Orthogonal polynomial bases for data-driven analysis and control of continuous-time systems

Rapisarda, P.; van Waarde, H. J.; Camlibel, M. K.

Published in:
IEEE Transactions on Automatic Control

DOI:
10.1109/TAC.2023.3321214

Orthogonal polynomial bases for data-driven analysis and control of continuous-time systems

P. Rapisarda, *Member, IEEE*, H.J. van Waarde, and M.K. Çamlibel, *Senior Member, IEEE*

***Abstract*— We use polynomial approximation theory to perform data-driven analysis and control of linear, continuous-time invariant systems. We transform the continuous-time input- and state trajectories into discrete sequences consisting of the coefficients of their orthogonal polynomial bases representations. We show that the dynamics of the transformed input- and state signals and those of the original continuous-time trajectories are described by the same system matrices. We investigate informativity, quadratic stabilization, and $\mathcal{H}_2$-performance problems for continuous-time systems. We deal with the case in which machine-precision accuracy in the representation of continuous-time signals can be achieved from the data using a finite number of basis elements, and the case in which the approximation error is non-negligible.**

***Index Terms*— Continuous-time linear systems, polynomial orthogonal basis, data-driven control, informativity, quadratic stabilization, $\mathcal{H}_2$-performance.**

I. Introduction

Data-driven control is a very active area of research, with considerable attention being given especially to the discrete-time case (see [1]–[8]). Recently, also continuous-time data-driven control problems have been investigated. The standard approach in this case involves sampling of at least the input and state trajectories (see e.g. Section 2.2 of [9]; Section II.A of [10]; Remark 2 p. 913 of [4]; Section 2 of [11]; Section I of [12]; Section I of [13]). It is often assumed that also the state derivative can be *directly measured* (see formula (6.d) of [9]; Section II.A of [10], Remark 2 p. 913 of [4]; Section F of [14]; Section I of [12]; formula (15) of [15]; Remark 3 of [16]; Section I of [13]), even if this is possible only in few practical situations (e.g. mechanical systems). If the state derivative is not directly measured, its values at the sampling instants are *estimated* using *discretization* methods (see e.g. Appendix A of the extended version of [11]). To bound the resulting approximation error, *assumptions* are made on the *system dynamics* (e.g. the intersample behavior, see formula (21) of the extended version of [11]; knowledge of bounds on the norms of the state and the input matrix, see the introduction to Appendix A therein), making the approach not *fully data-driven*.

A realistic data-driven framework for control of continuous-time systems should not require the direct measurement of the state derivative, impossible in most situations; and, at least in the noiseless case, it should not assume insight in the system dynamics in order to design controllers. It should be possible to design continuous-time controllers directly and only from data, with minimal assumptions thus exploiting the opportunities offered by continuous-time control (e.g. using analogue devices to implement the controller, and exploiting well-established design techniques). Our aim in this paper is to show that *polynomial orthogonal bases* (POBs in the following) provide useful tools to build such a framework for data-driven control of continuous-time systems. To the best of the authors' knowledge, only in a couple of recent publications dealing with the stochastic case (see [17], [18]) have POBs been used to solve data-driven control problems; ours is the first use of POBs for data-driven control in a deterministic setting. POBs are widely used in applied mathematics, engineering, and physics (see e.g. [19]–[21]), and in system and control theory (see [22]–[26]). Some features of POBs relevant for data-driven control are the following (formal statements and references for these properties are deferred to Section II, where the basics of POBs are summarized):

1) Numerically accurate and computationally efficient algorithms are available to compute POB representations of signals from their samples.
2) The convergence rate depends on the regularity of the function; approximation error bounds are available for truncated representations.
3) Differentiation can be reformulated as product of the (infinite) vector of representation coefficients with an (infinite) basis-dependent differentiation matrix.
4) Error bounds for the approximation error of the derivative computed using a truncated series are computable directly from the function itself.

A crucial conceptual step in our approach is the *transformation* of the continuous-time input- and state trajectories into *discrete* sequences consisting of the coefficients of their POB representations. We use the transformed input- and state trajectories, and exploit the properties of POBs listed above, to offer the following contributions:

1) We show that the (POB representation of the) state

H.J. van Waarde and M.K. Çamlibel are with the Bernoulli Institute for Mathematics, Computer Science, and Artificial Intelligence, University of Groningen, Nijenborgh 9, 9747 AG, Groningen, The Netherlands. (e-mail: {m.k.camlibel, h.j.van.waarde}@rug.nl).

P. Rapisarda is with the School of Electronics and Computer Science, University of Southampton, Great Britain (e-mail: pr3@ecs.soton.ac.uk).

derivative trajectory can be computed *directly* from the (POB representation of the) state trajectory. If the latter is sufficiently regular, the (POB representation of its) derivative can be computed up to machine precision. Otherwise, an approximation can be computed, with an error bound computable *directly* and *only* from the data.

2) We show that the sequences of coefficients of the input, state and state derivative trajectory satisfy equations involving the *same* system matrices as the continuous-time system. This structural identity is crucial: it allows to analyze system properties and to design controllers *directly* from the input- and state trajectories.
3) We exploit such equivalent dynamical representation involving POB representations to develop an *informativity* point of view (see [27]) for continuous-time systems. Ascertaining system properties such as controllability, stabilizability, and so forth can be done *directly* from data[1].
4) We leverage the matrix S-lemma of [6] to solve quadratic stabilization problems and $\mathcal{H}_2$-performance problems when the approximation error introduced by series truncation is non-negligible. Our solution is *entirely* data-driven and *no a priori knowledge* of the system dynamics is assumed. We also show how to deal with noisy data; in such case some insight into the dynamics of the unknown plant is necessary to solve the problem.

A. Structure of the paper

In the first part of the paper we summarize the foundations of orthogonal polynomial bases, and those properties used in the rest of the paper. In Section II we introduce some basic definitions and properties, and exemplify these for Chebyshev and Legendre bases. In Section III we introduce differential i-s-o representations, an equivalent representation of input-state equations in an orthogonal basis.

In the second part of the paper we consider systems analysis and control problems. We formulate the informativity problem in Section IV. In Section IV-A we study informativity for system identification. In Section IV-B characterize informativity for controllability. In Section IV-C we characterize informativity for state feedback stabilization, when the derivative of the state trajectory can be computed to machine precision. We study the problem of feedback stabilization under non-negligible approximation errors in Section V, and the design of controllers with $\mathcal{H}_2$-performance specifications in Section VI. We summarize our results in Section VII.

B. Notation

We denote by $\mathbb{N}$, $\mathbb{R}$ and $\mathbb{C}$ respectively the set of natural, real and complex numbers, and by $\mathbb{R}[s]$ the ring of polynomials with real coefficients. $\mathbb{R}^n$, respectively $\mathbb{C}^n$, denote the space of n-dimensional vectors with real, respectively complex, entries. $\mathbb{R}^{n\times m}$ denotes the set of $n\times m$ matrices with real entries; $\mathbb{R}^{n\times\infty}$ the set of real matrices with n rows and an infinite number of columns; and $\mathbb{R}^{\infty\times\infty}$ the set of real matrices with an infinite number of rows and columns. The transpose of a matrix M is denoted by $M^\top$, its complex conjugate transpose by M^*, and its pseudoinverse by $M^\dagger$.

We denote by $\mathbb{I}$ an interval (t_0,t_1), with $t_0,t_1\in\mathbb{R}$. We denote by $\mathcal{L}_2(\mathbb{I},\mathbb{R})$ the space of square-integrable real-valued functions defined on $\mathbb{I}$ equipped with the standard inner product $\langle\cdot,\cdot\rangle$. The inner product on $\mathcal{L}_2(\mathbb{I},\mathbb{R})$ defined by a *weight-function* w is denoted by $\langle\cdot,\cdot\rangle_w$:

$$\langle f,g\rangle_w := \int_{\mathbb{I}} f(t)g(t)w(t)dt\ .$$

The notation and definitions extend in a natural way to vector-valued functions. The space of real square-summable sequences is denoted by $\ell_2(\mathbb{N},\mathbb{R})$.

If $f\in\mathcal{L}_2(\mathbb{I},\mathbb{R})$ has a square-integrable derivative, the vector of coefficients of the orthogonal basis representation of $\frac{d}{dt}f$ is denoted by $\widetilde{f}^{(1)}$.

[1]A different approach to informativity for continuous-time system stabilization, based on samples of the input, state and state derivative trajectories, is proposed in [13].

II. Polynomial orthogonal bases

A. Fundamentals of polynomial orthogonal bases

For simplicity of exposition we concentrate on the case of scalar functions; definitions and properties can be straightforwardly generalized to the multivariable case, and are introduced only when necessary.

Let $\mathbb{I}=(t_0,t_1)$, with $t_0,t_1\in\mathbb{R}$; an orthogonal basis for $\mathcal{L}_2(\mathbb{I},\mathbb{R})$ is defined by

1) a set of *basis elements* $b_k\in\mathcal{L}_2(\mathbb{I},\mathbb{R})$, $k\in\mathbb{N}$;
2) a *weight function* $w:\mathbb{I}\to\mathbb{R}$;
3) an *inner product* on $\mathcal{L}_2(\mathbb{I},\mathbb{R})$ defined by

$$\langle f,g\rangle_w := \int_{\mathbb{I}} f(t)g(t)w(t)dt\ ,$$

such that $\langle b_j,b_k\rangle_w=\gamma_{jk}\delta_{j,k}$, $j,k\in\mathbb{N}$, where $\delta_{\cdot,\cdot}$ denotes the Kronecker delta, and $\gamma_{jk}>0$, $j,k=0,\ldots$.

A basis $\{b_k\}_{k\in\mathbb{N}}$ is *complete* if its linear span is dense in $\mathcal{L}_2(\mathbb{I},\mathbb{R})$.

For proofs of the following statements, see Section 6 of [28].

Theorem 1: The following statements are equivalent:

1) $\{b_k\}_{k\in\mathbb{N}}$ is complete;
2) If $f\in\mathcal{L}_2(\mathbb{I},\mathbb{R})$ and $\langle f,b_k\rangle_w=0\ \forall\ k\in\mathbb{N}$, then $f=0$;
3) If $f\in\mathcal{L}_2(\mathbb{I},\mathbb{R})$, there exist unique $\widetilde{f}_k\in\mathbb{R}$, $k\in\mathbb{N}$, such that the sequence $\left\{\sum_{k=0}^N \widetilde{f}_k b_k\right\}_{N\in\mathbb{N}}$ converges in the mean to f; moreover, $\widetilde{f}_k=\langle f,b_k\rangle_w$.

Given a complete orthogonal basis $\{b_k\}_{k\in\mathbb{N}}$ and $f\in\mathcal{L}_2(\mathbb{I},\mathbb{R})$, we call $f=\sum_{k=0}^\infty \widetilde{f}_k b_k$, where $\widetilde{f}_k\in\mathbb{R}$, $k\in\mathbb{N}$ is defined as in statement 3) of Theorem 1, the *orthogonal basis representation* of f. We call $\widetilde{f}_k$ the k*-th coefficient* of f in the orthogonal basis representation.

Let $\{b_k\}_{k\in\mathbb{N}}$ be a complete orthogonal basis on $\mathbb{I}$; then for all $f\in\mathcal{L}_2(\mathbb{I},\mathbb{R})$ Bessel's equality $\int_{\mathbb{I}} f(t)^2w(t)dt=\sum_{k=0}^\infty \widetilde{f}_k^2$ holds (see Theorem 23 in Section 6 of [28]). The following result is a straightforward consequence of completeness and Bessel's equality; it implies that $\lim_{k\to\infty}\widetilde{f}_k=0$.

Proposition 1: If $\{b_k\}_{k\in\mathbb{N}}$ is complete, then Π defined by

$$\begin{aligned}\Pi : \mathcal{L}_2(\mathbb{I},\mathbb{R}) &\rightarrow \ell_2(\mathbb{N},\mathbb{R})\\ f &\rightarrow \widetilde{f}\ ,\end{aligned} \tag{1}$$

is a bijective isometry between $\mathcal{L}_2(\mathbb{I},\mathbb{R})$ and $\ell_2(\mathbb{N},\mathbb{R})$.
If each element b_k of the basis is a polynomial, we call $\{b_k\}_{k\in\mathbb{N}}$ a POB. In the rest of this paper, we only consider POBs.

Example 1 (Chebyshev polynomials): Let $\mathbb{I} = (-1,1)$. The Chebyshev polynomials $C_k \in \mathbb{R}[t]$, $k \in \mathbb{N}$ are defined by (see [29]): $C_0(t) := 1$, $C_1(t) := t$, and $C_{n+1}(t) = 2tC_n(t) - C_{n-1}(t)$, $n \geq 2$. They are orthogonal with respect to the inner product defined by $w(t) = \frac{1}{\sqrt{1-t^2}}$, and form a complete basis for $\mathcal{L}_2(\mathbb{I},\mathbb{R})$. *Shifted* Chebyshev bases can be defined for intervals $\mathbb{I} = (t_0, t_1)$ with $t_0, t_1 \in \mathbb{R}$ by the transformation $t \rightarrow \frac{2}{t_1-t_0}t - \frac{t_1+t_0}{t_1-t_0}$. In the following, we only consider the case $\mathbb{I} = (-1,1)$.

Example 2 (Legendre polynomials): Let $\mathbb{I} = (-1,1)$; the Legendre polynomials $L_k \in \mathbb{R}[t]$, $k \in \mathbb{N}$ are defined by (see e.g. [30]) $L_0(t) = 1$, $L_1(t) = t$, and

$$\frac{d}{dt}L_{n+1} = \frac{d}{dt}L_{n-1} + (2n+1)L_n\ . \tag{2}$$

They are orthogonal in the standard $\mathcal{L}_2(\mathbb{I},\mathbb{R})$ inner product, and form a complete basis for $\mathcal{L}_2(\mathbb{I},\mathbb{R})$. *Shifted* Legendre bases can be defined for intervals $\mathbb{I} = (t_0, t_1)$ with $t_0, t_1 \in \mathbb{R}$ by the transformation $t \rightarrow \frac{2}{t_1-t_0}t - \frac{t_1+t_0}{t_1-t_0}$.

Remark 1 (Computation of coefficients): We substantiate statement 1) in the first list of Section I. For the Chebyshev and Legendre bases, computing the coefficients $\widetilde{f}_k$ does *not* require the evaluation of $\int_{\mathbb{I}} f(t)b_k(t)w(t)dt$: they can be computed by *interpolating* f on an appropriate *sampling grid* (see (3.4) p. 14 of [31], and Section 2.3 of [20], respectively). For the Chebyshev basis, the coefficients can be efficiently computed using the FFT (see Section 3.3 of [31])[2]. Consequently, the number of coefficients that can be computed is only limited by the size of the sampling grid, and by the computational power available to interpolate the associated data.

Remark 2 (Lipschitz continuity and uniform convergence): If the function f is *Lipschitz continuous* and the basis $\{b_k\}_{k\in\mathbb{N}}$ consists of the Chebyshev or Legendre polynomials, then statement 3) of Theorem 1 can be strengthened: the sequence $\left\{\sum_{k=0}^N \widetilde{f}_k b_k\right\}_{N\in\mathbb{N}}$ converges *absolutely* and *uniformly* (see Theorem 3.1 p. 17 in [31]).

Let $f \in \mathcal{L}_2(\mathbb{I},\mathbb{R})$, and $\{b_k\}_{k\in\mathbb{N}}$ be a POB; denote by $\mathfrak{b}$ the infinite vector of functions

$$\mathfrak{b} := \begin{bmatrix} b_0 & b_1 & \dots \end{bmatrix}^\top\ , \tag{3}$$

and by $\widetilde{f}$ the infinite vector defined by

$$\widetilde{f} := \begin{bmatrix} \widetilde{f}_0 & \widetilde{f}_1 & \dots \end{bmatrix}\ . \tag{4}$$

With these positions, we write

$$f = \sum_{k=0}^{\infty} \widetilde{f}_k b_k = \widetilde{f}\mathfrak{b}\ . \tag{5}$$

[2]A description of the process used to compute the coefficients of the Chebyshev expansion up to machine precision is given in pp. 18-20 of [31].

The meaning of this equality depends on the function space that f belongs to (almost everywhere for $f \in \mathcal{L}_2(\mathbb{I},\mathbb{R})$, pointwise for Lipschitz functions). The right-hand side of (5) is the *POB representation* of f, sometimes called also the *polynomial transform* of f (see p. 69 of [20]).

When dealing with *vector* functions, we use the following notation. Denote by f_i, $i = 1,\dots,n$ the i-th component of $f \in \mathcal{L}_2(\mathbb{I},\mathbb{R}^n)$, and let $f_i = \sum_{k=0}^{\infty} \widetilde{f}_{i,k} b_k$ be its POB representation; we write

$$f = \underbrace{\begin{bmatrix} \widetilde{f}_{1,0} & \widetilde{f}_{1,1} & \dots \\ \vdots & \vdots & \dots \\ \widetilde{f}_{n,0} & \widetilde{f}_{n,1} & \dots \end{bmatrix}}_{=:\widetilde{f}} \mathfrak{b}\ . \tag{6}$$

B. Approximation by projection; error bounds

Let $f \in \mathcal{L}_2(\mathbb{I},\mathbb{R})$ be represented by (5); we call

$$\pi_N(f) := \sum_{k=0}^{N} \widetilde{f}_k b_k\ , \tag{7}$$

the *truncation* or *projection* of (the series for) f *to degree* N. We associate with π_N the map

$$\begin{aligned}\Pi_N : \mathcal{L}_2(\mathbb{I},\mathbb{R}) &\rightarrow \mathbb{R}^{1\times(N+1)}\\ f &\rightarrow \begin{bmatrix} \widetilde{f}_0 & \dots & \widetilde{f}_N \end{bmatrix}\ .\end{aligned} \tag{8}$$

In view of statement 3) of Theorem 1 and of Remark 2, we conclude that the *approximation error*, defined by

$$f - \pi_N(f) = \sum_{k=N+1}^{\infty} \widetilde{f}_k b_k\ , \tag{9}$$

decays with N. For the Chebyshev and the Legendre basis, a general principle is that the smoother the function, the faster the approximation error goes to zero with N. Consequently, "well-behaved" functions can be represented up to machine precision by truncated series (see also Remark 1). The following remarks justify the first and second part of statement 2) in the first list in Section I, respectively.

Remark 3 (Rate of convergence): For ν-differentiable functions, the ∞-norm of the approximation error is $O(N^{-\nu})$ ("algebraic convergence"), see Theorem 7.2 p. 53 of [31] and Section 5.4.2 of [20], respectively. For analytic functions, the ∞-approximation error is $O(\rho^{-N})$ for some $0 < \rho < 1$ ("geometric convergence"), see Theorem 8.2 p. 57 of [31] for the Chebyshev basis. For $\mathcal{C}^\infty$-functions, the approximation error goes to zero faster than $O(N^{-k})$ for every finite k ("exponential convergence"), see p. 47 of [20]. Similar results can be established for less regular functions using Sobolev spaces (see Appendix A.11 of [20], and Sections 5.4.2 and 5.5.2 therein for the Legendre and Chebyshev cases, respectively).

Remark 4 (Error bounds): The upper bounds on the approximation error by truncation referred to in Remark 3 are expressed *only* in terms of properties of the function (and its derivatives). For differentiable functions (algebraic convergence), the bound is in terms of the total variation (1-norm) of the derivative of the function (see e.g. formula (7.4) p. 53 of [31]). For analytic functions (geometric convergence) the bound is in terms of the maximum value of the function in a bounded subset of $\mathbb{C}$ ("Bernstein ellipse", see p. 56 of [31]), see formula (8.1) p. 57 of [31]. For less regular functions in Sobolev spaces, see Sections 5.4.2 and 5.5.2 of [20].

C. Differentiation

We justify statement 3) in the first list of Section I. Assume that $f \in \mathcal{L}_2(\mathbb{I},\mathbb{R})$ is differentiable and that $\frac{d}{dt}f \in \mathcal{L}_2(\mathbb{I},\mathbb{R})$. Because of completeness and the unicity of the orthogonal basis representation (statement 3) in Theorem 1), the following *differentiation in the transform space* equality holds:

$$\frac{d}{dt}f = \sum_{k=0}^{\infty} \widetilde{f}_k \ \frac{d}{dt}b_k \ , \tag{10}$$

see p. 77 of [20]. Since b_k is a polynomial, $\frac{d}{dt}b_k$ is also a polynomial, and can be written as linear combination of the basis elements: there exist $d_{k,j} \in \mathbb{R}$ such that

$$\frac{d}{dt}b_k = \sum_{j=0}^{\infty} d_{k,j} b_j \ , \ k \in \mathbb{N} \ . \tag{11}$$

Recall (3) and define the infinite vector

$$\frac{d}{dt}\mathfrak{b} := \begin{bmatrix} \frac{d}{dt}b_0 & \frac{d}{dt}b_1 & \dots \end{bmatrix}^\top \ , \tag{12}$$

and from (11), define the infinite matrix

$$\mathcal{D}_{\mathfrak{b}} := \begin{bmatrix} d_{k,j} \end{bmatrix}_{k,j\in\mathbb{N}} \ . \tag{13}$$

With these positions, (11) can be written as $\frac{d}{dt}\mathfrak{b} = \mathcal{D}_{\mathfrak{b}}\mathfrak{b}$, and (10) is equivalent with

$$\frac{d}{dt}f = \widetilde{f}\frac{d}{dt}\mathfrak{b} \ = \widetilde{f}\mathcal{D}_{\mathfrak{b}}\mathfrak{b} = \underbrace{\begin{bmatrix} \widetilde{f}_0^{(1)} & \widetilde{f}_1^{(1)} & \dots \end{bmatrix}}_{=:\widetilde{f}^{(1)}=\widetilde{f}\mathcal{D}_{\mathfrak{b}}}\mathfrak{b} \ . \tag{14}$$

From Proposition 1 it follows that the operator $\frac{d}{dt}$ on $\mathcal{L}_2(\mathbb{I},\mathbb{R})$ induces an operator $\mathcal{D}_{\mathfrak{b}}$ defined by:

$$\begin{array}{rcl} \mathcal{D}_{\mathfrak{b}} : \ell_2(\mathbb{N},\mathbb{R}) & \to & \ell_2(\mathbb{N},\mathbb{R}) \\ \widetilde{f} & \to & \widetilde{f}\mathcal{D}_{\mathfrak{b}} \ , \end{array} \tag{15}$$

i.e. the POB representation of the derivative of a function is *directly* computed from the POB representation of the function itself.

Example 3 (Differentiation for Legendre polynomials): The following equalities can be proved with an induction argument:

$$\begin{array}{rcll} \frac{d}{dt}L_{2k+1} & = & \sum_{j=0}^{k}(4j+1)L_{2j} & k \geq 0 \ , \\ \frac{d}{dt}L_{2k} & = & \sum_{j=0}^{k-1}(4j+3)L_{2j+1} & k \geq 1 \ , \end{array}$$

(see also formulas (2.3.17)-(2.3.18) p. 77 of [20]). Denote $\mathfrak{L} := \begin{bmatrix} L_0 & L_1 & \dots \end{bmatrix}^\top$; then

$$\frac{d}{dt}\mathfrak{L} = \underbrace{\begin{bmatrix} 0 & 0 & 0 & 0 & 0 & \dots \\ 1 & 0 & 0 & 0 & 0 & \dots \\ 0 & 3 & 0 & 0 & 0 & \dots \\ 1 & 0 & 5 & 0 & 0 & \dots \\ 0 & 3 & 0 & 7 & 0 & \dots \\ 1 & 0 & 5 & 0 & 9 & \dots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots \end{bmatrix}}_{=:\mathcal{D}_{\mathfrak{L}}}\mathfrak{L} \ . \tag{16}$$

The nonzero entries of $\mathcal{D}_{\mathfrak{L}}$ increase linearly with the row- and column-indices. □

Example 4 (Differentiation for Chebyshev polynomials): Define $\mathfrak{C} := \begin{bmatrix} C_0 & C_1 & \dots \end{bmatrix}^\top$. Using formula (2.4.22) p. 87 of [20], it can be proved that the entries of the differentiation matrix for the Chebyshev polynomial basis are: $d_{0,k} = 0$ for all $k \in \mathbb{N}$; if ℓ is even, then $d_{\ell,k} = 2\ell$ if $k < \ell$ is even, 0 otherwise; and if ℓ is odd, then $d_{\ell,0} = \ell$, $d_{\ell,k} = 2\ell$ if $k \leq \ell - 1$ is even, and $d_{\ell,k} = 0$ otherwise (see also formula (2.4.22) p. 87 of [20]). Consequently,

$$\mathcal{D}_{\mathfrak{C}} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & \dots \\ 1 & 0 & 0 & 0 & 0 & \dots \\ 0 & 4 & 0 & 0 & 0 & \dots \\ 3 & 0 & 6 & 0 & 0 & \dots \\ 0 & 8 & 0 & 8 & 0 & \dots \\ 5 & 0 & 10 & 0 & 10 & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix} \ . \tag{17}$$

The nonzero entries of $\mathcal{D}_{\mathfrak{C}}$ increase linearly with the row- and column-indices. □

D. Approximation error for the derivative of the projection

We discuss statement 4) of the first list in Section I. Let $N \in \mathbb{N}$; using the notation introduced in (8), partition

$$\widetilde{f} =: \begin{bmatrix} \Pi_N(f) & \widetilde{f}' \end{bmatrix} \quad \text{and } \widetilde{f}^{(1)} =: \begin{bmatrix} \Pi_N(\frac{d}{dt}f) & \widetilde{f}^{(1)\prime} \end{bmatrix} \ , \tag{18}$$

where $\widetilde{f}', \widetilde{f}^{(1)\prime} \in \mathbb{R}^{1\times\infty}$. Partition $\mathcal{D}_{\mathfrak{b}}$ accordingly:

$$\mathcal{D}_{\mathfrak{b}} =: \begin{bmatrix} \mathcal{D}_{11} & \mathcal{D}_{12} \\ \mathcal{D}_{21} & \mathcal{D}_{22} \end{bmatrix} \ , \tag{19}$$

i.e. $\mathcal{D}_{11} \in \mathbb{R}^{(N+1)\times(N+1)}$, $\mathcal{D}_{12}, \mathcal{D}_{21}^\top \in \mathbb{R}^{(N+1)\times\infty}$, $\mathcal{D}_{22} \in \mathbb{R}^{\infty\times\infty}$. The nonzero coefficients of $\pi_N(\frac{d}{dt}f)$ are the entries of $\Pi_N(f)\mathcal{D}_{11} + \widetilde{f}'\mathcal{D}_{21}$; the nonzero coefficients of $\frac{d}{dt}(\pi_N(f))$ are the entries of $\Pi_N(f)\mathcal{D}_{11}$. It follows that *differentiation and projection do not commute*: $\frac{d}{dt} \circ \pi_N \neq \pi_N \circ \frac{d}{dt}$. We call $\frac{d}{dt}(\pi_N(f)) - \pi_N\left(\frac{d}{dt}f\right)$ the *derivative approximation error*. If $\widetilde{f}_k$ goes to zero quickly *relative* to the linear increase of the entries of $\mathcal{D}_{\mathfrak{b}}$ (see Remark 3 and Examples 3 and 4), a large enough N exists for which both $\widetilde{f}_k$ and $\widetilde{f}_k^{(1)}$ are below machine precision: in such cases $\widetilde{f}$ and $\widetilde{f}^{(1)}$ have compact support for practical purposes. Then $\pi_N(\frac{d}{dt}f) \simeq \frac{d}{dt}(\pi_N(f))$, and the differentiation and truncation operators "almost" commute on f. In all other cases, an upper bound on the norm of the commutator $\pi_N \circ \frac{d}{dt} - \frac{d}{dt} \circ \pi_N$ is needed. The following result

This article has been accepted for publication in IEEE Transactions on Automatic Control. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TAC.2023.3321214

holds for the Chebyshev basis, and it is stated under some technical assumptions discussed in Remark 5; some additional comments are given in Remark 6.

Proposition 2: Let $\{C_k\}_{k\in\mathbb{N}}$ be the Chebyshev basis. Let $f \in \mathcal{L}_2(\mathbb{I},\mathbb{R})$ and let $N \in \mathbb{N}$, $N \geq 1$. Assume that f and $f^{(1)}$ are absolutely continuous, and that $V\left(f^{(2)}\right) := \left\|\frac{d^2 f}{dt^2}\right\|_1 = \int_{-1}^{1}\left|\frac{d^2 f}{dt^2}(\tau)\right| d\tau < \infty$. Partition $\widetilde{f}$ and $\widetilde{f}^{(1)}$ as in (18) and $\mathcal{D}_{\mathfrak{C}}$ as in (19). Then the 2-norm of $\widetilde{f}'\mathcal{D}_{21}$ is $\leq \frac{2V\left(f^{(2)}\right)}{\sqrt{\pi}(N-1)}$.

Proof: Use Theorem 7.2 p. 53 of [31] to conclude that

$$\left\|f^{(1)} - \frac{d}{dt}\left(\pi_N(f)\right)\right\|_\infty \leq \frac{2V\left(f^{(2)}\right)}{\pi(N-1)} .$$

Use Proposition 1 and the partitions (18) and (19) to conclude that $\left\|f^{(1)} - \frac{d}{dt}\left(\pi_N(f)\right)\right\|_{\mathcal{L}_{2,w}}$ equals the ℓ_2-norm of the sequence $\left[\widetilde{f}'\mathcal{D}_{21} \quad \Pi_N(f)\mathcal{D}_{12} + \widetilde{f}'\mathcal{D}_{22}\right]$.

If $g \in \mathcal{L}_{2,w}(\mathbb{I},\mathbb{R})$, then $\|g\|_\infty$ is finite, and

$$\|g\|_{\mathcal{L}_{2,w}(\mathbb{I},\mathbb{R})} \leq \left(\int_{-1}^{1}\|g\|_\infty^2 w(\tau)d\tau\right)^{\frac{1}{2}} \leq \|g\|_\infty\sqrt{\pi} .$$

Consequently,

$$\begin{aligned}\frac{2V\left(f^{(2)}\right)}{\pi(N-1)} &\geq \left\|f^{(1)} - \frac{d}{dt}\left(\pi_N(f)\right)\right\|_\infty \\ &\geq \frac{1}{\sqrt{\pi}}\left\|f^{(1)} - \frac{d}{dt}\left(\pi_N(f)\right)\right\|_{\mathcal{L}_{2,w}(\mathbb{I},\mathbb{R})} .\end{aligned}$$

Observe that $\left\|f^{(1)} - \frac{d}{dt}\left(\pi_N(f)\right)\right\|_{\mathcal{L}_{2,w}(\mathbb{I},\mathbb{R})}$ equals the sum of the 2-norm of $\widetilde{f}'\mathcal{D}_{21}$ and of the ℓ_2-norm of the sequence $\Pi_N(f)\mathcal{D}_{12}+\widetilde{f}'\mathcal{D}_{22}$; since the latter is nonnegative, the desired inequality follows. ■

Remark 5 (On the assumptions of Proposition 2): The assumptions on f and $\frac{d}{dt}f$ being absolutely continuous can be relaxed by working in Sobolev spaces of square-integrable functions possessing a certain number of derivatives in the sense of distributions, see Section 2 of [32], especially Lemma 2.3 p. 75 therein.

Remark 6 (On the error bound of Proposition 2): In Proposition 2 the error bound is expressed *only* in terms of properties of the second-order derivative of f (see also Remark 4 and Section 2 of [32]). Consequently, it can be computed *directly* from the data. In view of the applications we consider in Section V of this paper, the fact that the bound is a truly data-driven one is an advantage over the results of [11] (see Appendix A therein), that require insight in the unknown system dynamics in the form of bounds on $\|A\|_2$ and $\|B\|_2$.

Two issues for further research are to ascertain how tight the error bound is (see e.g. p. 53 of [31]), and how to accurately estimate it. In Example 7 we illustrate one approach to the estimation of the bound: using the first $N+1$ coefficients we compute an approximation to $f^{(2)}$ as $\frac{d^2}{dt^2}\left(\Pi_N(f)\right)$; and we estimate $V(f^{(2)})$ by the total variation of $\frac{d^2}{dt^2}\left(\Pi_N(f)\right)$.

III. Orthogonal basis i-s representations

Consider the continuous-time input-state (i-s) equation

$$\frac{d}{dt}x = Ax + Bu , \tag{20}$$

where $x(t) \in \mathbb{R}^n$, $u(t) \in \mathbb{R}^m$. Since the system (20) is linear, given $u \in \mathcal{L}_2(\mathbb{I},\mathbb{R}^m)$, the trajectories x and $\frac{d}{dt}x$ satisfying (20) are also square-integrable on $\mathbb{I}$. We denote the corresponding POB representations by

$$\widetilde{x} := \{\widetilde{x}_k\}_{k\in\mathbb{N}} , \ \widetilde{x}^{(1)} := \{\widetilde{x}_k^{(1)}\}_{k\in\mathbb{N}} , \ \widetilde{u} := \{\widetilde{u}_k\}_{k\in\mathbb{N}} . \tag{21}$$

The following Theorem is crucial for the rest of the paper.

Theorem 2: Let $\{b_k\}_{k\in\mathbb{N}}$ be a complete POB for $\mathcal{L}_2(\mathbb{I},\mathbb{R})$. Let (20) be an i-s system. The following statements are equivalent:

1) u, x satisfy (20);
2) The sequences defined in (21) satisfy

$$\widetilde{x}^{(1)} = A\widetilde{x} + B\widetilde{u}$$

2a) The sequences defined in (21) satisfy

$$\widetilde{x}\mathcal{D}_{\mathfrak{b}} = A\widetilde{x} + B\widetilde{u} . \tag{22}$$

Proof: The equivalence of statements 2) and $2a$) follows from completeness of $\{b_k\}_{k\in\mathbb{N}}$ and (14). To prove the equivalence of 1) and $2a$), conclude from the completeness of $\{b_k\}_{k\in\mathbb{N}}$ and statement 3) of Theorem 1 that u, x satisfy (20) if and only if the basis representations of $\frac{d}{dt}x$, x, and u satisfy

$$\widetilde{x}\mathcal{D}_{\mathfrak{b}}\mathfrak{b} = A\widetilde{x}\mathfrak{b} + B\widetilde{u}\mathfrak{b} ,$$

from which (22) follows. ■

We call (22) the *POB representation* of (20).

Remark 7 (Why orthogonal bases?): By Theorem 2, we can associate to (20) an *equivalent* representation that involves the *same* matrices A, B of (20). In other words, the dynamics of the original system and the representation of the "transformed" system involving the POB representations of the state and input trajectories are described by the *same* matrices. The linearity of the state equations and of differentiation, and the equivalence (14) between differentiation in the time-domain and matrix multiplication in the orthogonal basis domain were crucial in establishing this result.

We set the stage for data-driven analysis and control of continuous-time systems using POBs. Let $\{b_k\}_{k\in\mathbb{N}}$ be a complete polynomial orthogonal basis; let the data-generating continuous-time system be described by (20). We assume that a *noiseless* (state, input) trajectory pair (x,u) has been measured, and that a finite number $N+1$ of coefficients of its orthogonal basis representation $(\widetilde{x},\widetilde{u})$ has been computed (see Remark 1). From the truncated coefficient sequences $\{\widetilde{x}\}_{k=0,\ldots,N}$ and $\{\widetilde{u}\}_{k=0,\ldots,N}$ we define the matrices

$$\begin{aligned}X &:= \begin{bmatrix}\widetilde{x}_0 & \ldots & \widetilde{x}_N\end{bmatrix} \in \mathbb{R}^{n\times(N+1)} \\ U &:= \begin{bmatrix}\widetilde{u}_0 & \ldots & \widetilde{u}_N\end{bmatrix} \in \mathbb{R}^{m\times(N+1)} .\end{aligned} \tag{23}$$

We compute an *approximation* of the truncation of the orthogonal basis representation of $\frac{d}{dt}x$ as $X\mathcal{D}_{11}$ (see Section II-D, and (19)). Note that since the state trajectory of a linear system consists also of exponential trajectories, that cannot be exactly represented by any finite combination of polynomials, such an approximation error always occurs. Two situations can arise, depending on the approximation error between the derivative of the projection and the projection of the derivative being negligible or not.

The first situation occurs when the trajectory x is sufficiently regular, i.e. its orthogonal basis coefficients decay faster than the linear increase of the entries of $\mathcal{D}$ (see Section II-D). x is smoother than, and $\frac{d}{dt}x$ as smooth as u: regularity of x ultimately depends on the differentiability of u. Since in many cases u can be chosen as an experimental input, this first situation is not necessarily unrealistic[3]. Our data-driven analysis of system properties in Section IV is performed under the assumption of negligible derivative approximation error.

We recognize that the second situation is of great practical importance, and in Sections V and VI we bring our error analysis of Section II-D and the matrix S-lemma of [6], [33] to bear on the problem of designing stabilizing state feedback laws when the derivative approximation error is non negligible. We also discuss the case when the trajectory x is affected by noise in Remark 9 in Section V.

IV. Data-driven analysis for exact data

Let Σ be a set of continuous-time systems containing a data-generating *unknown* system $\mathcal{S}$, and denote by $\mathfrak{D}$ a set of trajectories generated by $\mathcal{S}$. Denote by $\Sigma_{\mathfrak{D}} \subseteq \Sigma$ the set of all systems that could have generated the data; note that $\mathcal{S} \in \Sigma_{\mathfrak{D}}$. Denote by $\Sigma_{\mathcal{P}} \subseteq \Sigma$ the set consisting of all systems sharing a given property $\mathcal{P}$ (e.g. all stable, or controllable, or stabilizable systems in Σ).

Definition 1 (Informativity for a property $\mathcal{P}$): The data $\mathfrak{D}$ is *informative for* $\mathcal{P}$ if $\Sigma_{\mathfrak{D}} \subseteq \Sigma_{\mathcal{P}}$.
The *informativity problem* consists in providing necessary and sufficient conditions on $\mathfrak{D}$ under which the data are informative for property $\mathcal{P}$. In this section we solve it for various instances of the property $\mathcal{P}$, assuming that the *derivative approximation error* is *negligible*. We call this the *exact data* case.

In the rest of this paper, the data $\mathfrak{D}$ consists of the coefficients of the state and input-trajectory defined in (23):

$$\mathfrak{D} := (X, U) \ .$$

We call an input-state model associated with the matrices $(\widehat{A}, \widehat{B})$ *unfalsified* by $\mathfrak{D}$ if $X\mathcal{D}_{11} = \widehat{A}X + \widehat{B}U$. We denote by $\Sigma_{\mathfrak{D}}$ the set consisting of all pairs $(\widehat{A}, \widehat{B})$ *unfalsified* by $\mathfrak{D}$:

$$\Sigma_{\mathfrak{D}} := \left\{ (\widehat{A}, \widehat{B}) \mid X\mathcal{D}_{11} = \begin{bmatrix} \widehat{A} & \widehat{B} \end{bmatrix} \begin{bmatrix} X \\ U \end{bmatrix} \right\} . \tag{24}$$

A. Informativity for system identification

The data (X, U) are *informative for system identification* if $\Sigma_{\mathfrak{D}}$ contains only one element. Recall that an approximation error for the state derivative always occurs when working with truncated series. Consequently, even if the data are informative for identification, whether the unique element of $\Sigma_{\mathfrak{D}}$ is a good approximation of the data-generating system or not depends on how small the truncation error is (see also Example 7).

The following result is the differential version of Proposition 6 in [27]; we omit the proof since it is completely analogous.

Proposition 3: The data (23) are informative for system identification if and only if

$$\operatorname{rank} \begin{bmatrix} X \\ U \end{bmatrix} = n + m \ . \tag{25}$$

Moreover, if (25) holds, for any right inverse $\begin{bmatrix} X \\ U \end{bmatrix}^{\dagger}$ of $\begin{bmatrix} X \\ U \end{bmatrix}$ it holds that

$$\begin{bmatrix} A & B \end{bmatrix} = X\mathcal{D}_{11} \begin{bmatrix} X \\ U \end{bmatrix}^{\dagger} .$$

Example 5: Consider the unstable system

$$\frac{d}{dt}x = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} x + \begin{bmatrix} 1 \\ 0 \end{bmatrix} u \ ,$$

let $x(0) = \begin{bmatrix} 1 & 0 \end{bmatrix}^{\top}$ and $u(t) = e^{-2t}$; then $x_1(t) = -\frac{2}{3}e^{-2t} + e^{-t} + \frac{2e^t}{3}$ and $x_2(t) = \frac{e^{-2t}}{3} - e^{-t} + \frac{2e^t}{3}$. Since u is infinitely differentiable, also x is, and machine-precision accuracy of the Chebyshev representation of x, u and $\frac{d}{dt}x$ can be achieved with a relatively small amount of data. Indeed, using the `Chebfun` toolbox (see [34]), it can be verified that $N = 19$ samples are enough to compute a machine-precision Chebyshev representation of the signals, see Table I[4]. We can estimate *a posteriori* the accuracy of the derivative approximation by comparing the values of the coefficients of the derivative of the projections of x_1 and x_2 with those of the analytical derivatives $\frac{d}{dt}x_1(t) = \frac{4}{3}e^{-2t} - e^{-t} + \frac{2}{3}e^t$ and $\frac{d}{dt}x_2(t) = -\frac{2}{3}e^{-2t} + e^{-t} + \frac{2}{3}e^t$. The weighted 2-norms of the differences are $1.2157 \cdot 10^{-14}$ and $7.2397 \cdot 10^{-15}$, respectively.

TABLE I
Chebyshev coefficients for Example 5

$\widetilde{u}$	$\widetilde{x}_1$	$\widetilde{x}_2$
2.2796	$5.9039 \cdot 10^{-1}$	$3.3784 \cdot 10^{-1}$
-3.1813	1.7441	$8.2344 \cdot 10^{-1}$
1.3779	$-4.6611 \cdot 10^{-1}$	$3.6880 \cdot 10^{-1}$
$-4.2548 \cdot 10^{-1}$	$2.6887 \cdot 10^{-1}$	$-6.7932 \cdot 10^{-2}$
$1.0146 \cdot 10^{-1}$	$-5.8514 \cdot 10^{-2}$	$3.1994 \cdot 10^{-2}$
$-1.9651 \cdot 10^{-2}$	$1.2920 \cdot 10^{-2}$	$-5.6456 \cdot 10^{-3}$
$3.2003 \cdot 10^{-3}$	$-2.0586 \cdot 10^{-3}$	$1.0518 \cdot 10^{-3}$
$-4.4928 \cdot 10^{-4}$	$2.9845 \cdot 10^{-4}$	$-1.4443 \cdot 10^{-4}$
$5.5399 \cdot 10^{-5}$	$-3.6600 \cdot 10^{-5}$	$1.8400 \cdot 10^{-5}$
$-6.0884 \cdot 10^{-6}$	$4.0552 \cdot 10^{-6}$	$-2.0111 \cdot 10^{-6}$
$6.0339 \cdot 10^{-7}$	$-4.0134 \cdot 10^{-7}$	$2.0095 \cdot 10^{-7}$
$-5.4444 \cdot 10^{-8}$	$3.6288 \cdot 10^{-8}$	$-1.8106 \cdot 10^{-8}$
$4.5083 \cdot 10^{-9}$	$-3.0038 \cdot 10^{-9}$	$1.5024 \cdot 10^{-9}$
$-3.4490 \cdot 10^{-10}$	$2.2992 \cdot 10^{-10}$	$-1.1490 \cdot 10^{-10}$
$2.4520 \cdot 10^{-11}$	$-1.6344 \cdot 10^{-11}$	$8.1727 \cdot 10^{-12}$
$-1.6278 \cdot 10^{-12}$	$1.0852 \cdot 10^{-12}$	$-5.4261 \cdot 10^{-13}$
$1.0125 \cdot 10^{-13}$	$-6.7502 \cdot 10^{-14}$	$3.3751 \cdot 10^{-14}$
$-5.7732 \cdot 10^{-15}$	$3.8488 \cdot 10^{-15}$	$-1.9244 \cdot 10^{-15}$
$4.0246 \cdot 10^{-16}$	$-2.6830 \cdot 10^{-16}$	$1.3415 \cdot 10^{-16}$

[3]Assuming this situation is from the conceptual point of view as strong as assumption 2 and condition $ii)$ in Lemma 4 of [18], namely that the data are *exactly* representable by a *truncated* polynomial chaos expansion.

[4]N is chosen automatically by `Chebfun`, see Footnote 2. The number of samples equals the number of coefficients, see Remark 1.

The singular values of $\begin{bmatrix} X \\ U \end{bmatrix}$ are 4.3786, 1.6092 and $5.1243 \cdot 10^{-1}$, consequently condition (25) is satisfied and the data are informative for system identification. Using the formula for $\widehat{A}$ and $\widehat{B}$ in the second part of Prop. 3, one obtains

$$\begin{aligned} \widehat{A} &= \begin{bmatrix} -3.8195 \cdot 10^{-16} & 1 \\ 1 & -4.5317 \cdot 10^{-15} \end{bmatrix} \\ \widehat{B} &= \begin{bmatrix} 1 \\ -2.2438 \cdot 10^{-16} \end{bmatrix} , \end{aligned}$$

that coincide up to machine precision with the matrices (A, B) of the data-generating system. □

B. Informativity for controllability

Denote by Σ_{cont} the subset of $\Sigma_{\mathfrak{D}}$ consisting of all controllable pairs (A', B'). We call the data $\mathfrak{D}$ *informative for controllability* if every unfalsified model is controllable, equivalently $\Sigma_{\mathfrak{D}} \subseteq \Sigma_{\text{cont}}$. To characterize informativity for controllability, we first establish a characterization of controllability alternative to the classical Popov-Belevitch-Hautus (PBH) test.

Denote by $\mathfrak{B}_x$ the projection of the solution space of (20) onto the component x:

$$\mathfrak{B}_x := \{x : \mathbb{I} \to \mathbb{R}^n \mid \exists\ u : \mathbb{I} \to \mathbb{R}^m \text{ s.t. (20) holds}\} \ .$$

Proposition 4: The following statements are equivalent:

1) (A', B') is controllable;
2) For all $\lambda \in \mathbb{C}$, rank $\begin{bmatrix} A' - \lambda I & B' \end{bmatrix} = n$;
3) If $v \in \mathbb{C}^n$ and $\lambda \in \mathbb{C}$ are such that $v^* \left(\frac{d}{dt}x - \lambda x\right) = 0$ for all $x \in \mathfrak{B}_x$, then $v = 0$.

Proof: The equivalence of 1) and 2) is well known.

We prove that 3) $\Rightarrow$ 2). Let $v \in \mathbb{C}^n$ and $\lambda \in \mathbb{C}$ be such that $v^* \begin{bmatrix} A' - \lambda I & B' \end{bmatrix} = 0$. Left-multiply both sides of $\frac{d}{dt}x = A'x + B'u$ by v^*; conclude that $v^* \frac{d}{dt}x = v^* A'x + v^* B'u = v^* A'x = v^* \lambda x$, and consequently that $v^* \left(\frac{d}{dt}x - \lambda x\right) = 0$ for all $x \in \mathfrak{B}_x$. This implies $v = 0$, and consequently rank $\begin{bmatrix} A' - \lambda I & B' \end{bmatrix} = n$ for all $\lambda \in \mathbb{C}$.

To prove 2) $\Rightarrow$ 3), let $v \in \mathbb{C}^n$ and $\lambda \in \mathbb{C}$ be such that $v^* \left(\frac{d}{dt}x - \lambda x\right) = 0$ for all $x \in \mathfrak{B}_x$. It follows that $v^* \frac{d}{dt}x = v^* A'x + v^* B'u = v^* \lambda x$. This implies that

$$v^* \begin{bmatrix} A' - \lambda I & B' \end{bmatrix} \begin{bmatrix} x \\ u \end{bmatrix} = 0 ,$$

for every (x, u) satisfying $\frac{d}{dt}x = A'x + B'u$. We prove that for every $\begin{bmatrix} \overline{x}^\top & \overline{u}^\top \end{bmatrix}^\top \in \mathbb{R}^{n+m}$ there exists a trajectory of (20) such that $x(0) = \overline{x}$ and $u(0) = \overline{u}$. Let u' be any function such that $u'(0) = \overline{u}$. The differential equation $\frac{d}{dt}x' = A'x' + B'u'$, with initial condition $x'(0) = \overline{x}$ has a unique solution x'. Consequently (u', x') satisfies the differential equation, and $(u'(0), x'(0)) = (\overline{u}, \overline{x})$.

It follows that if $v^* \begin{bmatrix} A' - \lambda I & B' \end{bmatrix} \begin{bmatrix} x \\ u \end{bmatrix} = 0$ for every (x, u) satisfying (20), then $v^* \begin{bmatrix} A' - \lambda I & B' \end{bmatrix} \begin{bmatrix} \overline{x} \\ \overline{u} \end{bmatrix} = 0$ for every $\begin{bmatrix} \overline{x}^\top & \overline{u}^\top \end{bmatrix}^\top \in \mathbb{R}^{n+m}$. Consequently $v^* \begin{bmatrix} A' - \lambda I & B' \end{bmatrix} = 0$, and necessarily $v = 0$. ■

Theorem 3: The data (23) are informative for controllability if and only if

$$\text{rank}\,(X\mathcal{D}_{11} - \lambda X) = n \quad \text{for every} \quad \lambda \in \mathbb{C}\ . \qquad (26)$$

Proof: We first prove sufficiency. Let $\left(\widehat{A}, \widehat{B}\right) \in \Sigma_{\mathfrak{D}}$. Assume that there exists $v \in \mathbb{C}^n$ and $\lambda \in \mathbb{C}$ such that $v^* \begin{bmatrix} \widehat{A} - \lambda I & \widehat{B} \end{bmatrix} = 0$. Multiply both sides of this equation on the right by $\begin{bmatrix} X \\ U \end{bmatrix}$ to conclude that $v^* \left(\widehat{A}X + \widehat{B}U\right) = v^* X\mathcal{D}_{11} = v^* \lambda X$. Since rank $(X\mathcal{D}_{11} - \lambda X) = n$, it follows that $v = 0$. Consequently, rank $\begin{bmatrix} \widehat{A} - \lambda I & \widehat{B} \end{bmatrix} = n$ for every $\lambda \in \mathbb{C}$, and $\left(\widehat{A}, \widehat{B}\right)$ is controllable.

To prove necessity, let $v \in \mathbb{C}^n$ and $\lambda \in \mathbb{C}$ be such that $v^* (X\mathcal{D}_{11} - \lambda X) = 0$. Let $\left(\widehat{A}, \widehat{B}\right) \in \Sigma_{\mathfrak{D}}$; it follows that

$$v^* X\mathcal{D}_{11} = v^* \widehat{A}X + v^* \widehat{B}U = v^* \begin{bmatrix} \widehat{A} & \widehat{B} \end{bmatrix} \begin{bmatrix} X \\ U \end{bmatrix} = \lambda v^* X ,$$

equivalently $v^* \begin{bmatrix} \widehat{A} - \lambda I & \widehat{B} \end{bmatrix} \begin{bmatrix} X \\ U \end{bmatrix} = 0$, and consequently also $vv^* \begin{bmatrix} \widehat{A} - \lambda I & \widehat{B} \end{bmatrix} \begin{bmatrix} X \\ U \end{bmatrix} = 0$. We show that the last equation implies the existence of an unfalsified, *uncontrollable* model for (X, U). Assume first that $\lambda \in \mathbb{R}$. Then $v \in \mathbb{R}^n$; without loss of generality we can assume that $\|v\|^2 = 1$. Define

$$A' := \widehat{A} - vv^\top \left(\widehat{A} - \lambda I\right) \text{ and } B' := \widehat{B} - vv^\top \widehat{B}\ ; \qquad (27)$$

it is straightforward to check that $(A', B') \in \Sigma_{\mathfrak{D}}$. It is also straightforward to check that $v^\top \begin{bmatrix} A' - \lambda I & B' \end{bmatrix} = 0$, contradicting the fact that all unfalsified models for (X, U) are controllable, since (X, U) is informative for controllability. The proof for the case $\lambda \in \mathbb{C}$, $\text{Im}(\lambda) \neq 0$ is analogous to that of Theorem 8 in [27], and is omitted. ■

Remark 8: Informativity for stabilizability is defined in Section III of [27]. Its characterization can be obtained following arguments completely analogous to those used to prove Theorem 3. We do not enter into such details. □

C. Informativity for state feedback stabilization

The data (X, U) in (23) are *informative for state feedback stabilization* if there exists $K \in \mathbb{R}^{m \times n}$ such that

$$(A', B') \in \Sigma_{\mathfrak{D}} \quad \Longrightarrow \quad A' + B'K \ \text{ is Hurwitz}\ .$$

We call such a K a *stabilizing feedback* for $\Sigma_{\mathfrak{D}}$.

The following result is instrumental in establishing our characterization of informativity for state feedback stabilization.

Lemma 1: Assume that (X, U) are informative for state feedback stabilization, and let $K \in \mathbb{R}^{m \times n}$ be a stabilizing feedback for $\Sigma_{\mathfrak{D}}$. Then $\text{im} \begin{bmatrix} I_n \\ K \end{bmatrix} \subseteq \text{im} \begin{bmatrix} X \\ U \end{bmatrix}$.

Proof: Let $(A', B') \in \Sigma_{\mathfrak{D}}$. Let (A_0, B_0) be such that $\begin{bmatrix} A_0 & B_0 \end{bmatrix} \begin{bmatrix} X \\ U \end{bmatrix} = 0$; then $(A' + \alpha A_0, B' + \alpha B_0) \in \Sigma_{\mathfrak{D}}$ for all $\alpha \in \mathbb{R}$. Since K is a stabilizing feedback for $\Sigma_{\mathfrak{D}}$, it follows that for all $\alpha \in \mathbb{R}$

$$A' + \alpha A_0 + (B' + \alpha B_0)\, K = \underbrace{A' + B'K}_{=:F} + \alpha \underbrace{(A_0 + B_0 K)}_{=:F_0} ,$$

is Hurwitz. Now define $F_\alpha := \frac{1}{\alpha}(F + \alpha F_0)$. For larger positive values of α, F_α is Hurwitz, and consequently all eigenvalues of F_0 have nonpositive real part. For larger negative values of α, F_α is also Hurwitz, and consequently all eigenvalues of F_0 have nonnegative real part. It follows that all eigenvalues of F_0 are purely imaginary. Since $\begin{bmatrix} A_0 & B_0 \end{bmatrix} \begin{bmatrix} X \\ U \end{bmatrix} = \begin{bmatrix} F_0^\top A_0 & F_0^\top B_0 \end{bmatrix} \begin{bmatrix} X \\ U \end{bmatrix} = 0$, we conclude that also all eigenvalues of $F_0^\top F_0$ are purely imaginary. Since $F_0^\top F_0$ is symmetric, its eigenvalues are all real; we conclude that zero is the only eigenvalue of $F_0^\top F_0$. The only symmetric matrix with such property is the zero matrix. Conclude that $F_0 = 0$, and consequently, that $A_0 + B_0 K = 0$ for all (A_0, B_0) such that $\begin{bmatrix} A_0 & B_0 \end{bmatrix} \begin{bmatrix} X \\ U \end{bmatrix} = 0$. It follows that the orthogonal subspace to $\operatorname{im} \begin{bmatrix} X \\ U \end{bmatrix}$ is contained in the orthogonal subspace of $\operatorname{im} \begin{bmatrix} I_n \\ K \end{bmatrix}$, and consequently that $\operatorname{im} \begin{bmatrix} I_n \\ K \end{bmatrix} \subseteq \operatorname{im} \begin{bmatrix} X \\ U \end{bmatrix}$. ■

Theorem 4: (X, U) are informative for state feedback stabilization if and only if there exists $\Theta \in \mathbb{R}^{N \times n}$ such that the following linear matrix inequalities hold:

$$\begin{aligned} X\Theta = (X\Theta)^\top &\succ 0 \\ \Theta^\top \mathcal{D}_{11}^\top X^\top + X \mathcal{D}_{11} \Theta &\prec 0 \,. \end{aligned} \tag{28}$$

Moreover, $K \in \mathbb{R}^{m \times n}$ is a stabilizing feedback for $\Sigma_{\mathfrak{D}}$ if and only if $K = U\Theta (X\Theta)^{-1}$, for some $\Theta \in \mathbb{R}^{N \times n}$ satisfying (28).

Proof: We first prove sufficiency. The first LMI in (28) implies that X has full row rank; it is straightforward to check that $\Theta (X\Theta)^{-1}$ is a right-inverse of X. Define $K := U\Theta (X\Theta)^{-1}$. Let $(A', B') \in \Sigma_{\mathfrak{D}}$; multiply $X\mathcal{D}_{11} = A'X + B'U$ on the right by $\Theta (X\Theta)^{-1}$, obtaining $X\mathcal{D}_{11}\Theta (X\Theta)^{-1} = A' + B'K$. We show that $A' + B'K$ is Hurwitz. Multiply the second LMI in (28) on the left by $(X\Theta)^{-\top}$ and on the right by $(X\Theta)^{-1}$, and conclude that the matrix

$$\underbrace{(X\Theta)^{-\top} \Theta^\top \mathcal{D}_{11}^\top X^\top}_{=(A'+B'K)^\top} (X\Theta)^{-1} + (X\Theta)^{-\top} \underbrace{X \mathcal{D}_{11} \Theta (X\Theta)^{-1}}_{=(A'+B'K)}$$

is negative-definite. Define $P := (X\Theta)^{-1}$; the first LMI in (28) implies that P induces a Lyapunov function for $A' + B'K$. It follows that K stabilizes every $(A', B') \in \Sigma_{\mathfrak{D}}$. This argument also proves the sufficiency part of the second statement of the Theorem.

We prove the necessity of the first statement of Theorem 4. Let K be a stabilizing gain for $\Sigma_{\mathfrak{D}}$; from the inclusion proved in Lemma 1 it follows that X has full row rank, and consequently there exists a right-inverse $X^\dagger$ of X such that $K = UX^\dagger$. Moreover, for every $(A', B') \in \Sigma_{\mathfrak{D}}$ it holds that $A' + B'K = X\mathcal{D}_{11}X^\dagger$ is Hurwitz. It follows that $(A' + B'K)^\top$ is also Hurwitz, and consequently there exists $P = P^\top$, $P \succ 0$, such that $\left(X\mathcal{D}_{\text{fin}}X^\dagger\right) P + P \left(X\mathcal{D}_{\text{fin}}X^\dagger\right)^\top \prec 0$. Define $\Theta := X^\dagger P$; then the second inequality in (28) holds. Since $P = X\Theta \succ 0$ also the first inequality in (28) holds.

To prove the necessity of the second statement of Theorem 4, observe that $K = UX^\dagger = U\left(\Theta P^{-1}\right) = U\Theta (X\Theta)^{-1}$. ■

Example 6: We use the same setting and data of Ex. 5. Using `Yalmip` (see [35]) with `sedumi` solver, we obtain the following solution of the LMIs (28)

$$\Theta = \begin{bmatrix} 5.2887 \cdot 10^{-1} & -8.7668 \cdot 10^{-1} \\ 1.6588 \cdot 10^{-1} & 7.7412 \cdot 10^{-1} \\ -8.5485 \cdot 10^{-1} & 1.7862 \\ -5.0064 \cdot 10^{-17} & 1.7838 \cdot 10^{-7} \\ 1.9233 \cdot 10^{-15} & -5.4341 \cdot 10^{-8} \\ 2.1337 \cdot 10^{-16} & 1.0751 \cdot 10^{-8} \\ 1.5428 \cdot 10^{-15} & -1.8505 \cdot 10^{-9} \\ 1.0612 \cdot 10^{-15} & 2.6120 \cdot 10^{-10} \\ -4.6407 \cdot 10^{-16} & -3.2656 \cdot 10^{-11} \\ 1.3602 \cdot 10^{-15} & 3.6048 \cdot 10^{-12} \\ -2.0531 \cdot 10^{-15} & -3.7456 \cdot 10^{-13} \\ 1.3785 \cdot 10^{-15} & 4.3800 \cdot 10^{-14} \\ -5.1669 \cdot 10^{-16} & -6.9998 \cdot 10^{-15} \\ 6.3417 \cdot 10^{-16} & 5.5146 \cdot 10^{-15} \\ -1.6384 \cdot 10^{-16} & -1.3864 \cdot 10^{-15} \\ 2.3831 \cdot 10^{-16} & 1.9964 \cdot 10^{-15} \\ -1.3345 \cdot 10^{-15} & -1.1174 \cdot 10^{-14} \\ -6.3151 \cdot 10^{-16} & -5.2879 \cdot 10^{-15} \\ -1.7277 \cdot 10^{-16} & -1.4467 \cdot 10^{-15} \end{bmatrix} .$$

We conclude that the data is informative for stabilization. Using the second part of Theorem 4, we compute

$$K = U\Theta (X\Theta)^{-1} = \begin{bmatrix} -0.5 & -2 \end{bmatrix} .$$

Such K stabilizes the pair $(\widehat{A}, \widehat{B})$ identified from the data in Example 5: the eigenvalues of $\widehat{A} + \widehat{B}K$ are $-0.25 \pm 0.9683j$. It can be verified that K also stabilizes the data-generating pair (A, B), with the same closed-loop eigenvalues. □

V. Quadratic stabilization via the matrix S-lemma

In Section II-A we argued that given enough computational power and samples, in the noiseless case the Chebyshev representation of the signals x and u can be computed to high accuracy. A possible source of inaccuracy is the approximation of the Chebyshev coefficients of $\frac{d}{dt}x$ by those of the derivative of the projection of x (see Section II-D). In this section we show how to solve the quadratic stabilization problem in this case, using the matrix S-lemma (see [6], [33]) and the error analysis developed in Section II-D.

We briefly summarize the notation pertinent to, and the statement of, the matrix S-lemma. Let $\mathcal{M}, \mathcal{N} \in \mathbb{R}^{(q+r)\times(q+r)}$ be two symmetric matrices partitioned as

$$\mathcal{M} = \begin{bmatrix} \mathcal{M}_{11} & \mathcal{M}_{12} \\ \mathcal{M}_{12}^\top & \mathcal{M}_{22} \end{bmatrix} , \ \mathcal{N} = \begin{bmatrix} \mathcal{N}_{11} & \mathcal{N}_{12} \\ \mathcal{N}_{12}^\top & \mathcal{N}_{22} \end{bmatrix} , \tag{29}$$

where the $(1,1)$-blocks are $q \times q$ and the $(2,2)$-blocks are $r \times r$. Define $\mathcal{Z}_r(\mathcal{N})$ and $\mathcal{Z}_r^+(\mathcal{M})$ respectively by

$$\begin{aligned} \mathcal{Z}_r(\mathcal{N}) &:= \left\{ Z \in \mathbb{R}^{r\times q} \mid \begin{bmatrix} I_q \\ Z \end{bmatrix}^\top \mathcal{N} \begin{bmatrix} I_q \\ Z \end{bmatrix} \succeq 0 \right\} \\ \mathcal{Z}_r^+(\mathcal{M}) &:= \left\{ Z \in \mathbb{R}^{r\times q} \mid \begin{bmatrix} I_q \\ Z \end{bmatrix}^\top \mathcal{M} \begin{bmatrix} I_q \\ Z \end{bmatrix} \succ 0 \right\} . \end{aligned}$$

The following result is Corollary 4.13 p. 15 of [33].

Proposition 5: Let $\mathcal{M}$, $\mathcal{N}$ be symmetric matrices partitioned as in (29). Assume that $\mathcal{M}_{22} \preceq 0$, $\mathcal{N}_{22} \preceq 0$, that

$\mathcal{N}_{11}-\mathcal{N}_{12}\mathcal{N}_{22}^{\dagger}\mathcal{N}_{12}^{\top}\succeq 0$, and that $\ker\mathcal{N}_{22}\subseteq\ker\mathcal{N}_{12}$. The following statements are equivalent:

1) $\mathcal{Z}_r(\mathcal{N})\subseteq\mathcal{Z}_r^+(\mathcal{M})$;
2) There exist $\alpha\geq 0$ and $\beta>0$ such that

$$\mathcal{M}-\alpha\mathcal{N}\succeq\begin{bmatrix}\beta I & 0\\ 0 & 0\end{bmatrix}.$$

Using an analogous partition to (18), and recalling the definition of Π_N in (8), we write

$$\begin{aligned}\widetilde{x} &= \begin{bmatrix}\Pi_N(x) & \widetilde{x}'\end{bmatrix}=\begin{bmatrix}X & \widetilde{x}'\end{bmatrix}\\ \widetilde{u} &= \begin{bmatrix}\Pi_N(u) & \widetilde{u}'\end{bmatrix}=\begin{bmatrix}U & \widetilde{u}'\end{bmatrix}.\end{aligned}\tag{30}$$

Using the partition (19) of the differentiation matrix, write

$$\Pi_N(x)\mathcal{D}_{11}=A\Pi_N(x)+B\Pi_N(u)-\widetilde{x}'\mathcal{D}_{21};$$

define

$$W_-:=-\widetilde{x}'\mathcal{D}_{21}\in\mathbb{R}^{n\times(N+1)};\tag{31}$$

and write, analogously to equation (2) p. 163 of [6]:

$$\begin{aligned}W_- &= \Pi_N(x)\mathcal{D}_{11}-A\Pi_N(x)-B\Pi_N(u)\\ &= \begin{bmatrix}I & A & B\end{bmatrix}\begin{bmatrix}X\mathcal{D}_{11}\\ -X\\ -U\end{bmatrix}.\end{aligned}\tag{32}$$

Recall from Proposition 2 that there exists a constant c, determined by the total variation of the components of the second derivative of x and N, such that the 2-norm of $-\widetilde{x}'\mathcal{D}_{21}$ is less than or equal to c, equivalently

$$\begin{bmatrix}I & W_-\end{bmatrix}\begin{bmatrix}cI & 0\\ 0 & -I\end{bmatrix}\begin{bmatrix}I\\ W_-^{\top}\end{bmatrix}\succeq 0.\tag{33}$$

Redefine the set of explanatory models for the data (X,U) (see (24)) by

$$\Sigma_{\mathfrak{D},c}:=\Big\{(\widehat{A},\widehat{B})\mid X\mathcal{D}_{11}=\begin{bmatrix}\widehat{A} & \widehat{B}\end{bmatrix}\begin{bmatrix}X\\ U\end{bmatrix}+W_- \tag{34}$$
$$\text{for some } W_-\in\mathbb{R}^{n\times(N+1)} \text{ satisfying (33)}\Big\}.$$

Definition 2: (X,U) are *informative for quadratic stabilization* if there exist $K\in\mathbb{R}^{m\times n}$ and $P\in\mathbb{R}^{n\times n}$, $P=P^{\top}\succ 0$ such that

$$\left(\widehat{A}+\widehat{B}K\right)P+P\left(\widehat{A}+\widehat{B}K\right)^{\top}\succ 0,$$

for all $(\widehat{A},\widehat{B})\in\Sigma_{\mathfrak{D},c}$[5].

It is straightforward to check that Definition 2 is equivalent to the existence of a P and K such that:

$$\begin{aligned}P=P^{\top} &\succ 0\\ (A+BK)P+P(A+BK)^{\top} &\prec 0.\end{aligned}\tag{35}$$

The second inequality in (35) can be written as

$$\begin{bmatrix}I & A & B\end{bmatrix}\begin{bmatrix}0 & -P & -PK^{\top}\\ -P & 0 & 0\\ -KP & 0 & 0\end{bmatrix}\begin{bmatrix}I\\ A^{\top}\\ B^{\top}\end{bmatrix}\succ 0.$$

[5]As in [6], we use the "transpose" of the standard Lyapunov equation.

Define

$$\begin{aligned}\mathcal{N} &:= \begin{bmatrix}I & X\mathcal{D}_{11}\\ 0 & -X\\ 0 & -U\end{bmatrix}\begin{bmatrix}cI & 0\\ 0 & -I_{N+1}\end{bmatrix}\begin{bmatrix}I & X\mathcal{D}_{11}\\ 0 & -X\\ 0 & -U\end{bmatrix}^{\top}\\ &=: \begin{bmatrix}\mathcal{N}_{11} & \mathcal{N}_{12}\\ \mathcal{N}_{12}^{\top} & \mathcal{N}_{22}\end{bmatrix}.\end{aligned}\tag{36}$$

Use (32) to rewrite (33) as

$$\begin{bmatrix}I & A & B\end{bmatrix}\mathcal{N}\begin{bmatrix}I\\ A^{\top}\\ B^{\top}\end{bmatrix}\succeq 0.$$

Note that

$$\mathcal{N}_{22}=-\begin{bmatrix}X\\ U\end{bmatrix}\begin{bmatrix}X^{\top} & U^{\top}\end{bmatrix}\preceq 0;\tag{37}$$

moreover, $\mathcal{N}_{12}=X\mathcal{D}_{11}\begin{bmatrix}X^{\top} & U^{\top}\end{bmatrix}$, and consequently,

$$\ker\mathcal{N}_{12}\supseteq\ker\mathcal{N}_{22}.\tag{38}$$

Finally, observe that

$$\mathcal{N}_{11}-\mathcal{N}_{12}\mathcal{N}_{22}^{\dagger}\mathcal{N}_{12}^{\top}\succeq 0,\tag{39}$$

since $\Sigma_{\mathfrak{D},c}\neq\emptyset$ by Definition 2. Now define

$$\mathcal{M}:=\left[\begin{array}{c:cc}0 & -P & -PK^{\top}\\ \hdashline -P & 0 & 0\\ -KP & 0 & 0\end{array}\right]=:\begin{bmatrix}\mathcal{M}_{11} & \mathcal{M}_{12}\\ \mathcal{M}_{12}^{\top} & \mathcal{M}_{22}\end{bmatrix},$$

and observe that

$$\mathcal{M}_{22}\preceq 0.\tag{40}$$

Since the inequalities (37)-(40) hold, all the assumptions of Proposition 5 are satisfied. We obtain the following characterization of informativity for quadratic stabilization, and an LMI-based design procedure.

Theorem 5: Define $\mathcal{N}$ by (36). The following statements are equivalent:

1) (X,U) are informative for quadratic stabilization;
2) There exists $P\in\mathbb{R}^{n\times n}$, $L\in\mathbb{R}^{m\times n}$, $\alpha,\beta\in\mathbb{R}$ such that $P=P^{\top}\succ 0$, $\alpha\geq 0$, $\beta>0$ and the LMI

$$\begin{bmatrix}0 & -P & -L^{\top}\\ -P & 0 & 0\\ -L & 0 & 0\end{bmatrix}-\alpha\mathcal{N}\succeq\begin{bmatrix}\beta I & 0 & 0\\ 0 & 0 & 0\\ 0 & 0 & 0\end{bmatrix},\tag{41}$$

is satisfied.

Moreover, if P and L satisfy the LMI (41), then $K:=LP^{-1}$ is a stabilizing feedback gain for all $(A,B)\in\Sigma_{\mathfrak{D},c}$.

Example 7 (Example 6 revisited): We use the same data-generating system and continuous-time trajectories as in Examples 5 and 6. A non-negligible error in the derivative approximation arises when considering only $N=4$ coefficients.

To estimate the error bound of Proposition 2, we proceed as follows. We compute the Chebyshev coefficients of $\frac{d^2}{dt^2}\Pi_N(x_i)$ (the second derivative of the truncation of x_i) by matrix multiplication, as $\Pi_N(x_i)\mathcal{D}_{11}^2$, $i=1,2$. We compute the total variation of these two functions by integration over $(-1,1)$ of the absolute values of the corresponding polynomials; the resulting values are 6.99166 and 2.96517, respectively. The corresponding error bounds of Proposition 2 are 0.556379 and 0.235961, respectively.

To define $\mathcal{N}$ in (36), we first choose[6] $c_{\max} = 0.556379$. We use `Yalmip` (with `sedumi` solver) to set up the LMIs in statement 2) of Theorem 5. The system of LMIs is feasible, and the solutions are

$$\begin{aligned} P &= \begin{bmatrix} 9.4552 \cdot 10^{-1} & -1.9180 \cdot 10^{-1} \\ -1.9180 \cdot 10^{-1} & 4.2712 \cdot 10^{-2} \end{bmatrix} \\ L &= \begin{bmatrix} -4.7401 & 3.9634 \cdot 10^{-1} \end{bmatrix} . \end{aligned}$$

The state feedback gain corresponding to P and L is $K = LP^{-1} = \begin{bmatrix} -35.143 & -148.53 \end{bmatrix}$. The data-generating system (A, B) defined in Example 5 belongs to $\Sigma_{\mathfrak{D},c}$ defined by (34). The corresponding closed-loop eigenvalues are

$$\sigma(A + BK) = \{-30.269, -4.8740\} .$$

It can be verified that the data are such that rank $\begin{bmatrix} X \\ U \end{bmatrix} = 3$, and consequently are informative for identification, with explanatory system described by

$$\begin{aligned} A' &:= \begin{bmatrix} 2.5989 \cdot 10^{-1} & 8.3569 \cdot 10^{-1} \\ 8.6458 \cdot 10^{-1} & 6.9179 \cdot 10^{-2} \end{bmatrix} \\ B' &:= \begin{bmatrix} 9.4008 \cdot 10^{-1} \\ 6.9179 \cdot 10^{-2} \end{bmatrix} . \end{aligned}$$

$(A', B') \in \Sigma_{\mathfrak{D},c_{\max}}$, and it can be verified that

$$\sigma(A' + B'K) = \{-40.060, -2.9229\} .$$

The model described by

$$\begin{aligned} A'' &:= \begin{bmatrix} -2.3532 \cdot 10^{-1} & 1.8178 \\ 8.6458 \cdot 10^{-1} & 6.9179 \cdot 10^{-2} \end{bmatrix} \\ B'' &:= \begin{bmatrix} 9.2279 \cdot 10^{-1} \\ 6.9179 \cdot 10^{-2} \end{bmatrix} , \end{aligned}$$

also belongs to $\Sigma_{\mathfrak{D},c_{\max}}$, and it can be verified that

$$\sigma(A'' + B''K) = \{-39.819, -3.0515\} .$$

Choosing $c_{\min}$ and solving the matrix inequalities (41), we obtain

$$\begin{aligned} P' &= \begin{bmatrix} 7.1395 \cdot 10^{-1} & -1.4333 \cdot 10^{-1} \\ -1.4333 \cdot 10^{-1} & 3.9683 \cdot 10^{-2} \end{bmatrix} \\ L' &= \begin{bmatrix} -3.8543 & 2.8644 \cdot 10^{-1} \end{bmatrix} . \end{aligned}$$

The state feedback gain corresponding to P' and L' is $K' = L'P'^{-1} = \begin{bmatrix} -14.370 & -44.685 \end{bmatrix}$. The model (A'', B'') does not belong to $\Sigma_{\mathfrak{D},c_{\min}}$, but (A, B) and (A', B') do; moreover,

$$\begin{aligned} \sigma(A + BK') &= \{-10.002, -4.3676\} \\ \sigma(A' + B'K) &= \{-13.746, -2.5249\} . \end{aligned}$$

[6]We can verify *a posteriori* that this choice of values for c is correct. We first compute with `Chebfun` up to machine precision the coefficients of the Chebyshev representation of $\frac{d}{dt}x_i$, $i = 1, 2$. As in Example 5, 19 coefficients are enough. An estimate of the i-th row of $W_- \in \mathbb{R}^{2\times 10}$, $i = 1, 2$, is given by the coefficients of $x_i^{(1)}$ from the 4th to the 19th. With this definition of W_- we compute

$$W_- W_-^\top = \begin{bmatrix} 1.8468 \cdot 10^{-2} & -8.1446 \cdot 10^{-3} \\ -8.1446 \cdot 10^{-3} & 3.5952 \cdot 10^{-3} \end{bmatrix} .$$

This matrix has eigenvalues $2.7297 \cdot 10^{-6}$ and $2.2060 \cdot 10^{-2}$, and consequently the inequality (33) holds for $c_{\max} = 0.556379$, and also for $c_{\min} = 0.235961$.

Remark 9 (The noisy case): Assume that the trajectory x is affected by disturbances, i.e. that $\widehat{x} = x + \epsilon$ is measured instead of x, where ϵ is an unknown disturbance. It follows from the definition of the POB coefficients that

$$\widetilde{\widehat{x}} = \widetilde{x} + \widetilde{\epsilon} ,$$

and consequently that the coefficients of the POB representation of $\frac{d}{dt}\widehat{x}$ are

$$\begin{aligned} \widetilde{\widehat{x}}\mathcal{D} &= \widetilde{x}\mathcal{D} + \widetilde{\epsilon}\mathcal{D} = A\widetilde{x} + B\widetilde{u} + \widetilde{\epsilon}\mathcal{D} \\ &= A\widetilde{\widehat{x}} + B\widetilde{u} + (\widetilde{\epsilon}\mathcal{D} - A\widetilde{\epsilon}) . \end{aligned}$$

Define the matrices $\widehat{X}$ and U analogously to (23), and assume for simplicity of exposition that the derivative approximation error is negligible, i.e. the nonzero coefficients of $\frac{d}{dt}\widehat{x}$ are the entries of $\widehat{X}\mathcal{D}_{11}$. Under such assumptions, the matrix S-lemma approach illustrated in this section can be straightforwardly adapted to deal with this situation. Note that the definition of W_- depends on $(\widetilde{\epsilon}\mathcal{D} - A\widetilde{\epsilon})$, and consequently some insight into the dynamics of the unknown plant must be available to compute an error bound as in (33). The necessity of assumptions on the noise is established also in the approach developed in [13], see Example III.6, Lemma IV.10 and Corollary IV.11 therein.

VI. Quadratic Stabilization and Performance

We incorporate $\mathcal{H}_2$-performance specifications in the quadratic stabilization problem, under the assumption that the state derivative error $X\mathcal{D}_{11} - AX - BU$ is non-negligible. To cast the problem in the orthogonal representation setting of Theorem 2, define the p-dimensional continuous-time performance output z by

$$z = Cx + Du ;$$

using completeness of the basis $\{b_k\}_{k\in\mathbb{N}}$ and the fact that z is a linear function of x and u, conclude that $\widetilde{z}$ is a linear combination of $\widetilde{x}$ and $\widetilde{u}$, with the same matrices as z: $\widetilde{z} = C\widetilde{x} + D\widetilde{u}$.

The quadratic stabilization with $\mathcal{H}_2$-performance index γ problem is stated as follows: compute, if it exists, a state feedback law $u = Kx$ such that the closed loop dynamics $\frac{d}{dt}x = (A + BK)x$ are stable, and the 2-norm of the transfer function from u to z is less than a given $\gamma \in \mathbb{R}$:

$$\|(C + DK)(sI - (A + BK))^{-1}\|_2 < \gamma .$$

It is well known (see e.g. Proposition 3.13 p. 77 of [36]) that the problem is solvable if and only if there exists $P \in \mathbb{R}^{n\times n}$, $P = P^\top \succ 0$; $K \in \mathbb{R}^{m\times n}$; and $Z = Z^\top \in \mathbb{R}^{p\times p}$ such that

$$\begin{bmatrix} (A + BK)^\top P + P(A + BK) & P \\ P & -\gamma I_n \end{bmatrix} \prec 0 , \tag{42}$$

and

$$\begin{bmatrix} P & C^\top + K^\top D^\top \\ C + DK & Z \end{bmatrix} \succ 0 , \ \operatorname{trace}(Z) < \gamma . \tag{43}$$

This result leads to the following definition.

Definition 3: (X, U) are *informative for $\mathcal{H}_2$-performance* γ if there exist $K \in \mathbb{R}^{m\times n}$, $P \in \mathbb{R}^{n\times n}$, $P = P^\top \succ 0$,

and $Z = Z^\top \in \mathbb{R}^{p\times p}$ such that (42) and (43) hold for all $(\widehat{A}, \widehat{B}) \in \Sigma_{\mathfrak{D},c}$.
Define $Y := P^{-1}$ and $L = KY$; pre- and post-multiplying (42) by $\begin{bmatrix} P^{-1} & 0 \\ 0 & I_m \end{bmatrix}$ yields the equivalent matrix inequality

$$\begin{bmatrix} (AY+BL)+(AY+BL)^\top & I_n \\ I_n & -\gamma I_n \end{bmatrix} \prec 0 \; ; \qquad (44)$$

a Schur complement argument shows that this LMI is equivalent with $(AY+BL)+(AY+BL)^\top + \frac{1}{\gamma} I_n \prec 0$, equivalently with

$$\begin{bmatrix} I & A & B \end{bmatrix} \underbrace{\begin{bmatrix} -\frac{1}{\gamma} I_n & -Y & -L^\top \\ -Y & 0 & 0 \\ -L & 0 & 0 \end{bmatrix}}_{=:\mathcal{M}} \begin{bmatrix} I \\ A^\top \\ B^\top \end{bmatrix} \succ 0 \; ; \qquad (45)$$

note that $\mathcal{M}_{22} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} \preceq 0$. Pre- and post-multiplying the first LMI in (43) with $\begin{bmatrix} P^{-1} & 0 \\ 0 & I_p \end{bmatrix}$ one obtains

$$\begin{bmatrix} Y & (CY+DL)^\top \\ CY+DL & Z \end{bmatrix} \succ 0 \; . \qquad (46)$$

Define $\mathcal{N}$ by (36), $\mathcal{M}$ by (45), and recall (37)-(39). The assumptions of Proposition 5 are satisfied; we obtain the following characterization of informativity for $\mathcal{H}_2$-performance, and an LMI-based design procedure.

Theorem 6: Define $\mathcal{N}$ by (36) and $\mathcal{M}$ by (45). The following statements are equivalent:

1) (X, U) are informative for $\mathcal{H}_2$ performance γ;
2) There exists $Y \in \mathbb{R}^{n\times n}$, $Z = Z^\top \in \mathbb{R}^{p\times p}$, $L \in \mathbb{R}^{m\times n}$, $\alpha, \beta \in \mathbb{R}$ such that $Y = Y^\top \succ 0$, $\alpha \geq 0$, $\beta > 0$ and the LMIs

$$\begin{bmatrix} -\frac{1}{\gamma} I_n & -Y & -L^\top \\ -Y & 0 & 0 \\ -L & 0 & 0 \end{bmatrix} - \alpha\mathcal{N} - \begin{bmatrix} \beta I & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \succeq 0$$

$$\begin{bmatrix} Y & (CY+DL)^\top \\ CY+DL & Z \end{bmatrix} \succ 0 \; , \; \text{trace}(Z) < \gamma \; , \qquad (47)$$

are satisfied.

Moreover, if Y and L satisfy the LMIs (47), then $K := LY^{-1}$ is a feedback gain achieving $\mathcal{H}_2$-performance γ for all $(A, B) \in \Sigma_{\mathfrak{D},c}$.

Example 8 (Example 7 revisited): Define

$$C := \begin{bmatrix} 1 & 1 \end{bmatrix} \; , \; D := 1 \; .$$

We use the same data of Example 7, requiring a closed-loop $\mathcal{H}_2$-norm smaller than $\gamma = 2$. We first choose $c_{\max} = 0.556379$, and solve the LMIs (44), (46) and trace$(Z) < \gamma$ with `Yalmip`. We obtain

$$\begin{aligned} P &= Y^{-1} = \begin{bmatrix} 4.2505 \cdot 10^{11} & 9.4837 \cdot 10^{11} \\ 9.4837 \cdot 10^{11} & 4.3716 \cdot 10^{12} \end{bmatrix} \\ K &= LY^{-1} = \begin{bmatrix} -5.5372 & -12.153 \end{bmatrix} \; . \end{aligned}$$

The gain K stabilizes (A, B) and (A', B') and (A'', B''):

$$\begin{aligned} \sigma(A+BK) &= \{-2.7686 \pm 1.8676j\} \\ \sigma(A'+B'K) &= \{-2.8585 \pm 8.6227 \cdot 10^{-1} j\} \\ \sigma(A''+B''K) &= \{-3.8974, -2.2191\} \; . \end{aligned}$$

We compute the $\mathcal{H}_2$-norm with `Matlab`:

$$\begin{aligned} 1.6929 &= \|(C+DK)(sI-(A+BK))^{-1}\|_2 \\ 1.7274 &= \|(C+DK)(sI-(A'+B'K))^{-1}\|_2 \\ 1.7116 &= \|(C+DK)(sI-(A''+B''K))^{-1}\|_2 \; . \end{aligned}$$

Choosing $c_{\min} = 0.235961$ we obtain

$$\begin{aligned} P' &= Y'^{-1} = \begin{bmatrix} 4.6525 & 7.1667 \\ 7.1667 & 30.799 \end{bmatrix} \\ K' &= L'Y'^{-1} = \begin{bmatrix} -3.5299 & -5.9493 \end{bmatrix} \; , \end{aligned}$$

with closed-loop eigenvalues

$$\begin{aligned} \sigma(A+BK') &= \{-1.7649 \pm 1.3544j\} \\ \sigma(A'+B'K) &= \{-1.7004 \pm 1.0521j\} \; , \end{aligned}$$

and $\mathcal{H}_2$-norm

$$\begin{aligned} 1.2679 &= \|(C+DK')(sI-(A+BK'))^{-1}\|_2 \\ 1.2853 &= \|(C+DK')(sI-(A'+B'K'))^{-1}\|_2 \; . \end{aligned}$$

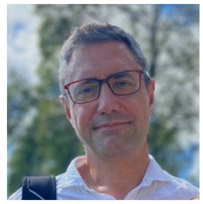
VII. Conclusions

We used polynomial approximation theory tools to transform the continuous-time input- and state signals to the sequences of coefficients of their polynomial orthogonal basis representation. In Section III we proved Theorem 2 stating that the dynamics of the original (continuous-time) input- and state signals and the representation of the transformed ones are associated with the same system matrices. We exploited this crucial result in Section IV, where we provided data-driven characterizations of continuous-time linear systems properties, and in Theorem 4 we designed state feedback controllers for continuous-time systems directly from data, without the need to directly measure the state derivative. When non-negligible derivative approximation errors occur, they can be modelled as disturbances, and treated effectively via the matrix S-lemma; we solved the quadratic stabilization problem (Theorem 5 in Section V), and the $\mathcal{H}_2$-performance problem in Section VI (Theorem 6).

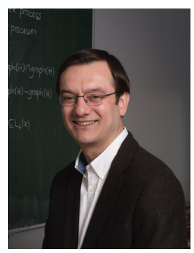
References

[1] J. Berberich and F. Allgöwer, "A trajectory-based framework for data-driven system analysis and control," in *2020 European Control Conference (ECC)*, 2020, pp. 1365–1370.
[2] J. Berberich, A. Römer, C. Scherer, and F. Allgöwer, "Robust data-driven state-feedback design," in *In Proc. Amer. contr. conf.*, 2020.
[3] J. Coulson, J. Lygeros, and F. Dörfler, "Data-enabled predictive control: In the shallows of the deepc," in *2019 18th European Control Conference (ECC)*, 2019, pp. 307–312.
[4] C. D. Persis and P. Tesi, "Formulas for data-driven control: Stabilization, optimality, and robustness," *IEEE Trans. Aut. Contr.*, vol. 65, no. 3, pp. 909–924, 2020.
[5] I. Markovsky and P. Rapisarda, "Data-driven simulation and control," *International Journal of Control*, vol. 81, no. 12, pp. 1946–1959, 2008.
[6] H. van Waarde, M. Çamlibel, and M. Mesbahi, "From noisy data to feedback controllers: non-conservative design via a matrix S-lemma," *IEEE Trans. Aut. Contr.*, vol. 67, no. 1, pp. 162–175, 2022.
[7] J. Berberich, J. Köhler, M. Müller, and F. Allgöwer., "Data-driven model predictive control with stability and robustness guarantees," *IEEE Transactions on Automatic Control*, vol. 66, no. 4, pp. 702–1717, 2021.

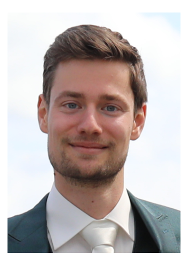
[8] D. Selvi, D. Piga, G. Battistelli, and A. Bemporad, "Optimal direct data-driven control with stability guarantees," *European Journal of Control*, vol. 59, pp. 175–187, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0947358020300741
[9] A. Bisoffi, C. D. Persis, and P. Tesi, "Data-driven control via Petersen's lemma," *Automatica*, vol. 145, p. 110537, 2022.
[10] R. Strässer, J. Berberich, and F. Allgöwer, "Data-driven control of nonlinear systems: Beyond polynomial dynamics," in *60th IEEE Conference on Decision and Control (CDC)*, 2021, pp. 4344–4351.
[11] J. Berberich, S. Wildhagen, M. Hertneck, and F. Allgöwer, "Data-driven analysis and control of continuous-time systems under aperiodic sampling," in *IFAC-PapersOnLine (Extended version at* `arXiv:2011.09221`*)*, vol. 54, no. 7, 2021, pp. 210–215.
[12] M. Bianchi, S. Grammatico, and J. Cortés, "Data-driven stabilization of switched and constrained linear systems," 2208.11392 arXiv, 2022.
[13] J. Eising and J. Cortés, "When sampling works in data-driven control: Informativity for stabilization in continuous time," 2301.10873 arXiv, 2023.
[14] T. Dai and M. Sznaier, "A semi-algebraic optimization approach to data-driven control of continuous-time nonlinear systems," *IEEE Control Systems Letters*, vol. 5, no. 2, pp. 487–492, 2021.
[15] J. Miller and M. Sznaier, "Data-driven gain scheduling control of linear parameter-varying systems using quadratic matrix inequalities," *IEEE Control Systems Letters*, vol. 7, pp. 835–840, 2023.
[16] A. Bisoffi, C. De Persis, and P. Tesi, "Data-based stabilization of unknown bilinear systems with guaranteed basin of attraction," *Systems & Control Letters*, vol. 145, p. 104788, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0167691120301705
[17] T. Mühlpfordt, R. Findeisen, V. Hagenmeyer, and T. Faulwasser, "Comments on truncation errors for polynomial chaos expansions," *IEEE Control Systems Letters*, vol. 2, no. 1, pp. 169–174, 2018.
[18] G. Pan, R. Ou, and T. Faulwasser, "On a stochastic fundamental lemma and its use for data-driven optimal control," *IEEE Transactions on Automatic Control*, pp. 1–16, 2022.
[19] J. P. Boyd, *Chebyshev and Fourier Spectral Methods*. Dover, 2001.
[20] C. Canuto, M. Y. Hussaini, A. Quarteroni, and T. A. Zang, *Spectral Methods: Fundamentals in Single Domains*, ser. Scientific Computation. Berlin Heidelberg: Springer-Verlag, 2006.
[21] J. Shen, T. Tang, and L.-L. Wang, *Spectral Methods: Algorithms, Analysis and Applications*, ser. Springer Series in Computational Mathematics. Springer-Verlag, 2011, vol. 41.
[22] A. Deb, G. Sarkar, and A. Sengupta, *Triangular orthogonal functions for the analysis of continuous time systems*. Anthem Press, 2011.
[23] H. Garnier, M. Mensler, and A. Richard, "Continuous-time model identification from sampled data: Implementation issues and performance evaluation," *International Journal of Control*, vol. 76, no. 13, pp. 1337–1357, 2003.
[24] P. Paraskevopoulos, "Legendre series approach to identification and analysis of linear systems," *IEEE Trans. Aut. Contr.*, vol. 30, no. 6, pp. 585–589, 1985.
[25] ——, "System analysis and synthesis via orthogonal polynomial series and Fourier series," *Mathematics and Computers in Simulation*, vol. 27, no. 5-6, pp. 453–469, 1985.
[26] J. Schoukens, "Modeling of continuous time systems using a discrete time representation," *Automatica*, vol. 26, no. 3, pp. 579–583, 1990.
[27] H. J. van Waarde, J. Eising, H. L. Trentelman, and M. K. Çamlibel, "Data informativity: A new perspective on data-driven analysis and control," *IEEE Transactions on Automatic Control*, vol. 65, no. 11, pp. 4753–4768, 2020.
[28] G. Sansone, *Orthogonal functions*, ser. Pure and Applied Mathematics. New York-London: Interscience Publishers, 1959, vol. 9.
[29] A. Gil, J. Segura, and N. Temme, *Numerical methods for special functions*. SIAM, 2007.
[30] W. Bell, *Special Functions for Scientists and Engineers*. Van Nostrand, 1968.
[31] L. N. Trefethen, *Approximation Theory and Approximation Practice, Extended Edition*. Philadelphia, PA: Society for Industrial and Applied Mathematics, 2019.
[32] C. Canuto and A. Quarteroni, "Approximation results for orthogonal polynomials in Sobolev spaces," *Mathematics of Computation*, vol. 38, no. 157, pp. 67–86, 1982.
[33] H. van Waarde, M. Çamlibel, J. Eising, and H. Trentelman. (2023) Quadratic matrix inequalities with application to data-based control. [Online]. Available: https://arxiv.org/pdf/2203.12959.pdf
[34] T. Driscoll, N. Hale, and L. Trefethen, *Chebfun Guide*. Oxford: Pafnuty Publications, 2014.
[35] J. Lofberg, "Yalmip: a toolbox for modelling and optimization in matlab," in *Proc. IEEE Int. Conf. Robot. Autom.*, 2004, pp. 284–289.
[36] C. Scherer and S. Weiland. (2015) Linear matrix inequalities in control. [Online]. Available: https://www.imng.uni-stuttgart.de/mst/files/LectureNotes.pdf



P. Rapisarda received his Ph.D. degree at the Department of Mathematics, University of Groningen, The Netherlands. He has held lecturing position at the University of Trieste, Italy; and at the University of Maastricht, The Netherlands. He is currently Professor in Control Theory at the School of Electronics and Computer Science of the University of Southampton, United Kingdom. His research interests include data-driven simulation and control, system identification, multidimensional systems, model reduction. He was associate editor of the IEEE Transactions on Automatic Control (2017-2022) and of Systems and Control Letters (2004-16). He is associate editor of Multidimensional Systems and Signal Processing from 2012, and of the IMA Journal of Mathematical Control and Information since 2016.



M.L. Çamlibel M. Kanat Çamlibel received the Ph.D. degree in mathematical theory of systems and control from Tilburg University in 2001. From 2001 to 2005, he held post-doctoral positions at the University of Groningen and Tilburg University. Between 2005 and 2007, he was an assistant professor at Eindhoven University of Technology. In 2007, he joined the University of Groningen where he is currently a full professor at the Bernoulli Institute for Mathematics, Computer Science, and Artificial Intelligence. His research interests include differential variational inequalities, complementarity problems, optimization, piecewise affine dynamical systems, switched linear systems, constrained linear systems, multi-agent systems, model reduction, geometric theory of linear systems, and data-driven control. Dr. Camlibel is an associate editor of the IEEE Transactions on Automatic Control, and is the past associate editor of International Journal of Robust and Nonlinear Control, SIAM Journal on Control and Optimization, and Systems and Control Letters.



Henk J. van Waarde is an assistant professor in the Bernoulli Institute for Mathematics, Computer Science and Artificial Intelligence at the University of Groningen in The Netherlands. During 2020-2021 he held postdoctoral researcher positions at the University of Cambridge, UK, and at ETH Zürich, Switzerland. He obtained the master degree *summa cum laude* and Ph.D. degree *cum laude* in Applied Mathematics from the University of Groningen in 2016 and 2020, respectively. He was also a visiting researcher at University of Washington, Seattle in 2019-2020. His research interests include data-driven control, system identification and identifiability, networks of dynamical systems, and robust and optimal control. He was the recipient of the 2021 IEEE Control Systems Letters Outstanding Paper Award.