{col}(u^N, x^N)$ for $N = 2, 3, 4, 5$.

6. Conclusions

We stated Gramian-based continuous-time versions of Willems et al.'s fundamental lemma in Theorem 22 and Corollary 24 in the case of input-output and input-state measurements, respectively. Then, we applied the derived results to the data-driven simulation problem in Corollaries 25 and 26.

The evaluation of the performance of our approach in the case of noisy data is of pressing importance. The extension of our approach to the nonlinear case, at least for specific classes of systems, is also a matter of pressing research, especially in the light of recent nonlinear extension of the discrete-time fundamental lemma, see [31, 32, 33, 34]. For more general classes of nonlinear systems, one may invoke recent results on the approximation error for the Koopman generator [35] and operator [36], which may, then, also be used for data-driven predictive control, see, e.g., [37, 38].

References

[1] R. Sepulchre, "Driven by Data: Part One of Two," *IEEE Control Systems Magazine*, vol. 43, no. 5, pp. 4–7, 2023.

[2] ——, "Data-Driven Control, Part II," *IEEE Control Systems Magazine*, vol. 43, no. 6, pp. 5–7, 2023.

[3] J. C. Willems, P. Rapisarda, I. Markovsky, and B. L. De Moor, "A note on persistency of excitation," *Systems & Control Letters*, vol. 54, no. 4, pp. 325–329, 2005.

[4] T. Faulwasser, R. Ou, G. Pan, P. Schmitz, and K. Worthmann, "Behavioral theory for stochastic systems? a data-driven journey from Willems to Wiener and back again," *Annual Reviews in Control*, vol. 55, pp. 92–117, 2023.

[5] J. Coulson, J. Lygeros, and F. Dörfler, "Data-enabled predictive control: In the shallows of the deepc," in *18th IEEE European Control Conference (ECC)*, 2019, pp. 307–312.

[6] C. De Persis and P. Tesi, "Formulas for data-driven control: Stabilization, optimality, and robustness," *IEEE Transactions on Automatic Control*, vol. 65, no. 3, pp. 909–924, 2019.

[7] I. Markovsky and F. Dörfler, "Behavioral systems theory in data-driven analysis, signal processing, and control," *Annual Reviews in Control*, vol. 52, pp. 42–64, 2021.

[8] P. Schmitz, A. Engelmann, T. Faulwasser, and K. Worthmann, "Data-driven mpc of descriptor systems: A case study for power networks," *IFAC-PapersOnLine*, vol. 55, no. 30, pp. 359–364, 2022.

[9] D. Bilgic, A. Koch, G. Pan, and T. Faulwasser, "Toward data-driven predictive control of multi-energy distribution systems," *Electric Power Systems Research*, vol. 212, p. 108311, 2022.

[10] I. Markovsky, L. Huang, and F. Dörfler, "Data-driven control based on the behavioral approach: From theory to applications in power systems," *IEEE Control Systems Magazine*, vol. 43, no. 5, pp. 28–68, 2023.

[11] C. Verhoek, H. S. Abbas, and R. Tóth, "Direct data-driven LPV control of nonlinear systems: An experimental result," *IFAC-PapersOnLine*, vol. 56, no. 2, pp. 2263–2268, 2023.

[12] V. G. Lopez and M. A. Müller, "On a continuous-time version of Willems' lemma," in *61st IEEE Conference on Decision and Control (CDC)*, 2022, pp. 2759–2764.

[13] V. G. Lopez, M. A. Müller, and P. Rapisarda, "An Input-Output Continuous-Time Version of Willems' Lemma," *IEEE Control Systems Letters*, vol. 8, pp. 916–921, 2024.

[14] P. Rapisarda, H. van Waarde, and M. Çamlibel, "A "fundamental lemma" for continuous-time systems, with applications to data-driven simulation," *Systems & Control Letters*, vol. 179, p. 105603, 2023.

[15] J. Polderman and J. Willems, *Introduction to mathematical systems theory: a behavioral approach, Texts in Applied Mathematics 26*. Berlin: Springer-Verlag, 1997.

[16] P. Rapisarda and J. C. Willems, "State maps for linear systems," *SIAM Journal on Control and Optimization*, vol. 35, no. 3, pp. 1053–1091, 1997.

[17] J. C. Willems, "From time series to linear system—Part I. Finite dimensional linear time invariant systems," *Automatica*, vol. 22, no. 5, pp. 561–580, 1986.

[18] H. L. Trentelman, A. A. Stoorvogel, and M. Hautus, *Control theory for linear systems*, ser. Commun. Control Eng. London: Springer, 2001.

[19] H. Sira-Ramirez and S. K. Agrawal, *Differentially flat systems*. Crc Press, 2018.

[20] J. Lévine and D. Nguyen, "Flat output characterization for linear systems using polynomial matrices," *Systems & Control Letters*, vol. 48, no. 1, pp. 69–75, 2003.

[21] C. Canuto, M. Y. Hussaini, A. Quarteroni, and T. A. Zang, *Spectral methods: fundamentals in single domains*. Springer Science & Business Media, 2007.

[22] L. A. Lyusternik, O. A. Chervonenkis, and A. R. Yanpol'skii, *Handbook for Computing Elementary Functions*. Pergamon Press, Oxford, 1965.

[23] L. L. Littlejohn and A. Zettl, "The Legendre equation and its self-adjoint operators," *Electronic Journal of Differential Equations*, vol. 2011, no. 69, pp. 1–33, 2011.

[24] K. Schmüdgen, *Unbounded self-adjoint operators on Hilbert space*. Springer Science & Business Media, 2012, vol. 265.

[25] J. Behrndt, S. Hassi, and H. De Snoo, *Boundary value problems, Weyl functions, and differential operators*. Springer Nature, 2020.

[26] J. C. Willems and H. L. Trentelman, "On quadratic differential forms," *SIAM Journal on Control and Optimization*, vol. 36, no. 5, pp. 1703–1749, 1998.

[27] I. Mareels, "Sufficiency of excitation," *Systems & control letters*, vol. 5, no. 3, pp. 159–163, 1984.

[28] P. Rapisarda, M. K. Çamlibel, and H. J. Van Waarde, "A persistency of excitation condition for continuous-time systems," *IEEE Control Systems Letters*, vol. 7, pp. 589–594, 2022.

[29] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge University Press, Oct. 2012.

[30] B. Chu and P. Rapisarda, "Data-driven iterative learning control for continuous-time systems," in *62nd IEEE Conference on Decision and Control (CDC)*, 2023, pp. 4626–4631.

[31] J. Berberich, J. Köhler, M. A. Müller, and F. Allgöwer, "Data-driven model predictive control with stability and robustness guarantees," *IEEE Transactions on Automatic Control*, vol. 66, no. 4, pp. 1702–1717, 2021.

[32] M. Alsalti, V. G. Lopez, J. Berberich, F. Allgöwer, and M. A. Müller, "Data-driven nonlinear predictive control for feedback linearizable systems," *IFAC-PapersOnLine*, vol. 56, no. 2, pp. 617–624, 2023.

[33] O. Molodchyk and T. Faulwasser, "Exploring the links between the fundamental lemma and kernel regression," *IEEE Control Systems Letters*, 2024.

[34] M. Lazar, "Basis functions nonlinear data-enabled predictive control: Consistent and computationally efficient formulations," *arXiv preprint arXiv:2311.05360*, 2023.

[35] F. Nüske, S. Peitz, F. M. Philipp, M. Schaller, and K. Worthmann, "Finite-data error bounds for Koopman-based prediction and control," *Journal of Nonlinear Science*, vol. 33, 14, 2023.

[36] F. Köhne, F. M. Philipp, M. Schaller, A. Schiela, and K. Worthmann, "L^∞-error bounds for approximations of the Koopman operator by kernel extended dynamic mode decomposition," *arXiv preprint arXiv:2403.18809*, 2024.

[37] M. Korda and I. Mezić, "Linear predictors for nonlinear dynamical systems: Koopman operator meets model predictive control," *Automatica*, vol. 93, pp. 149–160, 2018.

[38] L. Bold, L. Grüne, M. Schaller, and K. Worthmann, "Data-driven MPC with stability guarantees using extended dynamic mode decomposition," *IEEE Transactions on Automatic Control*, 2024, DOI: 10.1109/TAC.2024.3431169.