Nonparametric Control Koopman Operators

Petar Bevanda, Bas Driessen, Lucian Cristian Iacob, Stefan Sosnowski, Roland Tóth and Sandra Hirche

***Abstract*—This paper presents a novel Koopman (composition) operator representation framework for control systems in reproducing kernel Hilbert spaces (RKHSs) that is free of explicit dictionary or input parametrizations. By establishing fundamental equivalences between different model representations, we are able to close the gap of control system operator learning and infinite-dimensional regression, enabling various empirical estimators and the connection to well-understood learning theory in RKHSs under one unified framework. As a consequence, our proposed framework allows for arbitrary accurate finite-rank approximations in infinite-dimensional spaces and leads to finite-dimensional predictors without apriori restrictions to a finite span of functions or inputs. To enable applications to high-dimensional control systems, we improve the scalability of our proposed control Koopman operator estimates by utilizing sketching techniques. Numerical experiments demonstrate superior prediction accuracy compared to bilinear EDMD, especially in high dimensions. Finally, we show that our learned models are readily interfaced with linear-parameter-varying techniques for model predictive control.**

***Index Terms*—Data-driven modeling, Nonlinear control systems, Kernel methods, Machine learning, Koopman operators, Reproducing kernel Hilbert spaces**

I. Introduction

RECENT years have seen an ever-growing interest across different fields in constructing operator-theoretic models that can provide *global* insight into physical or biological characteristics of observed phenomena [1], facilitating tractable analysis and control design for nonlinear dynamics [2]–[5]. While, historically, physical insight based on *first principles* was the driving force in modeling, increasing system complexity [6], [7] limits their utility for modeling in engineering, necessitating the use of data-driven methods. A promising framework that has recently reemerged and gained traction in the data-driven modeling community is based on the Koopman operator [8], whose spectral decomposition can enable linear superposition of signals for possibly highly nonlinear systems [5]. This representational simplicity of dynamics inspired a bevy of applications from system identification [9]–[11], soft robotics [12], [13], optimal control [14]–[16], to name a few.

Koopman-based representations for control systems: As the Koopman framework was originally developed for autonomous systems, to accommodate control inputs, different methods have been proposed. These range from heuristically selecting a linear time-invariant (LTI) model class [17], having a finite set of input values and describing a switched model [18] or analytically deriving the lifted representation [19]. It has become established that control-affine systems can be written as bilinear lifted models under certain conditions, at least in continuous-time. The authors of [20] show that for both continuous- and discrete-time systems with inputs, an invariant Koopman form can be analytically derived, granted that the autonomous part is exactly embedded. The resulting model class contains a state and input-dependent input contribution, which is often not bilinear, especially in the discrete-time case. Thus, a globally exact finite-dimensional representation generally requires a non-affine nonlinear control effect or a recasting to a *linear parameter-varying* (LPV) model form. While this has been shown on an analytic level for finite-dimensional Koopman operator-based representations, it is an open question if nonlinear input terms are unavoidable in the infinite-dimensional case and if approximation errors could be handled under certain but general assumptions.

Data-driven operator-based approaches for control: A number of deep learning-enhanced, yet parametric, methods [27]–[29] have been proposed to capture nonlinear data relations, but commonly lack rigorous learning-theoretic foundations. In contrast, kernel-based operator learning provides a powerful alternative [30] that is mathematically and implementation-wise simple, but offers rigorously established avenues for approximation error analysis. Unsurprisingly, the aforementioned has lead to a recent increase in learning-theoretic understanding of nonparametric Koopman operator-based models [31]–[33] for autonomous systems. Nevertheless, their control system counterparts do not enjoy a comparable level of understanding. For example, Hilbert space Koopman operators for control systems are not defined to full generality in existing literature, often requiring restrictive arguments involving specific generator discretizations [24], [26] or state inflation [17]. This impedes a connection to strong approximation results and learning in a flexible nonparametric (dictionary-free) manner.

A major technical reason for the aforementioned theoretical gap can also can be traced back to a rather practical desire for finite-dimensional models. However, early discretization of the learning problem inspired by finite-section methods, e.g., by using an explicit and fixed feature or input dependence [25], [34]–[37], lead to a systematic loss of precision and inefficient exploitation of the data [38]. The use of data-independent finite-dimensional subspace is especially ill-suited when dealing with unknown large-scale systems that require a suitably large/rich feature or input space. As summarized in Table I, existing data-driven operator-theoretic control system models do not enable input and output spaces to be jointly infinite-dimensional and require multiple operator regressions. This exclude data-driven settings, e.g., learning from a dataset,

This work was supported by the European Union's Horizon program under grant agreement No. 101093822, "SeaClear2.0".

P. Bevanda, S. Sosnowski, and S. Hirche, are with Chair of Information-oriented Control, TU München, Germany. {petar.bevanda, sosnowski, hirche}@tum.de.

L.C. Iacob, and R. Tóth are with the Control Systems Group, TU/e, Eindhoven, The Netherlands {l.c.iacob@, r.toth@}tue.nl. R. Tóth is also affiliated with the Systems and Control Laboratory, HUN-REN Institute for Computer Science and Control, Budapest, Hungary.

TABLE I
COMPARISON OF EXISTING LINEAR OPERATOR LEARNING APPROACHES FOR CONTROL SYSTEMS. OUR CKOR APPROACH IS BASED ON RISK MINIMIZATION AND WORKS WITH INFINITE-DIMENSIONAL AUTONOMOUS AND CONTROL EFFECTS.

APPROACH	controls	$\dim$(input space)	$\dim$(output space)	|datasets|	risk notion	ERM	scalability
switched $\boldsymbol{u}$ [18]	quantized	finite	finite	$n_u{+}1$	✗	✗	✗
bEDMD: [21], [22];	arbitrary	finite	finite	1	✗	✗	✓
$\{\text{EDMD}(\boldsymbol{u}_i)\}_{i=0}^{n_u}$ [23]–[26]	constant	finite	finite	$n_u{+}1$	✗	✗	✗
cKOR (ours)	**arbitrary**	∞	∞	**1**	$\geq \|\text{error}\|_{\text{op}}^2$	✔	✔

ensuring sufficient exploration or safe data collection.

Our proposition: To alleviate the theoretical as well as practical limitations mentioned above, we formalize a novel dictionary-free learning approach using reproducing kernel Hilbert spaces. We connect infinite-dimensional regression with composition operators of control systems to provide a rigorous and self-contained nonparametric learning framework. This turns out to be crucial to get hold of both the approximation error as well as avoiding explicit tensor products of the dictionary and the control inputs [22], [39], leading to arbitrary accurate operator approximation. As our numerical experiments confirm, this has strong empirical implications: our nonparametric approach significantly outperforms classically used (bilinear) EDMD approaches, that commonly aim at approximating finite-element methods from data [17], [40]. Moreover, our framework does not need commonly unavailable derivative information, approximate interpolation of different fixed-input Koopman operators or multiple regressions found in existing works [18], [25], [26].

A. Contributions

- Derive the control-affine RKHSs (Theorem 1), establishing equivalent operator-theoretic models (Corollary 1).
- Connect regression risk (Lemma 1) to operator norm error and prove arbitrary accurate control system operator approximation under minimal assumptions (Theorem 2).
- Prove that any RKHS observable admits arbitrarily accurate prediction using finite-rank operators (Corollary 2).
- Turn finite-rank infinite-dimensional operators to closed-form finite-dimensional predictors (Proposition 1 and 2).
- Statistically confirm the advantage of our nonparametric framework on various prediction tasks.
- Show that our nonparametric models can be directly used in computationally efficient iterated LPV model predictive control (MPC) methods.

B. Notation

For non-negative integers n and m, $[m,n] = \{m, m{+}1, \ldots, n\}$ with $n \geq m$ gives an interval set of integers. We use the shorthand $[n] \triangleq [1,n]$. Given two separable Hilbert spaces $\mathcal{F}$ and $\mathcal{Y}$, we let $\text{HS}(\mathcal{F},\mathcal{Y})$ be a Hilbert space of Hilbert-Schmidt (HS) operators from $\mathcal{F}$ asfasdto $\mathcal{Y}$ endowed with the norm $\|A\|_{\text{HS}(\mathcal{F},\mathcal{Y})}^2 \equiv \sum_{i\in\mathbb{N}} \|Ae_i\|_{\mathcal{Y}}^2 = \text{Tr}(A^*A)$, where $\{e_i\}_{i\in\mathbb{N}}$ is an orthonormal basis of $\mathcal{F}$. For HS-operators from $\mathcal{F}$ to itself, we use the shorthand $\text{HS}(\mathcal{F})$. The operator norm of a linear operator $\mathcal{G} : \mathcal{F} \to \mathcal{Y}$ is denoted as $\|\mathcal{G}\|_{\text{op}} \triangleq \sup_{\|f\|_{\mathcal{F}}=1} \|\mathcal{G}f\|_{\mathcal{Y}}^2$. For simplicity, with a slight abuse of notation, adjoints of operators as well as (conjugate) transposes of matrices are denoted as $(\cdot)^*$. Lower/upper case symbols denote functions/operators while bold symbols reserved for matrices and vectors. The space of square-integrable functions is denoted as $L_\mu^2(\cdot)$ with an appropriate Lebesgue measure μ while the vector space of bounded continuous functions is denoted by $C_0(\cdot)$. We let $k : \mathbb{X}\times\mathbb{X} \to \mathbb{R}$ be a symmetric and positive definite kernel function and $\mathcal{H}$ the corresponding RKHS [41], with norm denoted as $\|\cdot\|_{\mathcal{H}} = \sqrt{\langle\cdot,\cdot\rangle_{\mathcal{H}}}$. For $\boldsymbol{x} \in \mathbb{X}$, we use $k_{\boldsymbol{x}} \equiv k(\cdot,\boldsymbol{x}) \in \mathcal{H}$ to denote the canonical feature map.

II. PROBLEM STATEMENT

Consider an unknown nonlinear control-affine system

$$\boldsymbol{x}_{k+1} = \boldsymbol{f}(\boldsymbol{x}_k) + \textstyle\sum_{j=1}^{n_u} \boldsymbol{g}_j(\boldsymbol{x}_k)u_j(k) \equiv \mathbf{f}(\boldsymbol{x}_k,\boldsymbol{u}_k) \quad \text{(NCAS)}$$

where $\boldsymbol{x}_k \in \mathbb{X} \subset \mathbb{R}^{n_x}$ is the state and $\boldsymbol{u}_k \triangleq [u_1(k)\cdots u_{n_u}(k)]^\intercal \in \mathbb{U} \subset \mathbb{R}^{n_u}$ denotes the control variable and $k \in \mathbb{N}_0^+$ is the discrete time. Throughout, we consider $\mathbf{f}$ to be continuously differentiable and the sets $\mathbb{X}$ and $\mathbb{U}$ to be compact. Control-affinity in (NCAS) is often sufficient to represent the dynamical behavior of systems in many engineering applications [42]. Additionally, more general representation of nonlinear systems characterized by unstructured $\mathbf{f}$, under mild conditions, can be rewritten in the control-affine form (NCAS) according to the procedure detailed in [42].

Objective: Contrasting with the classical representation in the (immediate) state-space $\mathbb{X}$, one may describe the dynamics of functions over $\mathbb{X}$ based on the so-called composition (often termed Koopman) operators. From the perspective of operator theory, (NCAS) only describes the dynamics of a single function over $\mathbb{X}$: the identity function. Nevertheless, the knowledge of (NCAS) allows us to capture more than the state-space dynamics, i.e., *describe the dynamics of functions* (observables) that evolve under the dynamics. Hence, we consider a *control Koopman operator* $\mathcal{G} = \left[\begin{smallmatrix}\mathcal{A}\\ \mathcal{B}\end{smallmatrix}\right] : L_{\mu'}^2(\mathbb{X}) \to L_\eta^2(\mathbb{X}\times\mathbb{U})$ describing tensor-product dynamics between Hilbert spaces

$$y_{k+1}(\boldsymbol{x}_k,\boldsymbol{u}_k) \triangleq \left\langle \mathcal{G}y_k, \left[\begin{smallmatrix}\varphi(\boldsymbol{x}_k)\\ \boldsymbol{u}_k\otimes\varphi(\boldsymbol{x}_k)\end{smallmatrix}\right]\right\rangle_{L_\eta^2} \quad \text{(TPS)}$$

for any observable $y \in L_{\mu'}^2$, where the sufficiency of a tensor-product basis is a common assumption [19], [21], [22], [43], [44] for a system in the form of (NCAS). For brevity, we often denote $\boldsymbol{z} \triangleq \left[\begin{smallmatrix}\boldsymbol{x}\\ \boldsymbol{u}\end{smallmatrix}\right]$ so that $\mathbb{Z} \triangleq \mathbb{X}\times\mathbb{U}$. Compared to the classical Koopman operator setting, the true *operator maps between fundamentally different spaces*, formally defined as follows.

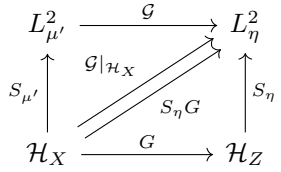


Fig. 1. The diagram for the approximation $\mathcal{G}|_{\mathcal{H}_X} \approx S_\eta G$.

Definition 1 (Control Koopman Operator): The linear operator $\mathcal{G} : L^2_{\mu'}(\mathbb{X}) \to L^2_\eta(\mathbb{X}\times\mathbb{U})$ maps *input-independent to control-affine observables* through a composition with (NCAS)

$$[\mathcal{G}f](\boldsymbol{x},\boldsymbol{u}) = f \circ \mathbf{f}(\boldsymbol{x},\boldsymbol{u}) \quad \text{for all } (\boldsymbol{x},\boldsymbol{u}) \in \mathbb{X}\times\mathbb{U}, \quad \text{(CKO)}$$

where $\eta \triangleq \mu \times \nu$ the product measure on $\mathbb{X}\times\mathbb{U}$.

Assumption 1: To ensure well-posedness, we require that (CKO) is bounded, i.e., $\exists \mathsf{L},\ \text{s.t. } \|\mathcal{G}f\|_{L^2_\eta} \le \mathsf{L}\,\|f\|_{L^2_{\mu'}}$.

Remark 1: For Lipschitz unforced dynamics $\boldsymbol{f}(\boldsymbol{x}) \equiv \mathbf{f}(\cdot,\mathbf{0})$, the above assumption is readily satisfied as they are locally invertible $|\det(\nabla \boldsymbol{f}(\boldsymbol{x}))| \neq 0$ (singularity-free) almost-everywhere, describing the transient behavior of large classes of cyber-physical systems [33], [45]. This extends to (NCAS) though Lipschitz continuity and the compactness of $\mathbb{X}$, $\mathbb{U}$, allowing a bounded *measure distortion* $\eta(\{\boldsymbol{z} \in \mathbb{X}\times\mathbb{U} : \mathbf{f}(\boldsymbol{z}) \in \mathbb{A}\}) \le \mathsf{L}^2\mu'(\mathbb{A})$ for any Borel set $\mathbb{A} \subseteq \mathbb{X}$, in turn, satisfying Assumption 1.

In practice, the operator $\mathcal{G}$ is only observed through the samples of its action on a hypothesis space. This renders the true domain and range unknown, as inferring one measure from the other would require complete governing equations—which are often unavailable. To make estimation over infinite-dimensional spaces feasible, we *restrict the operator* to a reproducing kernel Hilbert space (RKHS) $\mathcal{G}_{|_{\mathcal{H}_X}} : \mathcal{H}_X \to L^2_\eta$ to learn a *finite-rank* operator $G\colon \mathcal{H}_X \to \mathcal{H}_Z$ based on a data

$$\mathbb{D}_n = \left\{\boldsymbol{x}_{k_i+1} \equiv \mathbf{f}(\boldsymbol{x}_{k_i},\boldsymbol{u}_{k_i}), \left[\begin{smallmatrix}\boldsymbol{x}_{k_i}\\ \boldsymbol{u}_{k_i}\end{smallmatrix}\right]\right\}_{i\in[n]} = \left\{\boldsymbol{x}_+^{(i)},\boldsymbol{z}^{(i)}\right\}_{i\in[n]} \quad (1)$$

where an approximation $G \approx \mathcal{G}_{|_{\mathcal{H}}}$:

1) Rigorously follows from infinite-dimensional regression;
2) Avoids the curse of input dimensionality;
3) Allows for control-uniform observable prediction;
4) Integrates with various estimators for (CKOR);
5) Rigorously turns rank- into dimension-finite models;
6) Scales to large datasets $\mathbb{D}_n$ using sketching [46], [47].

III. Operator Learning in Infinite-Dimensions

Our approach is rooted in the well-studied theory of infinite-dimensional regression in (reproducing kernel) Hilbert spaces, leading to an RKHS-valued regression problem for learning control Koopman operators in a flexible yet principled manner.

RKHSs as subspaces: We consider RKHSs $\mathcal{H}_X$/$\mathcal{H}_Z$ that are a subset of $L^2_{\mu'}$/L^2_η-integrable functions [41, Chapter 4.3] with associated canonical feature maps $k_{\boldsymbol{x}} : \mathbb{X} \to \mathcal{H}_X$ and $k_{\boldsymbol{z}} : \mathbb{Z} \to \mathcal{H}_Z$, so that one can approximate the (CKO) *restriction* $\mathcal{G}|_{\mathcal{H}_X} : \mathcal{H}_X \to L^2_\eta$ with an operator $G : \mathcal{H}_X \to \mathcal{H}_Z$, as shown in Figure 1. Notice that despite $\mathcal{H}_X \subset L^2_{\mu'}, \mathcal{H}_Z \subset L^2_\eta$, the two spaces have different norms. To account for this discrepancy, we introduce *inclusions*

$$S_\eta : \mathcal{H}_Z \hookrightarrow L^2_\eta \quad \text{s.t.} \quad \mathcal{H}_Z \ni g \mapsto [g]_\sim \in L^2_\eta, \quad (2a)$$
$$S_{\mu'} : \mathcal{H}_X \hookrightarrow L^2_{\mu'} \quad \text{s.t.} \quad \mathcal{H}_X \ni f \mapsto [f]_\sim \in L^2_{\mu'} \quad (2b)$$

that send RKHS elements to their point-wise equal equivalence class $[\cdot]_\sim$ endowed with an appropriate $L^2_{(\cdot)}$-norm. Moreover, the inclusions admit adjoints $S^*_\eta : L^2_\eta \ni g \mapsto \int_{\mathbb{Z}} g(\boldsymbol{z})k_{\boldsymbol{z}}\eta(d\boldsymbol{z}) \in \mathcal{H}_Z$ and $S^*_{\mu'} : L^2_{\mu'} \ni f \mapsto \int_{\mathbb{X}} f(\boldsymbol{x})k_{\boldsymbol{x}}\mu'(d\boldsymbol{x}) \in \mathcal{H}_X$. To avoid being caught up in measurability and integrability issues in defining our learning approach, we impose the following technical assumptions.

Assumption 2: We impose the following requirements on the previously defined RKHSs and kernels:

1) $\mathcal{H}_Z, \mathcal{H}_X$ are separable: this satisfied if $\mathbb{X}$ and $\mathbb{Z}$ are Polish spaces and the kernels defining $\mathcal{H}_Z, \mathcal{H}_X$ are continuous;
2) $k_{\boldsymbol{x}}$ and $k_{\boldsymbol{z}}$ are measurable for μ'-almost all $\boldsymbol{x} \in \mathbb{X}$ and η-almost all $\boldsymbol{z} \in \mathbb{Z}$;
3) $k(\boldsymbol{x},\boldsymbol{x})$ and $k(\boldsymbol{z},\boldsymbol{z})$ are bounded for μ'-almost all $\boldsymbol{x} \in \mathbb{X}$ and η-almost all $\boldsymbol{z} \in \mathbb{Z}$, respectively.

The above assumptions are not restrictive in practice, as well-known kernels, e.g., Gaussian, Laplacian or Matérn kernels [41] satisfy all of the above assumptions over finite-dimensional Euclidean domains [48]. Moreover, is well-established that, under Assumptions 1, the inclusions S_η and $S_{\mu'}$ are Hilbert-Schmidt (HS) operators [41, Chapter 4.3], so $\mathcal{G}|_{\mathcal{H}_X} \in \mathrm{HS}\left(\mathcal{H}_X, L^2_\eta\right)$ by Assumption 2 [49].

Defining a notion of risk: To learn G, it is natural to measure the *squared* L^2_η *error* of the residual $\mathcal{G}|_{\mathcal{H}_X} - S_\eta G$ – as Figure 1 suggests – over a separable $\mathcal{H}_X$, i.e., its orthonormal basis $\sum_i \|[(\mathcal{G}|_{\mathcal{H}_X} - S_\eta G)h_i](\cdot)\|^2_{L^2_\eta}$. By the Hilbert-Schmidt norm definition, this amounts to

$$\begin{aligned}\mathsf{R}(G) &= \|\mathcal{G}|_{\mathcal{H}_X} - S_\eta G\|^2_{\mathrm{HS}(\mathcal{H}_X,L^2_\eta)} \qquad \text{(RISK)}\\ &= \underbrace{\|P_{\mathcal{H}_Z}\mathcal{G}|_{\mathcal{H}_X} - G\|^2_{\mathrm{HS}(\mathcal{H}_X,L^2_\eta)}}_{\text{projected risk}} + \underbrace{\|[I - P_{\mathcal{H}_Z}]\mathcal{G}|_{\mathcal{H}_X}\|^2_{\mathrm{HS}(\mathcal{H}_X,L^2_\eta)}}_{\text{representation risk}},\end{aligned}$$

where $P_{\mathcal{H}_Z}$ is the orthogonal projector in L^2_η onto $\mathcal{H}_Z$. The above decomposition is classical in learning theory [41], [50] and operator regression [31], [32], [49], [51]. While the *projected risk* depends on the learned G, mitigating the, $\mathcal{H}_Z$-dependent, *representation risk* is crucial for achieving statistical consistency.

While natural from an operator-theoretic perspective, (RISK) suggests computing the squared error over an orthonormal basis $\mathcal{H}_X$, which does not immediately reveal a loss function in the conventional sense. Nevertheless, our supervised learning label is equivalently *the target embedding* $k_{\mathbf{f}(\cdot)}\colon \mathbb{Z} \to \mathcal{H}_X$ of (CKO) in RKHS such that the *reproducing property* $[\mathcal{G}|_{\mathcal{H}_X}h](\boldsymbol{z}) = \langle h, k_{\mathbf{f}(\boldsymbol{z})}\rangle_{\mathcal{H}_X}$ holds. Then, by Fubini's theorem, the risk $\mathsf{R}(G)$ can rewritten via RKHS embeddings $k_{\mathbf{f}(\cdot)} \in \mathcal{H}_X, k_{(\cdot)} \in \mathcal{H}_Z$ w.r.t. the Bochner space of square-integrable $\mathcal{H}_X$-valued functions

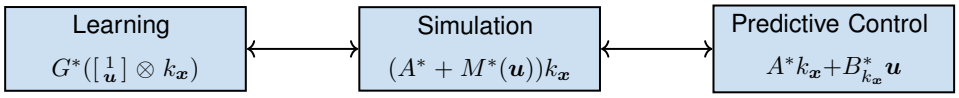


Fig. 2. Equivalent control operator-induced RKHS embeddings of the system dynamics for various tasks.

$\int_{\mathbb{Z}} \sum_i \langle h_i, k_{\mathbf{f}(\boldsymbol{z})} - G^* k_{\boldsymbol{z}} \rangle^2_{\mathcal{H}_X} \eta(d\boldsymbol{z}) \equiv \|k_{\mathbf{f}(\cdot)} - G^* k_{(\cdot)}\|^2_{L^2_\eta(\mathcal{H}_X)}$ to define a (canonical) feature-based risk minimization

$$\min_{G:\mathcal{H}_X \to \mathcal{H}_Z} \|k_{\mathbf{f}(\cdot)} - G^* k_{(\cdot)}\|^2_{L^2_\eta(\mathcal{H}_X)}, \tag{CKOR}$$

we refer to as *control Koopman operator regression* (cKOR). Notice that the above is essentially equivalent[1] to the *mean squared error* $\mathbb{E}_{Z\sim\eta}\big[\|k_{\mathbf{f}(Z)} - G^* k_Z\|^2_{\mathcal{H}_X}\big]$ minimization. This is consistent with standard formulations in learning theory where the risk is defined as an expected loss [41], here defined over the states and control inputs, acknowledging that the predictive performance is judged across η. Note that, solving (CKOR) using suitably defined infinite-dimensional RKHSs [41], the *representation risk* can vanish, enabling arbitrary accurate learning and, in turn, prediction of $\mathcal{H}_X$-observables under control inputs. Moreover, any continuous observables can be represented to arbitrary accuracy by their embedding in universal RKHSs, covering a large class of observables whose evolution can be approximated while providing a surrogate model for the state-space dynamics (NCAS) as a special case.

IV. Nonparametric Control Koopman Operator Approximations

The *flexibility* of a nonparametric approach comes from the fact that given a universal kernel, neither the feature map nor the feature space is uniquely determined – defining a dictionary-free approach [41]. Nonetheless, a *reproducing kernel Hilbert space* (RKHS) uniquely defines a kernel (and vice versa) [41], so we impose the structure of (TPS) on the feature space of an RKHS to derive a kernel that corresponds to it. As it turns out, such a kernel is crucial for practically working with infinite-dimensional spaces and getting a hold on the approximation error of the evolution of observables in $\mathcal{H}_X$. In the conditional expectation operator setting, it is established that working in infinite-dimensional RKHSs has various benefits, e.g., overcoming the weak convergence results [17], [23] of finite-section methods, cf. Mollenhauer and Koltai [49] for a discussion. In the context of linear operator learning for control systems, a similarly flexible framework is missing, which we propose here.

A. Reproducing kernel Hilbert space representations

By the structure of (TPS), tensor product spaces [51], [52] are of particular importance to endow the Hilbert space-valued Koopman operators with control effects in a principled manner.

[1]Probabilism in the data measure η comes, e.g., through random excitation or data-samples drawn from a probability distribution. We may normalize a Lebesgue measure w.l.o.g. to construct a probability measure on compact sets.

Hilbert tensor products: For $y \in \mathcal{Y}$ and $f \in \mathcal{F}$, the bounded operator $y \otimes f \in \mathcal{L}(\mathcal{F}, \mathcal{Y})$ is the *rank-one operator*

$$\mathcal{F} \ni h \mapsto [y \otimes f](h) \triangleq \langle f, h \rangle_{\mathcal{F}} y \in \mathcal{Y}. \tag{3}$$

The Hilbert tensor (outer) product $\mathcal{Y} \otimes \mathcal{F}$ is defined to be the completion of the linear span of all such rank-one operators with respect to the inner product

$$\langle y \otimes f, y' \otimes f' \rangle_{\mathcal{Y}\otimes\mathcal{F}} \triangleq \langle f, f' \rangle_{\mathcal{F}} \langle y, y' \rangle_{\mathcal{Y}} \tag{4}$$

We will interchangeably use the isometric isomorphisms $\mathrm{HS}(\mathcal{F}, \mathcal{Y}) \cong \mathcal{Y} \otimes \mathcal{F}$ and $L^2_{\mu'}(\mathcal{Y}) \cong L^2_{\mu'}(\mathbb{R}) \otimes \mathcal{Y}$ [52, Chapter 12], and treat such spaces as essentially identical.

Control-affine kernels: To appreciate the above tensor product construction, we start by recognizing that the dynamics (TPS) satisfy the following pairing for any $y \in L^2_{\mu'}$

$$\langle y, \mathcal{G}^*([\begin{smallmatrix}1\\ \boldsymbol{u}_k\end{smallmatrix}] \otimes \varphi(\boldsymbol{x}_k)) \rangle_{L^2_{\mu'}} = \langle \mathcal{G} y, [\begin{smallmatrix}1\\ \boldsymbol{u}_k\end{smallmatrix}] \otimes \varphi(\boldsymbol{x}_k) \rangle_{L^2_\eta} \tag{5}$$

for a bounded (CKO), revealing that the image space of $\mathcal{G}$ is essentially that of vector-valued $[\begin{smallmatrix}1\\ \boldsymbol{u}_k\end{smallmatrix}]$-affine L^2_η-functions, helping us uniquely define the image RKHS $\mathcal{H}_Z$ – equivalently kernel $k : \mathbb{Z} \times \mathbb{Z} \mapsto \mathbb{R}$ – for our hypothesis $G : \mathcal{H}_X \to \mathcal{H}_Z$.

Theorem 1 (Control-affine kernel): Let $\mathcal{H}_X$ be a separable RKHS with corresponding kernel $k : \mathbb{X} \times \mathbb{X} \mapsto \mathbb{R}$ and $\boldsymbol{v} \triangleq [\begin{smallmatrix}1\\ \boldsymbol{u}\end{smallmatrix}] \in \mathbb{V} \subseteq \mathbb{R}^{n_u+1}$. Then, the completion of $\boldsymbol{v}$-affine functions is the tensor product $\mathbb{V} \otimes \mathcal{H}_X$ is defined by the kernel

$$k(\boldsymbol{z}, \boldsymbol{z}') \triangleq k([\begin{smallmatrix}\boldsymbol{x}\\ \boldsymbol{u}\end{smallmatrix}], [\begin{smallmatrix}\boldsymbol{x}\\ \boldsymbol{u}\end{smallmatrix}]') = k(\boldsymbol{x}', \boldsymbol{x})(1 + \langle \boldsymbol{u}, \boldsymbol{u}' \rangle) \tag{6}$$

that corresponds to RKHS $\mathcal{H}_Z$ and equivalently defines $[\begin{smallmatrix}1\\ \boldsymbol{u}\end{smallmatrix}]$-affine $\mathcal{H}_X$-valued observables via the operator-valued kernel $K(\boldsymbol{x}, \boldsymbol{x}') \triangleq k(\boldsymbol{x}, \boldsymbol{x}')\mathrm{Id}_{\mathbb{V}} \in \mathcal{L}(\mathbb{V})$, where $\mathcal{L}(\mathbb{V})$ is the set of bounded operators from $\mathbb{V}$ to itself.

Proof: Let $V \triangleq K_{\boldsymbol{x}}\boldsymbol{v}, V' \triangleq K_{\boldsymbol{x}'}\boldsymbol{v}'$ belong to vector-valued RKHS $\mathscr{G}$ of $\boldsymbol{v}$-affine functions defined as

$$\mathscr{G} = \overline{\mathrm{span}\{K_{\boldsymbol{x}}\boldsymbol{v} \mid \boldsymbol{v} \in \mathbb{V}, \boldsymbol{x} \in \mathbb{X}\}} \tag{7}$$

where the closure $\overline{\mathrm{span}\{\cdot\}}$ is completed w.r.t. the inner product $\langle K_{\boldsymbol{x}}\boldsymbol{v}, K_{\boldsymbol{x}'}\boldsymbol{v}' \rangle_{\mathscr{G}}$. Thus, we have $\langle V, V' \rangle_{\mathscr{G}} = \langle K_{\boldsymbol{x}}\boldsymbol{v}, K_{\boldsymbol{x}'}\boldsymbol{v}' \rangle_{\mathscr{G}} = \langle K^*_{\boldsymbol{x}'}K_{\boldsymbol{x}}\boldsymbol{v}, \boldsymbol{v}' \rangle_{\mathbb{V}} = \langle K(\boldsymbol{x}', \boldsymbol{x})\boldsymbol{v}, \boldsymbol{v}' \rangle_{\mathbb{V}} = \langle \boldsymbol{v}' \otimes k_{\boldsymbol{x}'}, \boldsymbol{v} \otimes k_{\boldsymbol{x}} \rangle_{\mathbb{V}\otimes\mathcal{H}_X} = \langle k_{\boldsymbol{x}'}, k_{\boldsymbol{x}} \rangle_{\mathcal{H}_X} \langle \boldsymbol{v}, \boldsymbol{v}' \rangle_{\mathbb{V}} = k(\boldsymbol{z}, \boldsymbol{z}')$. □

The structured kernel (6), should not be confused with parametric models that work with a finite set of observables. Namely, using an C_0-universal $\mathcal{H}_X$, guarantees that (6) induces an infinite-dimensional span of functions (7) that is dense in the space of bounded continuous as well as square-integrable control-affine functions [41].

Remark 2 (Beyond finite-dimensional controls): Theorem 1 and subsequent results directly hold for infinite-dimensional control kernels in place of $\langle \boldsymbol{u}, \boldsymbol{u}' \rangle$, e.g., by changing the kernel to $K(\boldsymbol{x}, \boldsymbol{x}') = k(\boldsymbol{x}, \boldsymbol{x}')\mathrm{Id}_{\mathcal{H}_U^1 \triangleq 1 \oplus \mathcal{H}_U}$ or equivalently $k(\boldsymbol{z}, \boldsymbol{z}') = k(\boldsymbol{x}', \boldsymbol{x})(1 + k(\boldsymbol{u}, \boldsymbol{u}'))$ where $k(\boldsymbol{u}, \boldsymbol{u}')$ may be an infinite-dimensional RKHS $\mathcal{H}_U$. This is particularly useful for

the case where an unstructured nonlinear system admits a $\left[\begin{smallmatrix}1\\ \boldsymbol{u}\end{smallmatrix}\right]$-affine reformulation [42] which is not *apriori* known.

Representational equivalence: While our hypothesized dynamics are modeled by linear operator G^*, the explicit form of the simulation model is not immediately obvious due to an "input-evolving" RKHS. Intuitively, given a fixed control value would collapse the tensor product in (5) and make the system autonomous in $\mathcal{H}_X$. We make such intuition rigorous, revealing equivalent linear parameter-varying (LPV) operator formulations of our hypothesis $G^* \in \mathrm{HS}(\mathcal{H}_X, \mathcal{H}_Z)$.

Corollary 1: Let $\mathcal{H}_X, \mathcal{H}_U$ be separable and let $\mathcal{H}_U^1 \triangleq 1 \oplus \mathcal{H}_U$ so that $\mathcal{H}_Z = \mathcal{H}_U^1 \otimes \mathcal{H}_X$ in $G^* \in \mathrm{HS}(\mathcal{H}_Z, \mathcal{H}_X)$ with $(e_i)_{i\in\mathbb{N}}$ the orthonormal basis of $\mathcal{H}_U$ and $(e_i^*)_{i\in\mathbb{N}}$ is its dual basis. Then, the following isometry

$$G^* \longleftrightarrow \left[\begin{smallmatrix}1\\ 0\end{smallmatrix}\right]^* \otimes A^* + \textstyle\sum_{i\in\mathbb{N}} \left[\begin{smallmatrix}0\\ e_i\end{smallmatrix}\right]^* \otimes B^*(e_i), \tag{8}$$

explicitly establishes the isometric isomorphisms between $\mathrm{HS}\left(\mathcal{H}_U^1 \otimes \mathcal{H}_X, \mathcal{H}_X\right) \cong {\mathcal{H}_U^1}^* \otimes \mathrm{HS}(\mathcal{H}_X) = \mathrm{HS}(\mathcal{H}_X) \oplus \mathcal{H}_U^* \otimes \mathrm{HS}(\mathcal{H}_X)$ where $A^* \triangleq A^*\left(\left[\begin{smallmatrix}1\\ 0\end{smallmatrix}\right]\right)$ and $B^*(e_i) \triangleq A^*\left(\left[\begin{smallmatrix}0\\ e_i\end{smallmatrix}\right]\right)$. Moreover $\mathrm{HS}(\mathcal{H}_X) \oplus \mathcal{H}_U^* \otimes \mathrm{HS}(\mathcal{H}_X) \cong \mathrm{HS}(\mathcal{H}_X) \oplus \mathrm{HS}(\mathcal{H}_U, \mathrm{HS}(\mathcal{H}_X))$, inducing equivalent RKHS embeddings in Figure 2.

Proof: The isometric isomorphism is a direct consequence of [52, Theorem 12.3.2. & Proposition 12.3.1.] where we factored out the control input-independent span. Applying the former results once more, we have that $\mathrm{HS}(\mathcal{H}_X) \oplus \mathcal{H}_U^* \otimes \mathrm{HS}(\mathcal{H}_X) \cong \mathrm{HS}(\mathcal{H}_X) \oplus \mathrm{HS}(\mathcal{H}_U, \mathrm{HS}(\mathcal{H}_X))$. To see the equivalence of representations in Figure 2, consider a finite-dimensional kernel $\langle \boldsymbol{u}, \boldsymbol{u}' \rangle$ for $\mathcal{H}_U$ whose orthonormal basis is the standard basis $(\boldsymbol{e}_i)_{i\in[n_u]}$ of $\mathbb{R}^{n_u}$. Then, since $\boldsymbol{u} = [\boldsymbol{e}_i \otimes \boldsymbol{e}_i](\boldsymbol{u}) = \sum_{i=1}^{[n_u]} \langle \boldsymbol{e}_i^*, \boldsymbol{u} \rangle \boldsymbol{e}_i$, we have

$$G^*(\boldsymbol{v} \otimes k_{\boldsymbol{x}}) = A^* k_{\boldsymbol{x}} + \textstyle\sum_{i=1}^{[n_u]} \langle \boldsymbol{e}_i^*, \boldsymbol{v} \rangle B^*(\boldsymbol{e}_i) k_{\boldsymbol{x}}, \tag{9a}$$

$$= (A^* + [\textstyle\sum_{i=1}^{[n_u]} B^*(\boldsymbol{e}_i) \langle \boldsymbol{e}_i^*, \boldsymbol{u} \rangle]) k_{\boldsymbol{x}} \triangleq (A^* + M^*(\boldsymbol{u})) k_{\boldsymbol{x}}, \tag{9b}$$

$$= A^* k_{\boldsymbol{x}} + \left(\textstyle\sum_{i=1}^{[n_u]} B^*(\boldsymbol{e}_i) k_{\boldsymbol{x}} \otimes \boldsymbol{e}_i^*\right) \boldsymbol{u} \triangleq A^* k_{\boldsymbol{x}} + B_{k_{\boldsymbol{x}}}^* \boldsymbol{u}, \tag{9c}$$

leading to Figure 2, and concluding the proof. □

While the above result may seem technical, *the established equivalence is of central importance* in practice. Namely, once the isometric isomorphism between two spaces is established, one typically *works with whichever space is more convenient for the problem* at hand, as those shown in Figure 2. For example, formalizing a regression problem is shown to be cumbersome with input-parameterized operators [23]–[26], particularly in infinite dimensions, while it is extremely helpful to build predictors with LPV models in mind. On the other hand, the operator-/vector-valued regression formulation via (RISK) and Theorem 1 enables straightforward and flexible regression, but it may not be immediately apparent how to explicitly form multi-step predictors. As we demonstrate, the established equivalence allows one to use the best of both perspectives: *vector-valued regression for learning and analysis* and the *LPV forms for prediction and control*. Not only do our results describe the equivalence between different representations, they imply that implicit representations for control operators are completely described through scalar-valued kernels, as summarized in Table II. Hence, our equivalence results help bridge an important gap in understanding linear control operators, enabling the full utilization of the available RKHS structure, both for analysis and learning.

B. Control Koopman operator approximations in RKHS

After constructively establishing different equivalent RKHS-based hypotheses, we now study their approximation capabilities. Recall that, under Assumptions 1 and 2, *the operator restriction is Hilbert-Schmidt* $\mathcal{G}_{|\mathcal{H}_X} \triangleq \mathcal{G} S_{\mu'} \in \mathrm{HS}\left(\mathcal{H}_X, L_\eta^2\right)$ and approximation of $\mathcal{G}$ over functions in $\mathcal{H}_X$ in *operator norm* is feasible with finite-rank operators. This is critical, as $\|\cdot\|_{\mathrm{op}}$ which measures the worst-case difference between two operators in a target (image) space, leading to bounds quantifying the maximum possible error effect on any observable from the domain $\mathcal{H}_X$. However, direct estimation in operator norm comes at a price, requiring a supremum over the unit ball in $\mathcal{H}_X$, which we cannot consistently estimate [49]. To make estimation feasible, we restrict the hypothesis space to Hilbert–Schmidt operators $\mathrm{HS}(\mathcal{H}_X, \mathcal{H}_Z)$ via reproducing kernels and interpret the image space as L_η^2, which yields a principled surrogate problem for infinite-dimensional linear regression. Building on the analysis of Mollenhauer and Koltai [49], we can guarantee the well-posedness of our (RISK) objective, demonstrating that reducing the search space to Hilbert-Schmidt operators in (CKOR) instead of a larger class, e.g., of bounded operators, (RISK) sharply bounds the $\|\cdot\|^2_{L_\eta^2 \leftarrow \mathcal{H}_X}$-error.

Lemma 1: Under Assumptions 1 and 2, (RISK) sharply satisfies $\|\mathcal{G} - G\|^2_{L_\eta^2 \leftarrow \mathcal{H}_X} \leq \mathsf{R}(G)$ for every $G \in \mathrm{HS}(\mathcal{H}_X, \mathcal{H}_Z)$.

Proof: Let $G \in \mathrm{HS}(\mathcal{H}_X, \mathcal{H}_Z)$. We have that

$$\|\mathcal{G} - G\|^2_{L_\eta^2 \leftarrow \mathcal{H}_X} = \sup_{\|f\|_{\mathcal{H}_X}=1} \|\mathcal{G}|_{\mathcal{H}_X} f - Gf\|^2_{L_\eta^2} \tag{10a}$$

$$= \sup_{\|f\|_{\mathcal{H}_X}=1} \|[\mathcal{G}|_{\mathcal{H}_X} f](\cdot) - [Gf](\cdot)\|^2_{L_\eta^2} \tag{10b}$$

$$= \sup_{\|f\|_{\mathcal{H}_X}=1} \| \left\langle f, k_{\mathbf{f}(\cdot)} \right\rangle_{\mathcal{H}_X} - \left\langle f, G^* k_{(\cdot)} \right\rangle_{\mathcal{H}_X} \|^2_{L_\eta^2} \tag{10c}$$

$$= \sup_{\|f\|_{\mathcal{H}_X}=1} \mathbb{E}_{Z\sim\eta} \left[\left\langle f, k_{\mathbf{f}(Z)} - G^* k_Z \right\rangle^2_{\mathcal{H}_X} \right] \tag{10d}$$

$$\leq \sup_{\|f\|_{\mathcal{H}_X}=1} \mathbb{E}_{Z\sim\eta} \left[\|f\|^2_{\mathcal{H}_X} \left\| k_{\mathbf{f}(Z)} - G^* k_Z \right\|^2_{\mathcal{H}_X} \right] \tag{10e}$$

where we use the definition (CKO) and the reproducing property in (10b)-(10c) together with the Cauchy-Schwarz inequality in (10e). We say the upper bound is sharp if it holds for all $f \in \mathcal{H}_X$ and $\exists h \in \mathcal{H}_X$ so $\mathsf{R}(G) = \|\mathcal{G} - G\|^2_{L_\eta^2 \leftarrow \mathcal{H}_X}$. Hence, the above bound is sharp by considering that we have η-a.e. $k_{\mathbf{f}(Z)} - G^* k_Z = e$ for some constant $e \in \mathcal{H}_X$ so the equality is attained by setting $f = e/\|e\|_{\mathcal{H}_X}$ in the supremum. □

With the help of Lemma 1, we obtain an operator norm approximation guarantee, as the error can be made arbitrarily small when using a universal RKHS $\mathcal{H}_X$.

Theorem 2 (Arbitrary accuracy): Let Assumptions 1 and 2 hold. Also let $k : \mathbb{Z} \times \mathbb{Z} \mapsto \mathbb{R}$ be associated with an RKHS $\mathcal{H}_Z \cong \mathbb{V} \otimes \mathcal{H}_X$, and let $P_{\mathcal{H}_Z}$ be the orthogonal projector onto the closure of $\mathrm{Im}(S_\eta) \subseteq L_\eta^2(\mathbb{Z})$. Then, for every $\delta > 0$, there is a *finite-rank* $G \in \mathrm{HS}(\mathcal{H}_X, \mathcal{H}_Z)$, such that

1) $\underbrace{\mathsf{E}(G) \triangleq \|\mathcal{G} - G\|_{L_\eta^2 \leftarrow \mathcal{H}_X}}_{\text{operator norm error}} < \underbrace{\|[I - P_{\mathcal{H}_Z}]\mathcal{G}\|_{L_\eta^2 \leftarrow \mathcal{H}_X}}_{\text{representation bias } \mathsf{B}(\mathcal{H}_Z)} + \delta$,
2) $\mathsf{E}(G) < \delta$ if and only if $\mathsf{B}(\mathcal{H}_Z) = 0$. Consequently, when $\mathcal{H}_X$ *is a* C_0*-universal RKHS*, $\mathsf{B}(\mathcal{H}_Z) = 0$.

TABLE II
POPULAR EXISTING OPERATOR-THEORETIC REPRESENTATIONS FOR CONTROL SYSTEMS. cKOR INCLUDES ALL CONTROL-AFFINE REPRESENTATIONS WITH STRAIGHTFORWARD MODIFICATION OF SAMPLING/EMBEDDING OPERATORS.

METHOD	embedding dynamics	output kernel $\langle\cdot,\cdot\rangle_{\mathcal{H}_Z}$	input kernel $\langle\cdot,\cdot\rangle_{\mathcal{H}_X}$
DMD [53]	$\boldsymbol{Ax}$	$\langle \boldsymbol{x},\boldsymbol{x}'\rangle$	$\langle \boldsymbol{x},\boldsymbol{x}'\rangle$
DMDc [35]	$\boldsymbol{Ax}+\boldsymbol{Bu}$	$\langle \boldsymbol{x},\boldsymbol{x}'\rangle+\langle \boldsymbol{u},\boldsymbol{u}'\rangle$	$\langle \boldsymbol{x},\boldsymbol{x}'\rangle$
kEDMD [54], [55]	$A^*k_{\boldsymbol{x}}$	$k(\boldsymbol{x},\boldsymbol{x}')$	$k(\boldsymbol{x},\boldsymbol{x}')$
kEDMDc [56]	$A^*k_{\boldsymbol{x}}+B^*\boldsymbol{u}$	$k(\boldsymbol{x},\boldsymbol{x}')+\langle \boldsymbol{u},\boldsymbol{u}'\rangle$	$k(\boldsymbol{x},\boldsymbol{x}')$
cKOR	$A^*k_{\boldsymbol{x}}+B^*(k_{\boldsymbol{u}}\otimes k_{\boldsymbol{x}})$	$k(\boldsymbol{x},\boldsymbol{x}')+k(\boldsymbol{u},\boldsymbol{u}')k(\boldsymbol{x},\boldsymbol{x}')$	$k(\boldsymbol{x},\boldsymbol{x}')$

Proof: Let $\delta>0$. By the triangle inequality

$$\begin{aligned}\mathsf{E}(G) &\le \|[I-P_{\mathcal{H}_Z}]\mathcal{G}\|_{L^2_\eta\leftarrow\mathcal{H}_X}+\|P_{\mathcal{H}_Z}\mathcal{G}-G\|_{L^2_\eta\leftarrow\mathcal{H}_X}\\ &\le \mathsf{B}(\mathcal{H}_Z)+\|S_\eta\|\|P_{\mathcal{H}_Z}\mathcal{G}-G\|_{\mathcal{H}_Z\leftarrow\mathcal{H}_X}\end{aligned}$$

where the error splits into *representation bias* $\mathsf{B}(\mathcal{H}_Z)$ and *rank reduction error* $\|P_{\mathcal{H}_Z}\mathcal{G}-G\|_{\mathcal{H}_Z\leftarrow\mathcal{H}_X}$. By the fact that finite-rank operators from $\mathcal{H}_X\to\mathcal{H}_Z$ are dense in $\mathrm{HS}(\mathcal{H}_X,\mathcal{H}_Z)$, we have that $\|S_\eta\|\|P_{\mathcal{H}_Z}\mathcal{G}-G\|_{\mathcal{H}_Z\leftarrow\mathcal{H}_X}<\delta$ so that $\mathsf{E}(G)<\mathsf{B}(\mathcal{H}_Z)+\delta$.
" $\Rightarrow$ ": For a C_0-universal RKHS $\mathcal{H}_X$ inducing $\mathcal{H}_Z$, (6) is dense in L^2_η, i.e., $\mathrm{Im}(\mathcal{G}|_{\mathcal{H}_X})\subseteq\mathrm{cl}(\mathrm{Im}(S_\eta))$ so that $\|[I-P_{\mathcal{H}_Z}]\mathcal{G}\|_{L^2_\eta\leftarrow\mathcal{H}_X}=\mathsf{B}(\mathcal{H}_Z)=0$ [41, Chapter 4].
" $\Leftarrow$ ": Let $\mathscr{T}$ be the vector-valued RKHS generated by $\mathrm{HS}(\mathcal{H}_Z,\mathcal{H}_X)\cong\mathcal{H}_Z\otimes\mathcal{H}_X$:

$$\mathscr{T}\triangleq\{T\colon\mathbb{Z}\to\mathcal{H}_X\mid T{=}G^*k_{(\cdot)},G^*{\in}\mathrm{HS}(\mathcal{H}_Z,\mathcal{H}_X)\},\quad(11)$$

induced by the kernel $K(\boldsymbol{z},\boldsymbol{z}')\triangleq k(\boldsymbol{z},\boldsymbol{z}')I_{\mathcal{H}_X}$. Then following [49, Corollary 4.5], every operator $G^*\in\mathrm{HS}(\mathcal{H}_Z,\mathcal{H}_X)$ corresponds to a function $G^*k_{\boldsymbol{z}}$ for all $\boldsymbol{z}\in\mathbb{Z}$ and vice versa. Using a C_0-universal RKHS $\mathcal{H}_X$ to define (7) that is isometrically isomorphic to $\mathcal{H}_Z$ makes the embedding $\mathscr{T}$ (11) densely embedded into $L^2_\eta(\mathcal{H}_X)\equiv L^2_\eta(\mathbb{Z},\mathcal{H}_X)\cong\mathrm{HS}\left(L^2_\eta(\mathbb{Z}),\mathcal{H}_X\right)$, and, for every $\delta>0$ we therefore have an operator $G^*\in\mathrm{HS}(\mathcal{H}_Z,\mathcal{H}_X)$ such that the bound $\left\|T(\cdot)-G^*k_{(\cdot)}\right\|^2_{L^2_\eta(\mathcal{H}_X)}\equiv\mathsf{R}(G)<\delta$ holds. Applying Lemma 1 we have $\mathsf{E}(G)<\mathsf{B}(\mathcal{H}_Z)+\delta=\delta$, completing the proof. □

The above result reveals that whenever the RKHS $\mathcal{H}_X$ used to define (6) is C_0-universal, then there is no representation bias/error $\mathsf{B}(\mathcal{H}_Z)=0$ and one can find arbitrarily good finite-rank approximations of control Koopman operators. Note that Assumption 2 on the RKHS $\mathcal{H}_X$ is non-restrictive and not actually an assumption on the problem – it solely depends on the choice of the kernel and is readily satisfied by popular kernels such as Gaussian, Laplacian or Matérn kernels [41]. Moreover, notice that we do not require the true operator $\mathcal{G}:L^2_{\mu'}\to L^2_\eta$ to be compact let alone Hilbert-Schmidt for Theorem 2 to hold.

C. Approximating the dynamics of observables

One of the practical appeals of control operator models is the *ability to forecast of any observable belonging to the hypothetical domain* $\mathcal{H}_X$. For that, operator norm approximation is critical, allowing arbitrarily accurate prediction of any measurement/observable $y\in\mathcal{H}_X$. We formalize this in the following corollary.

Corollary 2: Let the conditions of Theorem 2 hold and denote the true one-step evolution of an observable $y\in\mathcal{H}_X$ as $y_+(\boldsymbol{z})\triangleq[\mathcal{G}S_{\mu'}y](\boldsymbol{z})$. Then, for any $y\in\mathcal{H}_X$ and every $\varepsilon>0$, there exists a $G\in\mathrm{HS}(\mathcal{H}_X,\mathcal{H}_Z)$, so that

$$\|y_+-S_\eta Gy\|_{L^2_\eta}<\varepsilon\qquad(12)$$

Proof: Applying the Cauchy-Schwartz inequality gives

$$\|(\mathcal{G}S_{\mu'}-S_\eta G)y\|_{L^2_\eta}\le\mathsf{E}(G)\|y\|_{\mathcal{H}_X}.$$

which is well-defined by $y\in\mathcal{H}_X\Longrightarrow\|y\|_{\mathcal{H}_X}{<}\infty$. After setting $\delta=\varepsilon/\|y\|_{\mathcal{H}_X}$ in Theorem 2, the assertion follows. □
The main added assumption in the above result is $y\in\mathcal{H}_X$, which may be straightforwardly fulfilled for known observables of interest, e.g., components of the full-state observable, by adding a linear kernel component to the hypothetical domain to include the state: $\mathcal{H}^{\mathsf{id}}=\mathcal{H}\oplus\mathsf{Id}(\mathbb{X})$, which is induced by the kernel $k^{\mathsf{id}}(\boldsymbol{x},\boldsymbol{x}')=k_X(\boldsymbol{x},\boldsymbol{x}')+\langle\boldsymbol{x},\boldsymbol{x}'\rangle$. Still, such a composite RKHS is universal if $k_X(\boldsymbol{x},\boldsymbol{x}')$ is C_0-universal, allowing Theorem 2 to hold with a vanishing representation bias $\mathsf{B}(\cdot)=0$. However, the existing approaches append the state to a data-independent and finite dictionaries [17], [39] or evaluate an empirical inner product that is only an approximation of the canonical $k^{\mathsf{id}}(\boldsymbol{x},\boldsymbol{x}')$. As such, there is no guarantee for unbiased representations $\mathsf{B}(\cdot)=0$ (cf. Theorem 2), which is crucial for achieving statistical consistency [41]. Note that we do not require (CKO) to be compact (let alone Hilbert-Schmidt), while our hypothesis does not need to be invariant w.r.t. (CKO) to admit arbitrary accurate prediction.

Remark 3 (Improved analysis): In contrast to our nonparametric and discrete-time setting, existing operator approximation error analyses for control use pre-RKHS hypothesis approaches [25] even when using kernels [26] and do not directly benefit from the structure of an RKHS. Additionally, the analysis therein comes with an irreducible time-discretization error and explicitly depends on control input dimensionality, e.g., using infinite-dimensional control spaces (cf. Remark 2) would render regression and analysis infeasible for existing approaches [23]–[26]. Although the aforementioned works do focus on probabilistic finite-data error bounds for a set of constant input Koopman operators, our theoretical analysis indicates that sharper and more flexible results should be readily available. Exploring this further is, however, out of the scope of this work.

V. ESTIMATING CONTROL KOOPMAN OPERATORS OVER RKHSS

Regularized risk minimization: After establishing the Hilbert-Schmidt (HS) representations over RKHSs, we continue with formulating a well-posed regression problem to

solve (CKOR). Recognizing the fact that $L^2_\eta(\mathcal{H}_X) \cong L^2_\eta \otimes \mathcal{H}_X \cong \mathrm{HS}\left(L^2_\eta, \mathcal{H}_X\right)$, it is easy to see that (RISK) has an equivalent, Hilbert-Schmidt norm, formulation

$$\|k_{\mathbf{f}(\cdot)} - G^* k_{(\cdot)}\|^2_{L^2_\eta(\mathcal{H}_X)} \equiv \|\mathcal{G}|_{\mathcal{H}_X} - S_\eta G\|^2_{\mathrm{HS}(\mathcal{H}_X, L^2_\eta)}, \quad (13)$$

so it is apparent that it admits a unique minimizer in the form of a Galerkin projection $\left(S^*_\eta S_\eta\right)^\dagger S^*_\eta \mathcal{G}|_{\mathcal{H}_X}$, where $\dagger$ is the Moore-Penrose pseudoinverse operator. However, the latter often leads to a poorly-conditioned system of equations. To ensure well-posedness, we use Tikhonov regularization

$$\begin{aligned} G_\gamma &\triangleq \operatorname*{arg\,min}_{G \in \mathrm{HS}(\mathcal{H}_X, \mathcal{H}_Z)} \boldsymbol{R}(G) + \gamma \|G\|^2_{\mathrm{HS}} \\ &= (C_{ZZ} + \gamma \operatorname{Id}_{\mathcal{H}_X})^{-1} C_{ZX_+}, \quad \gamma > 0. \end{aligned} \quad (\mathsf{rRM})$$

where the cross-covariance C_{ZX_+} and covariance C_{ZZ} are

$$C_{ZZ} \triangleq S^*_\eta S_\eta = \int_{\mathbb{Z}} k_{\boldsymbol{z}} \otimes k_{\boldsymbol{z}} \eta(d\boldsymbol{z}) : \mathcal{H}_Z \to \mathcal{H}_Z, \quad (14a)$$

$$C_{ZX_+} \triangleq S^*_\eta \mathcal{G}|_{\mathcal{H}_X} = \int_{\mathbb{Z}} k_{\boldsymbol{z}} \otimes k_{\mathbf{f}(\boldsymbol{z})} \eta(d\boldsymbol{z}) : \mathcal{H}_X \to \mathcal{H}_Z. \quad (14b)$$

It is easy to confirm that the objective in (RRM) is continuous, coercive and strictly convex, making (RRM) a *unique* minimizer [57] of the regularized (RISK), incurring no loss of precision as existing finite-dimensional finite-section approaches [26], [58].

A. Estimating control Koopman operators from data

Since we do not have access to population level quantities to compute (RRM) in practice, we have to rely on data samples from a dataset (1). For that, we define *sampling* operators

$$\widehat{S}_Z : \mathcal{H}_Z \to \mathbb{R}^n, \quad \widehat{S}_Z g = [g(\boldsymbol{z}^{(1)}) \cdots g(\boldsymbol{z}^{(n)})]^\intercal \quad (15a)$$

$$\widehat{S}_X : \mathcal{H}_X \to \mathbb{R}^n, \quad \widehat{S}_X f = [f(\boldsymbol{x}^{(1)}) \cdots f(\boldsymbol{x}^{(n)})]^\intercal \quad (15b)$$

$$\widehat{S}_+ : \mathcal{H}_X \to \mathbb{R}^n, \quad \widehat{S}_+ f = [f(\boldsymbol{x}^{(1)}_+) \cdots f(\boldsymbol{x}^{(n)}_+)]^\intercal. \quad (15c)$$

so their adjoints, the sampled *embedding* operators [59], are defined as $\widehat{S}^*_Z : \mathbb{R}^n \to \mathcal{H}_Z$ s.t. $\boldsymbol{a} \mapsto \sum_{i=1}^n k_{\boldsymbol{z}^{(i)}}(\boldsymbol{a})_i$, $\widehat{S}^*_X : \mathbb{R}^n \to \mathcal{H}_X$ s.t. $\boldsymbol{b} \mapsto \sum_{i=1}^n k_{\boldsymbol{x}^{(i)}}(\boldsymbol{b})_i$ and $\widehat{S}^*_+ : \mathbb{R}^n \to \mathcal{H}_X$ s.t. $\boldsymbol{c} \mapsto \sum_{i=1}^n k_{\boldsymbol{x}^{(i)}_+}(\boldsymbol{c})_i$.

Now we formulate a control system operator representation from a dataset (1) leading to the empirical risk $\widehat{\boldsymbol{R}}(G)$

$$\tfrac{1}{n} \sum_{i=1}^n \|k_{\boldsymbol{x}^{(i)}_+} - G^* k_{\boldsymbol{z}^{(i)}}\|^2_{\mathcal{H}_X} \equiv \tfrac{1}{n} \|\widehat{S}_+ - \widehat{S}_Z G\|^2_{\mathrm{HS}(\mathcal{H}_Z, \mathbb{R}^n)} \quad (\widehat{\mathsf{RISK}})$$

and the regularized empirical risk minimization

$$\begin{aligned} \widehat{G}_\gamma &\triangleq \operatorname*{arg\,min}_{G \in \mathrm{HS}(\mathcal{H}_X, \mathcal{H}_Z)} \widehat{\boldsymbol{R}}(G) + \gamma \|G\|^2_{\mathrm{HS}} = (\widehat{C}_{ZZ} + \gamma \operatorname{Id}_{\mathcal{H}_X})^{-1} \widehat{C}_{ZX_+} \\ &= \left(\tfrac{1}{n} \widehat{S}^*_Z \widehat{S}_Z + \gamma \operatorname{Id}_{\mathcal{H}_X}\right)^{-1} \left(\tfrac{1}{n} \widehat{S}^*_Z \widehat{S}_+\right), \end{aligned} \quad (\widehat{\mathsf{rRM}})$$

called *kernel ridge regression* (KRR) [31], [41], [60]. While this estimator is predominantly used in this work, Appendix A includes other popular estimators from dynamical systems literature, such as *principal component* (PCR) [55], [61] and recently proposed *reduced rank* (RRR) regression [31].

Finite-dimensional predictors: When $\mathcal{H}_X$ and $\mathcal{H}_Z$ are infinite-dimensional universal RKHSs, we can not directly compute the estimate in $(\widehat{\mathsf{rRM}})$. Still, the finiteness of the data makes $(\widehat{\mathsf{rRM}})$ finite in rank, allowing practical finite-dimensional computations. Before we show a formal result on this, we introduce the *input sampling operator* $\widehat{U} : \boldsymbol{u} \mapsto [\langle \boldsymbol{u}^{(1)}, \boldsymbol{u} \rangle \cdots \langle \boldsymbol{u}^{(n)}, \boldsymbol{u} \rangle)]^\intercal$ and kernel matrices $\boldsymbol{K} \triangleq \widehat{S}_X \widehat{S}^*_X = [k(\boldsymbol{x}^{(i)}, \boldsymbol{x}^{(j)})]_{i,j \in [n]}$, $\boldsymbol{K}_Z \triangleq \boldsymbol{K} + (\widehat{U}\widehat{U}^*) \odot \boldsymbol{K}$ and $\boldsymbol{K}_+ \triangleq \widehat{S}_+ \widehat{S}^*_X = [k(\boldsymbol{x}^{(i)}_+, \boldsymbol{x}^{(j)})]_{i,j \in [n]}$, where $\odot$ denotes the element-wise (Hadamard) product.

Proposition 1 (Control Koopman Operator Models): The minimizer of $(\widehat{\mathsf{rRM}})$ equals to

$$\widehat{G}_\gamma \triangleq \widehat{S}^*_Z \boldsymbol{W} \widehat{S}_+ = \begin{bmatrix} \widehat{A}_\gamma \\ \widehat{B}_\gamma \end{bmatrix} \in \mathrm{HS}\left(\mathcal{H}_X, \mathcal{H}_Z\right)$$

where $\boldsymbol{W} \triangleq (\boldsymbol{K}_Z + n\gamma \boldsymbol{I})^{-1} \in \mathbb{R}^{n \times n}$ and the $(k \geq 1)$-step evolution of any observable $\boldsymbol{y} \in \mathcal{H}^{n_y}$ under control inputs $\boldsymbol{u}_{k \in \mathbb{N}_0} \in \mathbb{U}$ are exactly represented by the predictor

$$\mathbf{z}_1 = (\boldsymbol{I} + \boldsymbol{M}_{\boldsymbol{u}_0}) \boldsymbol{k}_{\boldsymbol{X}}(\boldsymbol{x}_0), \quad (16a)$$

$$\mathbf{z}_{k+1} = (\boldsymbol{A} + \textstyle\sum_{i=1}^{n_u} \langle \boldsymbol{e}_i, \boldsymbol{u}_k \rangle \boldsymbol{B}_i) \mathbf{z}_k, \quad (16b)$$

$$\widehat{\boldsymbol{y}}_k = \boldsymbol{C} \mathbf{z}_k, \quad (16c)$$

where $\mathbf{z}_{k \geq 1} \in \mathbb{R}^n$, $\boldsymbol{k}_{\boldsymbol{X}}(\boldsymbol{x}) = [k(\boldsymbol{x}, \boldsymbol{x}^{(1)}) \cdots k(\boldsymbol{x}, \boldsymbol{x}^{(n)})]^\intercal$ and

$$\boldsymbol{M}_{(\cdot)} \triangleq \operatorname{diag}(\widehat{U}(\cdot)), \quad \boldsymbol{Y}_+^\intercal \triangleq [\boldsymbol{y}(\boldsymbol{x}^{(1)}_+) \cdots \boldsymbol{y}(\boldsymbol{x}^{(n)}_+)] \quad (17a)$$

$$\boldsymbol{A} \triangleq (\boldsymbol{W}\boldsymbol{K}_+)^\top, \quad \boldsymbol{B}_i \triangleq \boldsymbol{M}_{\boldsymbol{e}_i} \boldsymbol{A}, \quad \boldsymbol{C} \triangleq (\boldsymbol{W}\boldsymbol{Y}_+)^\intercal. \quad (17b)$$

Proof: By the *push-through identity* and the reproducing property, we rewrite $(\widehat{\mathsf{rRM}})$ as $(\widehat{S}^*_Z \widehat{S}_Z + n\gamma I_{\mathcal{H}_Z})^{-1} \widehat{S}^*_Z \widehat{S}_+ = \widehat{S}^*_Z (\widehat{S}_Z \widehat{S}^*_Z + n\gamma \boldsymbol{I})^{-1} \widehat{S}_+ = \widehat{S}^*_Z \boldsymbol{W} \widehat{S}_+ \triangleq \widehat{G}_\gamma$ so that

$$\widehat{A}_\gamma = \widehat{S}^*_X \boldsymbol{W} \widehat{S}_+ \in \mathrm{HS}\left(\mathcal{H}_X, \mathcal{H}_X\right), \quad (18a)$$

$$\widehat{B}_\gamma = (\widehat{U} \circledast \widehat{S}_X)^* \boldsymbol{W} \widehat{S}_+ \in \mathrm{HS}\left(\mathcal{H}_X, \mathbb{R}^{n_u} \otimes \mathcal{H}_X\right), \quad (18b)$$

where $\circledast, \circledcirc$ denote the column- and row-wise Kronecker product. By definition of sampling operators (15) and their adjoints. Moreover, $\boldsymbol{K}_Z \triangleq \widehat{S}_Z \widehat{S}^*_Z \in \mathbb{R}^{n \times n}$ and equals to $\boldsymbol{K} + (\widehat{U}\widehat{U}^*) \odot \boldsymbol{K}$ after using $(\widehat{U} \circledast \widehat{S})(\widehat{U} \circledast \widehat{S})^* = (\widehat{U} \circledast \widehat{S})(\widehat{U}^* \circledcirc \widehat{S}^*) = (\widehat{U}\widehat{U}^*) \odot (\widehat{S}\widehat{S}^*)$. For $\widehat{G}^*_\gamma = \widehat{S}^*_+ \boldsymbol{W} \widehat{S}_Z$, or equivalently, $\widehat{A}^*_\gamma, \widehat{B}^*_\gamma$, we can identify the weighted forward embedding $\widehat{F}^* = \widehat{S}^*_+ \boldsymbol{W}$ so that $\boldsymbol{A} = \widehat{S}_X \widehat{F}^* \in \mathbb{R}^{n \times n}$ and consider propagating a scalar-valued observable $y \in \mathcal{H}_X$, which by definition amounts to

$$\widehat{y}_+(\boldsymbol{z}) \triangleq [\widehat{G}y](\boldsymbol{z}) = \langle \widehat{G}_\gamma y, k_{\boldsymbol{z}} \rangle_{\mathcal{H}_Z} = \langle y, \widehat{G}^*_\gamma k_{\boldsymbol{z}} \rangle_{\mathcal{H}_X} \quad (19)$$

$$= \langle y, \widehat{A}^* k_{\boldsymbol{x}} + \widehat{B}^* (\boldsymbol{u} \otimes k_{\boldsymbol{x}}) \rangle_{\mathcal{H}_X} \quad (20)$$

where by propagating canonical features for one step we obtain

$$\mathcal{H}_X \ni \widehat{k}^{t+1}_{\boldsymbol{x}} = \widehat{A}^* k^t_{\boldsymbol{x}} + \widehat{B}^* (\boldsymbol{u}^t \otimes k^t_{\boldsymbol{x}}) \quad (21)$$

$$= \widehat{F}^* (\widehat{S}_X k^t_{\boldsymbol{x}} + (\widehat{U} \circledast \widehat{S}_X)(\boldsymbol{u}^t \otimes k^t_{\boldsymbol{x}})) \quad (22)$$

$$= \widehat{F}^* (\boldsymbol{k}_{\boldsymbol{X}}(\boldsymbol{x}) + \widehat{U} \boldsymbol{u}^t \odot \widehat{S}_X k^t_{\boldsymbol{x}}) \quad (23)$$

$$= \widehat{F}^* (\boldsymbol{k}_{\boldsymbol{X}}(\boldsymbol{x}) + \widehat{U} \boldsymbol{u}^t \odot \boldsymbol{k}_{\boldsymbol{X}}(\boldsymbol{x})) \quad (24)$$

$$= \widehat{F}^* (\boldsymbol{k}_{\boldsymbol{X}}(\boldsymbol{x}) + \operatorname{diag}(\widehat{U} \boldsymbol{u}^t) \boldsymbol{k}_{\boldsymbol{X}}(\boldsymbol{x})) \quad (25)$$

$$= \widehat{F}^* (\boldsymbol{I} + \operatorname{diag}(\widehat{U} \boldsymbol{u}^t)) \boldsymbol{k}_{\boldsymbol{X}}(\boldsymbol{x}) \quad (26)$$

$$= \widehat{F}^* (\boldsymbol{I} + \boldsymbol{M}_{\boldsymbol{u}^t}) \boldsymbol{k}_{\boldsymbol{X}}(\boldsymbol{x}) \triangleq \widehat{F}^* \mathbf{z}_t \quad (27)$$

after some some algebraic manipulation and the definition of sampling and embedding operators. For $t+2$, we plug in the previous solution for $\widehat{k}_{\boldsymbol{x}}^{t+1}$

$$\mathcal{H}_X \ni \widehat{k}_{\boldsymbol{x}}^{t+2} = \widehat{A}^*\widehat{k}_{\boldsymbol{x}}^{t+1} + \widehat{B}^*(\boldsymbol{u}^{t+1}\otimes\widehat{k}_{\boldsymbol{x}}^{t+1}) \tag{28}$$
$$= \widehat{F}^*(\boldsymbol{A}\mathbf{z}_t + \widehat{U}\boldsymbol{u}^{t+1}\odot\boldsymbol{A}\mathbf{z}_t) \tag{29}$$
$$= \widehat{F}^*(\boldsymbol{A}\mathbf{z}_t + \mathrm{diag}(\widehat{U}\boldsymbol{u}^{t+1})\boldsymbol{A}\mathbf{z}_t) \tag{30}$$
$$= \widehat{F}^*(\boldsymbol{I} + \mathrm{diag}(\widehat{U}\boldsymbol{u}^{t+1}))\boldsymbol{A}\mathbf{z}_t \tag{31}$$
$$= \widehat{F}^*(\boldsymbol{I} + \boldsymbol{M}_{\boldsymbol{u}^{t+1}})\boldsymbol{A}\mathbf{z}_t. \tag{32}$$

It is straightforward to verify by induction that $\widehat{k}_{\boldsymbol{x}}^{t+H} = \widehat{F}^*(\boldsymbol{M}_{\boldsymbol{u}^{t+H-1}}\boldsymbol{A})^{H-1}\mathbf{z}_t$ so that

$$\widehat{y}_{t+H}(\boldsymbol{z}_t) = \langle y, \widehat{F}^*((\boldsymbol{I} + \boldsymbol{M}_{\boldsymbol{u}^{t+H-1}})\boldsymbol{A})^{H-1}\mathbf{z}_t\rangle_{\mathcal{H}_X} \tag{33}$$
$$= \langle \widehat{F}y, ((\boldsymbol{I} + \boldsymbol{M}_{\boldsymbol{u}^{t+H-1}})\boldsymbol{A})^{H-1}\mathbf{z}_t\rangle \tag{34}$$
$$= \langle \boldsymbol{W}\widehat{S}_+ y, ((\boldsymbol{I} + \boldsymbol{M}_{\boldsymbol{u}^{t+H-1}})\boldsymbol{A})^{H-1}\mathbf{z}_t\rangle \tag{35}$$

which directly extends to vector-valued observables $\boldsymbol{y} \in (\mathcal{H}_X)^{n_y}$ as $\langle \boldsymbol{W}\boldsymbol{Y}_+, ((\boldsymbol{I} + \boldsymbol{M}_{\boldsymbol{u}^{t+H-1}})\boldsymbol{A})^{H-1}\mathbf{z}_t\rangle$, leading to (16) after plugging in (17). □

Embracing infinite-dimensions with RKHS: The above result show that, given a sequence of controls and the initial feature map $k_{\boldsymbol{x}_0}$, the prediction involves only matrix-vector multiplication. While the above predictor is finite-dimensional, it follows from an infinite-dimensional finite-rank control Koopman operator after meticulously applying algebra in RKHS. In turn, Proposition 1 eliminates auxillary regression problems, e.g., for state reconstruction, as the full-state is readily reconstructed by setting $\boldsymbol{y} = \mathrm{id}$ in the above derivation. To appreciate the gap the above result closes in the context of learning linear operators of nonlinear control systems, a the following remark is in order.

Remark 4 (Finite-rank ≠ finite-dimensional): In contrast to our results, existing kernel-based approaches for control systems operators [26], [58], [62] use *finite-dimensional* hypothesis spaces defined over data-based RKHS subspaces and sidestep infinite-dimensional regression [31], [49], [51], [60]. Technically, this amounts to learning a matrix between pre-RKHSs, incurring additional approximation errors w.r.t. truly RKHS-bound hypotheses.

VI. Efficient Approximations via Sketching

Like any other nonparametric approach, our cKOR algorithm is only suitable for small-scale systems due to the computational time-complexity of order $\mathcal{O}(n^3)$ w.r.t. the data cardinality n. For approximating kernel methods, random Fourier features (RFFs) stand out as a popular and straightforward way to reduce the time-complexity of estimation [63]. Recently, they have been utilized for control system identification [26]. Unfortunately, the algorithm in [26] is hardly useful in data-driven applications as it requires data gathered under constant inputs to estimate the Koopman operator for a few fixed input levels – prohibiting most realistic system identification scenarios that involve rich excitation signals or safe data collection, e.g., under an auxiliary controller. Moreover, RFFs being data-independent, they may not adapt well to the data at hand, limiting performance for an equivalent complexity as sketching schemes [64]. Random sketching or Nyström approximations estimate the kernel matrix by selecting a small subset of m data points known as inducing points or Nyström centres, that define a low-dimensional subspace of the RKHS the dataset is projected to [46], [47]. The Nyström approximation is accurate under the assumptions that an appropriate sampling is carried out and the kernel matrix has a low rank, where the latter is often satisfied, e.g., for Gaussian kernels whose Gram matrix eigenvalue spectrum rapidly decays [65]. We remark that, concurrent to our work, [56] proposes sketched operator estimation, but only for the restrictive case of LTI RKHS dynamics that is recovered as a special case of our approach (cf. Table II).

To put our developments in perspective, recall that the popular parametric bEDMD [12] or [15] have the time-complexity of $\mathcal{O}(n_{\mathbf{z}}^3(n_u+1)^3)$ – where $n_{\mathbf{z}}$ is the dictionary dimension – due to a cubic magnification of complexity in case of control systems based on the dimensionality of the inputs. This is primarily due to a lack of *kernel trick* for the control-affine effects – an inherent limitation of a parametric model. In contrast, the proposed combination of our nonparametric cKOR framework and random sketching that we will work out in this section preserves the "kernel trick" at a reduced set of samples of cardinality m to deliver a handy complexity reduction that is *independent of input-dimensionality*, in turn, amounting to the time-complexity $\mathcal{O}(m^3 + m^2 n)$, where $m \ll n$ is the number of inducing points – identical to the one known for autonomous systems [66].

We will consider uniform random sampling because it is a simple algorithm that is generally applicable and it is well-known that random projections are suitable for extracting a low rank matrix approximation [67]. There exist more advanced sampling approaches, which have the potential to minimize the number of inducing points necessary, but reviewing these is out of scope for this work. Still, our sketched scalable estimators are not limited to uniform random sampling. We note that the concept of Nyström approximation is rigorously studied in context of autonomous operators with KRR, PCR and RRR estimators, in [66]. In the remainder of this section, we extend these results to *control Koopman operator regression*. This approach starts by sampling a small subset of the data matrices with m datapoints/columns, where $m \ll n$ is the number of inducing points. With these datasets, the *subsampling operators* $\widetilde{S}_X, \widetilde{S}_+ : \mathcal{H}_X \to \mathbb{R}^m$ and $\widetilde{S}_Z : \mathcal{H}_Z \to \mathbb{R}^m$ are introduced to explicitly represent the orthogonal projections

$$\widetilde{P}_+ \triangleq \widetilde{S}_+^*(\widetilde{S}_+^*)^\dagger = \widetilde{S}_+^*\boldsymbol{K}_{\widetilde{+}}^\dagger\widetilde{S}_+ \triangleq \widetilde{E}_+\widetilde{E}_+^* : \mathcal{H}_X \to \mathcal{H}_X \tag{36}$$
$$\widetilde{P}_Z \triangleq \widetilde{S}_Z^*(\widetilde{S}_Z^*)^\dagger = \widetilde{S}_Z^*\boldsymbol{K}_{\widetilde{Z}}^\dagger\widetilde{S}_Z = \widetilde{E}_Z\widetilde{E}_Z^* : \mathcal{H}_Z \to \mathcal{H}_Z. \tag{37}$$

with kernel matrix shorthands $\boldsymbol{K}_{\widetilde{+}} \triangleq \widetilde{S}_+\widetilde{S}_+^* = [k(\boldsymbol{x}_+^{(i)}, \boldsymbol{x}_+^{(j)})]_{i,j\in[m]}$, $\boldsymbol{K}_{\widetilde{Z}} \triangleq \widetilde{S}_Z\widetilde{S}_Z^* = [k(\boldsymbol{z}^{(i)}, \boldsymbol{z}^{(j)})]_{i,j\in[m]}$ and $\widetilde{E}_{(\cdot)}$ partial isometries[2], satisfying $\widetilde{E}_{(\cdot)}^*\widetilde{E}_{(\cdot)} = \boldsymbol{I}$. With these projection operators, the original variational problem

[2]The input and output space partial isometries are $\widetilde{E}_+ = \widetilde{S}_+^*(\widetilde{S}_+\widetilde{S}_+^*)^{\dagger/2} \triangleq \widetilde{S}_+^*\boldsymbol{K}_{\widetilde{+}}^{\dagger/2}$, $\widetilde{E}_Z = \widetilde{S}_Z^*(\widetilde{S}_Z\widetilde{S}_Z^*)^{\dagger/2} \triangleq \widetilde{S}_Z^*\boldsymbol{K}_{\widetilde{Z}}^{\dagger/2}$, respectively.

(RRM) and ($\widehat{\text{RRM}}$) are projected following [66] to a lower-dimensional subspace, leading to *sketched* risk minimization

$$G_{m,\gamma} \triangleq \underset{G\in\mathrm{HS}(\mathcal{H}_X,\mathcal{H}_Z)}{\arg\min}\ \widetilde{R}(\widetilde{P}_Z G \widetilde{P}_+)+\gamma\|G\|_{\mathrm{HS}}^2 \qquad \text{(SRRM)}$$

and its empirical risk counterpart

$$\begin{aligned}\widehat{G}_{m,\gamma} &\triangleq \underset{G\in\mathrm{HS}(\mathcal{H}_X,\mathcal{H}_Z)}{\arg\min}\ \widehat{R}(\widetilde{P}_Z G \widetilde{P}_+)+\gamma\|G\|_{\mathrm{HS}}^2 \\ &= (\widetilde{P}_Z \widehat{S}_Z^* \widehat{S}_Z \widetilde{P}_Z + n\gamma I)^{-1} \widetilde{P}_Z \widehat{S}_Z^* \widehat{S}_+ \widetilde{P}_+. \qquad (\widehat{\text{SRRM}})\end{aligned}$$

Though the above expression may not be directly computable, we will arrive at a finite-dimensional form for them after meticulously applying algebra. Let us first introduce the *input subsampling operator* $\widetilde{U} : \boldsymbol{u} \mapsto [\langle \boldsymbol{u}^{(1)},\boldsymbol{u}\rangle \cdots \langle \boldsymbol{u}^{(m)},\boldsymbol{u}\rangle)]^\intercal$ and kernel matrices $\boldsymbol{K}_{\widetilde{X}} \triangleq \widetilde{S}_X \widetilde{S}_X^* = [k(\boldsymbol{x}^{(i)},\boldsymbol{x}^{(j)})]_{i,j\in[m]}$, $\boldsymbol{K}_{\widetilde{Z}} \triangleq \boldsymbol{K}_{\widetilde{X}} + (\widetilde{U}\widetilde{U}^*)\odot \boldsymbol{K}_{\widetilde{X}}$, $\boldsymbol{K}_{\widetilde{+}} \triangleq \widetilde{S}_+ \widetilde{S}_+^* = [k(\boldsymbol{x}_+^{(i)},\boldsymbol{x}_+^{(j)})]_{i,j\in[m]}$, $\boldsymbol{K}_{Z\widetilde{Z}} \triangleq \boldsymbol{K}_{X\widetilde{X}} + (\widehat{U}\widetilde{U}^*)\odot \boldsymbol{K}_{X\widetilde{X}}$ with $\boldsymbol{K}_{X\widetilde{X}} \triangleq [k(\boldsymbol{x}^{(i)},\boldsymbol{x}^{(j)})]_{i\in[n],j\in[m]}$ and $\boldsymbol{K}_{+\widetilde{+}} \triangleq [k(\boldsymbol{x}_+^{(i)},\boldsymbol{x}_+^{(j)})]_{i\in[n],j\in[m]}$.

Proposition 2: The minimizer of ($\widehat{\text{SRRM}}$) equals to

$$\widehat{G}_{m,\gamma} \triangleq \widetilde{S}_Z \boldsymbol{W} \widetilde{S}_+ = \begin{bmatrix} \widehat{A}_{m,\gamma} \\ \widehat{B}_{m,\gamma} \end{bmatrix} \in \mathrm{HS}(\mathcal{H}_X,\mathcal{H}_Z)$$

where $\boldsymbol{W} \triangleq (\boldsymbol{K}_{Z\widetilde{Z}}^\intercal \boldsymbol{K}_{Z\widetilde{Z}} + n\gamma \boldsymbol{K}_{\widetilde{Z}})^\dagger \boldsymbol{K}_{Z\widetilde{Z}}^\intercal \boldsymbol{K}_{+\widetilde{+}} \boldsymbol{K}_{\widetilde{+}}^\dagger \in \mathbb{R}^{m\times m}$ and the $(k \geq 1)$-step evolution of any observable $\boldsymbol{y} \in \mathcal{H}^{n_y}$ under control inputs $\boldsymbol{u}_{k\in\mathbb{N}_0} \in \mathbb{U}$ are exactly represented by the following predictor

$$\mathbf{z}_1 = (\boldsymbol{I} + \boldsymbol{M}_{\boldsymbol{u}_0})\boldsymbol{k}_{\widetilde{\boldsymbol{X}}}(\boldsymbol{x}_0), \tag{38a}$$

$$\mathbf{z}_{k+1} = (\boldsymbol{A} + \textstyle\sum_{i=1}^{n_u} \langle \boldsymbol{e}_i,\boldsymbol{u}_k\rangle \boldsymbol{B}_i)\mathbf{z}_k, \tag{38b}$$

$$\widehat{\boldsymbol{y}}_k = \boldsymbol{C}\mathbf{z}_k, \tag{38c}$$

where $\boldsymbol{k}_{\widetilde{\boldsymbol{X}}}(\boldsymbol{x}) = [k(\boldsymbol{x},\boldsymbol{x}^{(1)}) \cdots k(\boldsymbol{x},\boldsymbol{x}^{(m)})]^\intercal$, $\boldsymbol{M}_{(\cdot)} \triangleq \mathrm{diag}(\widetilde{U}\cdot)$, $\widetilde{\boldsymbol{Y}}_+^\intercal \triangleq [\boldsymbol{y}(\boldsymbol{x}_+^{(1)}) \cdots \boldsymbol{y}(\boldsymbol{x}_+^{(m)})]$, $\boldsymbol{A} \triangleq (\boldsymbol{W}\boldsymbol{K}_{\widetilde{+}})^\intercal$, $\boldsymbol{B}_i \triangleq \boldsymbol{M}_{\boldsymbol{e}_i}\boldsymbol{A}$, $\boldsymbol{C} \triangleq (\boldsymbol{W}\widetilde{\boldsymbol{Y}}_+)^\intercal$.

Proof: Writing out the estimator ($\widehat{\text{SRRM}}$) gives

$$(\widetilde{P}_Z \widehat{S}_Z^* \widehat{S}_Z \widetilde{P}_Z + n\gamma I)^{-1} \widetilde{P}_Z \widehat{S}_Z^* \widehat{S}_+ \widetilde{P}_+ \tag{39a}$$

$$= \widetilde{E}_Z (\boldsymbol{K}_{\widetilde{Z}}^{\dagger/2} \boldsymbol{K}_{Z\widetilde{Z}}^\intercal \boldsymbol{K}_{Z\widetilde{Z}} \boldsymbol{K}_{\widetilde{Z}}^{\dagger/2} + n\gamma I)^{-1} \widetilde{E}_Z^* \widehat{S}_Z^* \widehat{S}_+ \widetilde{E}_+ \widetilde{E}_+^* \tag{39b}$$

$$= \widetilde{S}_Z^* (\boldsymbol{K}_{Z\widetilde{Z}}^\intercal \boldsymbol{K}_{Z\widetilde{Z}} + n\gamma \boldsymbol{K}_{\widetilde{Z}})^\dagger \boldsymbol{K}_{Z\widetilde{Z}}^\intercal \boldsymbol{K}_{+\widetilde{+}} \boldsymbol{K}_{\widetilde{+}}^\dagger \widetilde{S}_+ \tag{39c}$$

where we used $P_Z = \widetilde{E}_Z \widetilde{E}_Z^*$ and the push-through identity together with $\widetilde{E}_Z^* \widetilde{E}_Z = \boldsymbol{I}$, By following the proof of [66, Proposition C.2] we have $\boldsymbol{W} \triangleq (\boldsymbol{K}_{Z\widetilde{Z}}^\intercal \boldsymbol{K}_{Z\widetilde{Z}} + n\gamma \boldsymbol{K}_{\widetilde{Z}})^\dagger \boldsymbol{K}_{Z\widetilde{Z}}^\intercal \boldsymbol{K}_{+\widetilde{+}} \boldsymbol{K}_{\widetilde{+}}^\dagger$. Finally, we arrive at (38) following the proof of Proposition 1. □

The inverses in Ny-cKOR only take $\mathcal{O}(m^3)$, which is a significant improvement compared to $\mathcal{O}(n^3)$, since $m \ll n$.

VII. Model Order Reduction

For cKOR and Ny-cKOR, the lifted state dimension scales with the dataset cardinality n and inducing points m, respectively. For relatively few inducing points, the lifted state dimension can still be high—posing challenges for efficient control design and real-time execution on low-level hardware.

A compelling approach for ordering and reducing the lifted states is based on proper orthogonal mode decomposition (POD), because it has been successfully applied in obtaining low-dimensional representations based on large-scale datasets in many applications [68]. The dynamic mode decomposition (DMD) algorithm actually makes use of this reduction [69] where it involves taking an SVD of the state "data matrix" – analogous to the SVD of the empirical embedding operator $\widehat{S}_X^*$ in our RKHS setting, which ranks the orthogonal structures of this matrix based on the singular values. With this ranking, the r dominant modes/coordinates can be selected to describe the dynamical behavior of the underlying system.

The aforementioned reduction approach in the context of cKOR (and Ny-cKOR), starts by taking the r-truncated SVD of the kernel matrix $[\![\boldsymbol{K}]\!]_r = \boldsymbol{V}_r \boldsymbol{\Sigma}_r \boldsymbol{V}_r^\intercal$, with the POD modes $\boldsymbol{V}_r \in \mathbb{R}^{N_{\boldsymbol{z}}\times r}$ and $\boldsymbol{\Sigma}_r = \mathrm{diag}(\sigma_1,\ldots,\sigma_r)$.[3] With these POD modes, the original bilinear lifted system can be transformed into the following reduced bilinear lifted form

$$\mathbf{z}_1 = \boldsymbol{V}_r^\intercal (\boldsymbol{I} + \boldsymbol{M}_{\boldsymbol{u}_0})\boldsymbol{k}_{\boldsymbol{X}}(\boldsymbol{x}_0), \tag{40a}$$

$$\mathbf{z}_{k+1} = (\boldsymbol{V}_r^\intercal \boldsymbol{A} \boldsymbol{V}_r + \textstyle\sum_{i=1}^{n_u} \langle \boldsymbol{e}_i,\boldsymbol{u}_k\rangle \boldsymbol{V}_r^\intercal \boldsymbol{B}_i \boldsymbol{V}_r)\mathbf{z}_k, \tag{40b}$$

$$\widehat{\boldsymbol{y}}_k = \boldsymbol{C}\boldsymbol{V}_r\mathbf{z}_k, \tag{40c}$$

In context of cKOR, this method is coined as reduced cKOR (r-cKOR). For Ny-cKOR, the reduction approach is identical to cKOR with the difference that the truncated SVD is applied to the kernel matrix $[\![\widetilde{\boldsymbol{K}}]\!]_r$. In this case, the method is coined as reduced Ny-cKOR (r-Ny-cKOR).

VIII. Numerical Examples

In this section, we present numerical studies to illustrate the implications of the theoretical results and showcase the advantages of our cKOR approach in practice. For comparison, the bilinear EDMD baseline [21], [22] is used with the same data- or subsample-based dictionary $\mathrm{span}\{k(\boldsymbol{x}^{(1)},\boldsymbol{x}),\ldots,k(\boldsymbol{x}^{(n)},\boldsymbol{x})\}$.

A. Model learning for the Duffing oscillator

As the first example, a controlled damped Duffing oscillator described by the state-space representation

$$\dot{\boldsymbol{x}} = \begin{bmatrix} x_2 \\ x_1 - x_1^3 - 0.5x_2 \end{bmatrix} + \begin{bmatrix} 0 \\ 2 + \sin(x_1) \end{bmatrix} u, \tag{41}$$

is used, where the state is simulated using RK4 integrator measured at sampling time $T_\mathrm{s} = 0.01s$ while the input is actuated in a synchronized zero-order-hold (ZOH) manner.

Prediction performance w.r.t. hyperparameters: First we compare the cKOR, Ny-cKOR and bEDMDc approaches over a range of the hyperparameters $\mu \in \mathbb{R}_+$ of the Gaussian (RBF) kernel $k(\boldsymbol{x},\boldsymbol{x}') = \mathrm{e}^{-\|\boldsymbol{x}-\boldsymbol{x}'\|_2^2/\mu}$ in terms of the T_test-ahead prediction performance, quantified using the root mean square error (RMSE) $(1/T_\mathrm{test} \sum_{t=1}^{T_\mathrm{test}} \|\boldsymbol{y}_t - \widehat{\boldsymbol{y}}_t\|_2^2)^{1/2}$, where T_test denotes the number of steps and $\boldsymbol{y}_t - \widehat{\boldsymbol{y}}_t$ is the difference between the true system response and the predicted solution. For the training dataset, 200 trajectories with $n = 1000$ samples (10 sec) are generated, starting from a 14 by 14 grid of initial conditions within the limits $|x_1| \leq 2.25$ and $|x_2| \leq 2.25$. For

[3] We select r thorough a threshold τ such that $\sum_{i=1}^r \sigma_i^2 / \sum_{j=1}^n \sigma_j^2 \cdot 100\% \leq \tau$, but there are various other methods for choosing r, see [70].

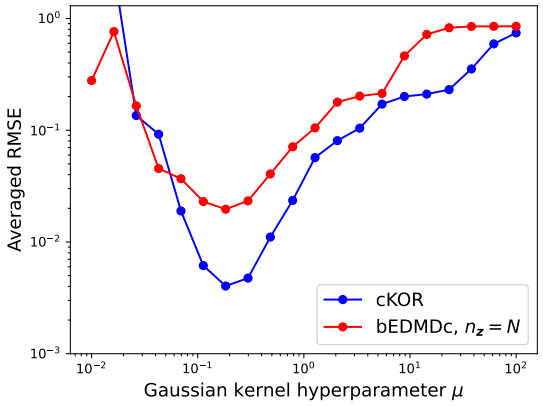


Fig. 3. Averaged RMSE of the 1-step-ahead prediction for the cKOR, Ny-cKOR and bEDMDc models over the test set for various choices of the kernel width μ for the case of predictor dimension $n_z = 1000$.

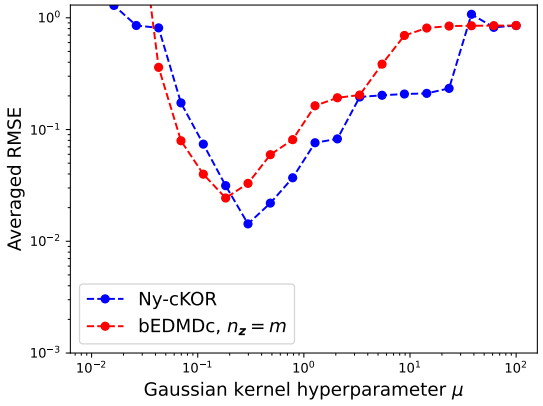


Fig. 4. Averaged RMSE of the 1-step-ahead prediction for the cKOR, Ny-cKOR and bEDMDc models over the test set for various choices of the kernel width μ for the case of predictor dimension $n_z = 200$.

the test dataset, 40 trajectories of $T_{\text{test}} = 100$ samples $(1s)$ are generated, starting from random initial conditions sampled within the limits $|x_1| \leq 2$ and $|x_2| \leq 2$ using a uniform distribution. Both datasets are generated using uniform random input sequences within the interval $[-2, 2]$. From the training dataset, $m = 200$ inducing points are randomly sampled. The considered approaches are used with a regularization parameter $\gamma = 10^{-9}$. Figure 3 confirms the superior accuracy of our nonparametric cKOR estimator as it reaches a significantly lower error then bEDMDc across μ values – showing a greater hyperparameter range of increased accuracy. Additionally, when lowering the regularization, the accuracy of our cKOR estimate increases, achieving up to an *order of magnitude better accuracy than bEDMDc*, cf. Appendix B. Even when we reduce complexity by projecting on a subset of data for our sketched Ny-cKOR estimator, a similar, but slightly reduced advantage can be observed in Figure 4.

Statistical performance & time-complexity evaluation: Next, the training data is varied in terms of the number of samples. To generate the data snapshots, the initial conditions

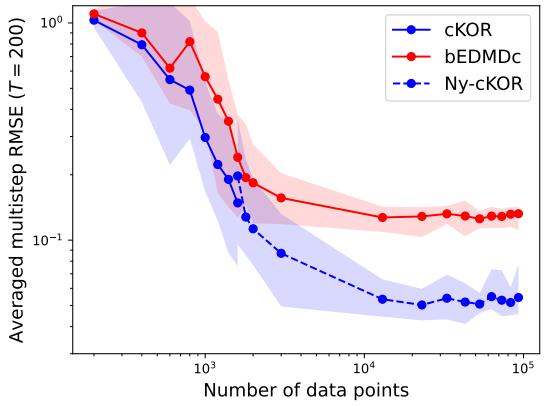


Fig. 5. RMSE of the 200-step-ahead prediction RMSE for the cKOR, Ny-cKOR and bEDMDc models w.r.t. increasing training data size. Our (Ny)-cKOR estimators attain significantly lower errors.

are sampled from a square and equidistant grid within the limits $|x_1|, |x_2| \leq 2$ and the system is driven for $2.0s$ by uniformly at random generated control inputs within the limits $|u| \leq 2$. The test data consists of 20 trajectories of length 2.0 sec ($T_{\text{test}} = 200$) with the same initial condition generation, but driven by an input sequence of $u_t = 2\sin(10\pi t)$. All the approaches use the Gaussian kernel to construct the lifted states with hyperparameter $\mu = 0.25$ and a regularization parameter of 10^{-7}. These values are empirically determined as "optimal" for the prediction RMSE and choosing the same settings allows for a fair comparison between cKOR, Ny-cKOR and bEDMDc. Note that, the bEDMDc approach takes the inducing points as centers, which are 200 uniformly randomly sampled points from the training dataset. Figure 5 illustrates the RMSE of the T_{test}-step-ahead prediction averaged over the test trajectories versus the number of training datapoints. The solid lines represent the average RMSE over 20 runs and the shaded area gives the variation of the RMSE per run. For each run, a new training data set and inducing points are generated to provide statistically relevant results. Figure 6 shows the computation times, i.e., the estimation time of the predictor along with the n-step-ahead prediction/rollout time. Strikingly, both the average RMSE of cKOR and Ny-cKOR stays below of bEDMDc, confirming the inherent advantages of estimators derived using a nonparametric paradigm. This also illustrates the common bottleneck of full KRR estimators well over the number of datapoints, since the full cKOR scales with $\mathcal{O}(n^3)$, compared to bEDMDc with $\mathcal{O}((n_z(n_u+1))^3)$ and Ny-cKOR $\mathcal{O}(m^3)$. In line with our expectation, Ny-cKOR continues the trend of cKOR for larger datasets, as the computation time becomes intractable for the full cKOR estimator. For a single input, the bEDMDc complexity is comparable to Ny-cKOR, however it is important to stress that Ny-cKOR is substantially more computationally efficient than bEDMDc for higher input dimensions – by a factor of $(n_u+1)^3$ – as it does not require taking a tensor product of features and inputs.

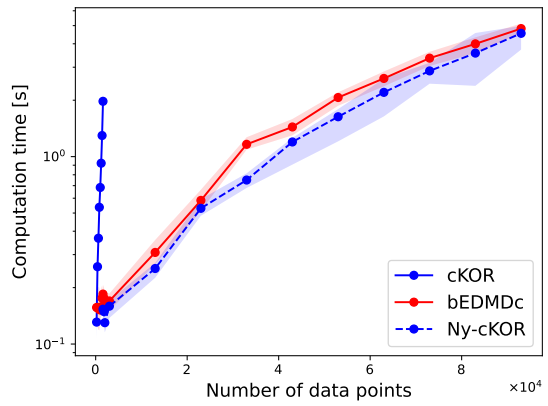


Fig. 6. Computation time for the cKOR, Ny-cKOR and bEDMDc models over increasing training data size. Our Ny-cKOR estimator effectively attains the complexity of the parametric approach.

B. Learning the high-dimensional Kalman vortex street

Tackling high-dimensional systems in a parametric manner, often comes with challenges as the suitable basis for representation is of critical importance. Through this simulation study, we want to showcase the superiority of our nonparametric learning paradigm even when it is based on a fraction of the data points, which is highly important for scalability. Figure 8 schematically illustrates the considered nonlinear system which generates a flow as a result of transverse non-slip movement of an oscillating cylinder as input. This flow exhibits vortex shedding causing vortex-induced vibrations on the structure, which accelerate material fatigue and may lead to failure [71]. In [72], the considered system is created and simulated using the Computational Fluid Dynamics (CFD) environment OpenFOAM. For the data generation, we refer to [72]. The system identification procedure is performed with the same setting as in [73].

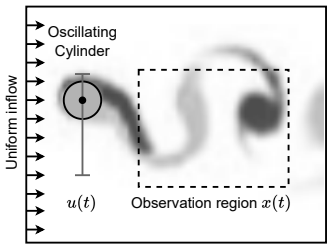


Fig. 8. An oscillating cylinder in a uniform flow.

For completeness, these settings are repeated here: the pressure, horizontal velocity and vertical velocity are observed in the rectangular wake region of 41×45, which amounts to a high state dimension of $n_x = 5535$. The flow conditions correspond to a Reynolds number of 100 and a Strouhal number of 0.167. The dataset consists of 11 timeseries of length $T = 1520$, with samples measured at 50Hz. These 11 timeseries correspond to 11 input sequences with a swept-sine input profile with different amplitudes and cylinder diameter ratios. The timeseries are randomly split into: 6 series for training, 3 for validation, and 2 for testing. For statistically significant comparison, we assess the prediction performance of the methods on the test data over 20 randomly assigned splits. The training, validation and test datasets are normalized to constrain the states and inputs to values ≤ 1. For a fair comparison, all approaches use the same 400 inducing points for learning, meaning bEDMDc uses 400 RBF centres based on the inducing points of Ny-cKOR (and r-Ny-cKOR). Also all models are fitted using a hyperparameter and regularizer grid search on the validation data with the grids $\mu \in \{0.1, 0.5, 1, 10, 20, \dots 60, 150, 175, \dots 400\}$ and $\gamma \in \{10^{-11}, 10^{-10}, \dots, 10^{-6}\}$ for the multi-step ($T_{\text{valid}} = 100$) state prediction RMSE. The rank r of the POD reduction is obtained for $\tau = 99.99\%$ (cf. Section VII). As shown in Table III, on average, both models from r-Ny-cKOR and Ny-cKOR significantly outperform those of bEDMDc. Strikingly, *the error variance* compared to bEDMDc *for our Ny-cKOR and r-Ny-cKOR models is* $20\times$ *and* $10\times$ *smaller*, respectively. Figure 7 shows an examples of the significant performance loss due to the large error variance of bEDMDc models. By just comparing the flow plots, it becomes clear that bEDMDc quickly deteriorates and does not resemble any aspect of the flow, as opposed to Ny-cKOR, which stays quite accurate over the entire horizon. The reduced model of r-Ny-cKOR comes with an offset to Ny-cKOR, but does not exhibit the performance loss of bEDMDc. This example demonstrates the superiority of the nonparametric paradigm, through the significantly better prediction accuracy of our (r-)Ny-cKOR models for an unknown high-dimensional control system.

TABLE III
TEST NORMALIZED RMSE (NRMSE) FOR THE ESTIMATED MODELS ON THE ACTUATED KARMAN VORTEX STREET FOR MULTI-STEP PREDICTION OVER $T_{\text{TEST}} = 3T_{\text{VALID}}$ FROM 20 TEST-VALIDATION-TRAIN SPLITS.

	Ny-cKOR	**r-Ny-cKOR**	bEDMDc
NRMSE	**0.157±0.049**	**0.206±0.109**	0.412±0.976

C. Model predictive control with cKOR predictors

Here we integrate our control operator predictors from (Proposition 1 and 2) in an *iterated* LPV-MPC scheme [74], that we call *control Koopman operator LPV-MPC* (cKOLPV-MPC), whose description is delegated to the Appendix C. In a nutshell, we solve a single QP at every time-instant where the cKOLPV-MPC updates the scheduling iteratively over the simulation time in a receding horizon manner. This may cause some loss of performance, but convergence is still observed in practice, similar to SQP schemes.

Damped Duffing oscillator: First we consider the MPC design for the Duffing oscillator (41). The autonomous dynamics have two stable equilibrium points while the origin is unstable. The vector field of the system is plotted in Figure 9 to illustrate if the control trajectories leverage the dynamics to gain performance. In this example, the cKOLPV-MPC (cf. Appendix C), the sequential linearization approach based we name LPV-MPC[4], and a linear MPC (LMPC) based on a linearized model of (41) are compared. The nonlinear MPC

[4]This scheme is equivalent to applying the *LTV-MPC* from [75] in a non-robustifed ("tube-free") fashion.

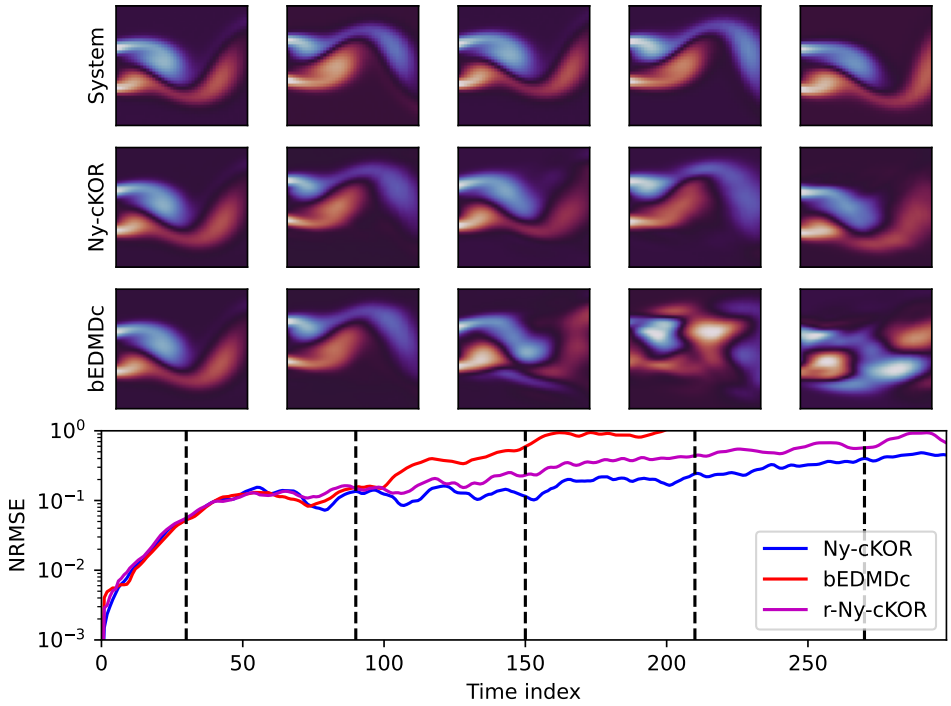


Fig. 7. A prediction of the flow in the Kalman Vortex Street example, showing the reduced prediction error of our higher fidelity Ny-cKOR models.

(NMPC) with the exact nonlinear model of (41) is considered as *the ground truth*.

The initial condition of the simulated scenario is set to $[1\ 1]^\intercal$ and the state reference consists of the two equilibra: $[-1, 0]^\intercal$ and $[0, 0]^\intercal$, for $9s$ and $3s$, respectively. The weighting matrices are chosen as $\boldsymbol{Q} = \text{diag}(6, 1)$, $R = 5$ and $\boldsymbol{Q}_T = 100\boldsymbol{Q}$. The horizon is set to 100 steps, i.e. $T = 100$ with a sampling time of $T_\text{s} = 0.01s$. Lastly, the constraints are sets as $-2 \leq u \leq 2$, $-3 \leq x_1 \leq 3$ and $-3 \leq x_2 \leq 3$. These settings are such chosen such that knowing nonlinear system behavior accurately is rewarded and that the desired setpoints are reached.

The linear MPC (LMPC) model is obtained by linearizing the system equations around the origin. LPV-MPC linearizes the system along the prior predicted states and optimized inputs where its initial guess is the initial condition over the horizon and zero inputs. The same initial guess is used for NMPC, which also has access to the system equations. Our cKoLPV-MPC does not have access to the system equations, instead, it uses training data ($n = 704$) in a state domain of $-2 \leq x_1 \leq 2$ and $-2 \leq x_2 \leq 2$ and an input domain of $[-2, 2]$. The bilinear model is constructed via r-Ny-cKOR with 100 uniformly randomly sampled inducing points, $r = 29$ and a hyperparameter grid search over validation data.

From Figures 9 and 10, it can be observed that cKoLPV-MPC, LPV-MPC and NMPC are almost identical, while they showcase a clear performance improvement compared to LMPC. Specifically, LMPC requires extra input effort between $1s$ and $3s$, because around $x_1 = -0.2$ and $x_2 = -1.25$, the LMPC solution goes against the vector field. The other approaches use the vector field to reach the setpoint and thus requiring less input. In addition, there is an large offset

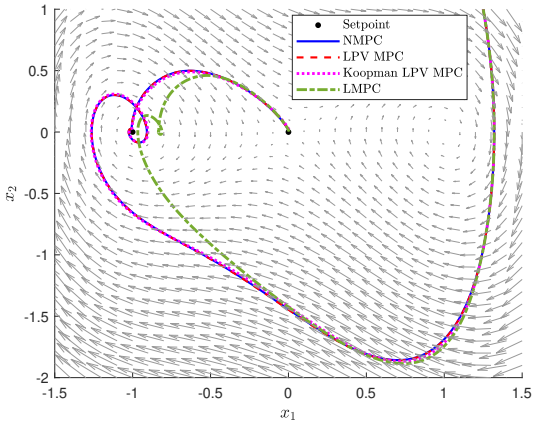


Fig. 9. State-space response of the simulated Duffing oscillator under piecewise constant reference tracking for the MPC.

between the settled state of LMPC and the setpoint, which then leads to an additional input effort to stabilize in the origin. The other approaches are almost identical, which implies that: 1) the system and/or control task is not challenging enough and that; 2) the bilinear Koopman model accurately identifies the nonlinear system.

Unstable system with linearly uncontrollable origin: For the next example, the following Van der Pol oscillator

$$\dot{\boldsymbol{x}} = \begin{bmatrix} x_2 \\ -x_1 - \frac{1}{2}x_2(1 - x_1^2) \end{bmatrix} + \begin{bmatrix} 0 \\ x_1 u \end{bmatrix}, \tag{42}$$

where the state is measured with sampling time $T_\text{s} = 0.05s$ and the input is actuated in a synchronized ZOH manner.

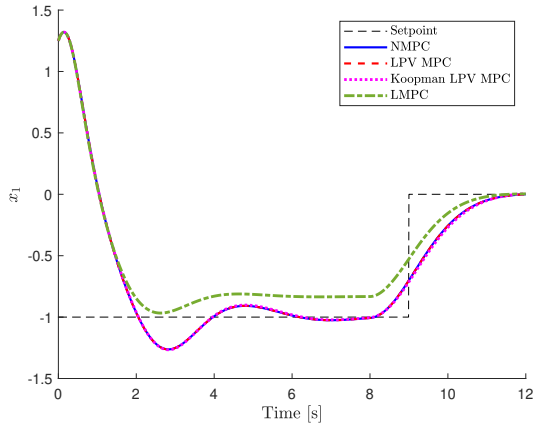


Fig. 10. Comparing position responses of LMPC, LPV-MPC, cKoLPV-MPC and NMPC for piecewise constant references on the simulated Duffing oscillator.

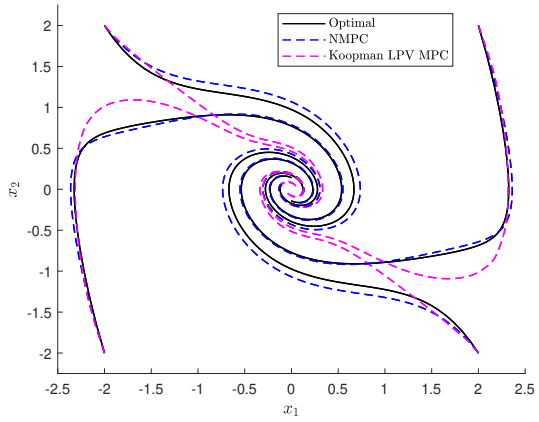


Fig. 11. State-space responses of the simulated Van der Pol oscillator for the cKoLPV-MPC vs NMPC.

This is an interesting example for three reasons: the origin is linearly uncontrollable, the optimal solution to drive the system from arbitrary initial condition to the origin is known, i.e., the infinite-horizon optimal controller for the cost function $J(\boldsymbol{x}, u) = x_2^2 + u^2$ is $u = -x_1 x_2$ [76]. Thus, the MPC control task is to minimize the aforementioned cost function over a finite-horizon. In other words, this system can showcase the potential of nonlinear control techniques as opposed to controllers based on linear state-space models. We evaluate the performance on, the initial conditions $[-2, -2]^\intercal$, $[-2, 2]^\intercal$, $[2, -2]^\intercal$ and $[2, 2]^\intercal$ with $\boldsymbol{Q} = \text{diag}(0, 1)$, $\boldsymbol{Q}_T = \boldsymbol{Q}$ and $R = 1$ to match the above cost function. The horizon is chosen as the minimal one such that NMPC stabilizes the origin. This resulted in a horizon of 100 steps, which is relatively big and thus another indicator of a difficult to control system. The latter in combination with being open-loop unstable in the considered region, complicates the data-gathering step for learning. For the training and the validation data, the NMPC controller is used to control the system to the origin with an exploratory uniform random disturbance within the interval $[-2, 2]$. The hyperparameter and regularization parameter are obtained by employing a grid search on the validation data. The same initialization of the scheduling is used as in the previous example. Figure 11 shows the resulting control trajectories for NMPC and cKoLPV-MPC. For these settings, the LPV-MPC and LMPC controllers *fail to stabilize the origin, due to linearization limitations and the linearly uncontrollable property*. The latter clearly highlights an advantage for the cKoLPV-MPC scheme. However, the control trajectories of cKoLPV-MPC deviate from the trajectories of the NMPC with full exact model knowledge. The aforementioned deviations are quantified with RMSE: RMSE of cKoLPV-MPC is $2.83 \cdot 10^{-1}$ and of NMPC: $1.01 \cdot 10^{-1}$. Due to its data-driven nature, the model is inherently approximative, while the NMPC works with perfect system knowledge. The latter is illustrated in Figure 12 by requiring a higher input as opposed to following the vector field. Note that cKoLPV-MPC *solves a single QP at every timestep and does not require any initial guess for the scheduling or employment of model-based planners*.

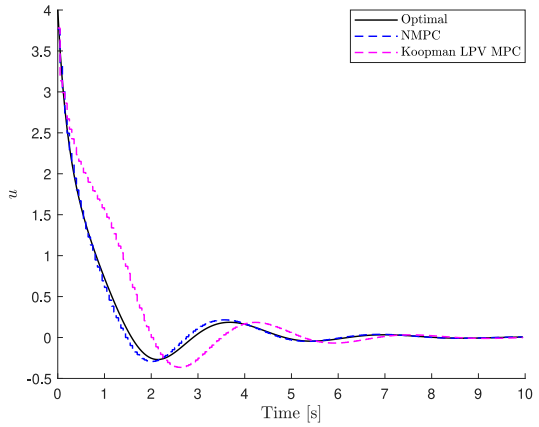


Fig. 12. Comparing cKoLPV-MPC and NMPC control input signals form the initial condition $\boldsymbol{x}_0 = [2, -2]^\intercal$ for the simulated Van der Pol oscillator.

IX. Conclusion

We introduce a novel framework for learning Koopman operators for control-affine systems in reproducing kernel Hilbert spaces (RKHS), grounded in risk minimization and infinite-dimensional regression. By establishing the equivalence of various operator formulations, we enable the simultaneous use of vector-valued regression for learning and LPV Koopman forms for prediction and control. This equivalence demonstrates that control operators can be fully described using scalar-valued kernels, bridging a critical gap in existing operator representations and fully leveraging the available RKHS structure. Our proposed empirical estimators based on data samples form finite-rank operators over RKHSs that are independent of feature and input dimensions, and form finite-dimensional predictors without any loss of precision. Furthermore, we prove that our approach allows arbitrarily accurate operator norm approximations under minimal assumptions using finite-rank operators. Additionally, we propose sketched estimators to improve the scalability of our method by reducing computational complexity to $\mathcal{O}(m^2)$ for large-scale problems, where m may be much smaller than the data cardinality. As

implied by our theoretical analysis, the numerical experiments demonstrate superior prediction accuracy compared to bilinear EDMD, especially for high-dimensional systems. Finally, our learned models integrate seamlessly with LPV techniques for model predictive control, offering a viable alternative to standard MPC approaches.

ACKNOWLEDGMENTS

The authors thank Nicolas Hoischen and Max Beier for their feedback on the manuscript and Jan Decuyper for sharing the details of the Karman vortex example.

REFERENCES

[1] K.-J. Engel and R. Nagel, *One-Parameter Semigroups for Linear Evolution Equations*. Springer-Verlag, 2000, vol. 194.

[2] P. Bevanda, S. Sosnowski, and S. Hirche, "Koopman operator dynamical models: Learning, analysis and control," *Annual Reviews in Control*, vol. 52, pp. 197–212, 2021.

[3] S. L. Brunton, M. Budišić, E. Kaiser, and J. N. Kutz, "Modern koopman theory for dynamical systems," *SIAM Review*, vol. 64, no. 2, pp. 229–340, 2022.

[4] S. E. Otto and C. W. Rowley, "Koopman operators for estimation and control of dynamical systems," *Annual Review of Control, Robotics, and Autonomous Systems*, vol. 4, pp. 59–87, 5 2021.

[5] M. Budišić, R. Mohr, and I. Mezić, "Applied koopmanism," *Chaos*, vol. 22, 10 2012.

[6] I. Mezić and A. Banaszuk, "Comparison of systems with complex behavior," *Physica D: Nonlinear Phenomena*, vol. 197, no. 1, pp. 101–133, 2004.

[7] I. Mezic, "Spectral properties of dynamical systems, model reduction and decompositions," *Nonlinear Dynamics*, vol. 41, pp. 309–325, 08 2005.

[8] B. O. Koopman, "Hamiltonian systems and transformation in hilbert space." *Proceedings of the National Academy of Sciences of the United States of America*, vol. 17 5, pp. 315–8, 1931.

[9] Q. Li, F. Dietrich, E. M. Bollt, and I. G. Kevrekidis, "Extended dynamic mode decomposition with dictionary learning: A data-driven adaptive spectral decomposition of the koopman operator," *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 27, p. 103111, 2017.

[10] S. E. Otto and C. W. Rowley, "Linearly recurrent autoencoder networks for learning dynamics," *SIAM Journal on Applied Dynamical Systems*, vol. 18, pp. 558–593, 1 2019.

[11] O. Azencot, N. B. Erichson, V. Lin, and M. Mahoney, "Forecasting sequential data using consistent koopman autoencoders," H. D. III and A. Singh, Eds., vol. 119. PMLR, 7 2020, pp. 475–485.

[12] D. Bruder, X. Fu, R. B. Gillespie, C. D. Remy, and R. Vasudevan, "Data-driven control of soft robots using koopman operator theory," *IEEE Transactions on Robotics*, vol. 37, pp. 948–961, 6 2021.

[13] D. A. Haggerty, M. J. Banks, E. Kamenar, A. B. Cao, P. C. Curtis, I. Mezić, and E. W. Hawkes, "Control of soft robots with inertial dynamics," *Science Robotics*, vol. 8, 8 2023.

[14] M. E. Villanueva, C. N. Jones, and B. Houska, "Towards global optimal control via koopman lifts," *Automatica*, vol. 132, p. 109610, 10 2021.

[15] Y. Guo, B. Houska, and M. E. Villanueva, "A tutorial on pontryagin-koopman operators for infinite horizon optimal control," in *2022 IEEE 61st Conference on Decision and Control (CDC)*, 2022, pp. 6800–6805.

[16] B. Houska, "Convex operator-theoretic methods in stochastic control," 5 2023. [Online]. Available: http://arxiv.org/abs/2305.17628

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

[18] S. Peitz and S. Klus, "Koopman operator-based model reduction for switched-system control of pdes," *Automatica*, vol. 106, pp. 184–191, 2019.

[19] D. Goswami and D. A. Paley, "Global bilinearization and controllability of control-affine nonlinear systems: A koopman spectral approach," in *2017 IEEE 56th Annual Conference on Decision and Control (CDC)*, 2017, pp. 6107–6112.

[20] L. C. Iacob, R. Tóth, and M. Schoukens, "Koopman form of nonlinear systems with inputs," *Automatica*, vol. 162, p. 111525, 2024.

[21] D. Bruder, X. Fu, and R. Vasudevan, "Advantages of bilinear koopman realizations for the modeling and control of systems with unknown dynamics," *IEEE Robotics and Automation Letters*, vol. PP, pp. 1–1, 03 2021.

[22] S. E. Otto, S. Peitz, and C. W. Rowley, "Learning bilinear models of actuated koopman generators from partially-observed trajectories," 2023.

[23] S. Peitz, S. Otto, and C. Rowley, "Data-driven model predictive control using interpolated koopman generators," *SIAM Journal on Applied Dynamical Systems*, vol. 19, p. 2162–2193, 09 2020.

[24] M. Schaller, K. Worthmann, F. Philipp, S. Peitz, and F. Nüske, *IFAC-PapersOnLine*, vol. 56, no. 1, pp. 169–174, 2023, 12th IFAC Symposium on Nonlinear Control Systems NOLCOS 2022.

[25] F. Nüske, S. Peitz, F. Philipp, M. Schaller, and K. Worthmann, "Finite-data error bounds for koopman-based prediction and control," *Journal of Nonlinear Science*, vol. 33, p. 14, 2 2023.

[26] F. Philipp, M. Schaller, K. Worthmann, S. Peitz, and F. Nüske, "Error analysis of kernel edmd for prediction and control in the koopman framework," 12 2023. [Online]. Available: http://arxiv.org/abs/2312.10460

[27] Y. Guo, M. Korda, I. G. Kevrekidis, and Q. Li, "Learning parametric koopman decompositions for prediction and control," 10 2023. [Online]. Available: http://arxiv.org/abs/2310.01124

[28] M. Haseli and J. Cortés, "Modeling nonlinear control systems via koopman control family: Universal forms and subspace invariance proximity," 7 2023. [Online]. Available: http://arxiv.org/abs/2307.15368

[29] P. Bevanda, M. Beier, S. Heshmati-Alamdari, S. Sosnowski, and S. Hirche, "Towards data-driven lqr with koopmanizing flows," *IFAC-PapersOnLine*, vol. 55, pp. 13–18, 2022.

[30] P. Batlle, M. Darcy, B. Hosseini, and H. Owhadi, "Kernel methods are competitive for operator learning," *Journal of Computational Physics*, vol. 496, p. 112549, 2024.

[31] V. Kostic, P. Novelli, A. Maurer, C. Ciliberto, L. Rosasco, and M. Pontil, "Learning dynamical systems via koopman operator regression in reproducing kernel hilbert spaces," in *Advances in Neural Information Processing Systems*, vol. 35, 2022, pp. 4017–4031.

[32] V. Kostic, K. Lounici, P. Novelli, and M. Pontil, "Sharp spectral rates for koopman operator learning," in *Advances in Neural Information Processing Systems*, vol. 36, 2023, pp. 32 328–32 339.

[33] P. Bevanda, M. Beier, A. Lederer, S. Sosnowski, E. Hüllermeier, and S. Hirche, "Koopman kernel regression," in *Advances in Neural Information Processing Systems*, vol. 36, 2023, pp. 16 207–16 221.

[34] J. L. Proctor, S. L. Brunton, and J. N. Kutz, "Generalizing koopman theory to allow for inputs and control," *SIAM Journal on Applied Dynamical Systems*, vol. 17, pp. 909–930, 1 2018.

[35] J. Proctor, S. Brunton, and J. Kutz, "Generalizing koopman theory to allow for inputs and control," *SIAM Journal on Applied Dynamical Systems*, vol. 17, 02 2016.

[36] E. Kaiser, J. N. Kutz, and S. L. Brunton, "Data-driven discovery of koopman eigenfunctions for control," *Machine Learning: Science and Technology*, vol. 2, p. 035023, 9 2021.

[37] F. Nüske and S. Klus, "Efficient approximation of molecular kinetics using random Fourier features," *The Journal of Chemical Physics*, vol. 159, no. 7, p. 074105, 08 2023.

[38] M. Khosravi, "Representer theorem for learning koopman operators," *IEEE Transactions on Automatic Control*, vol. 68, no. 5, pp. 2995–3010, 2023.

[39] R. Strässer, J. Berberich, and F. Allgöwer, "Robust data-driven control for nonlinear systems using the koopman operator*," *IFAC-PapersOnLine*, vol. 56, no. 2, pp. 2257–2262, 2023.

[40] C. Zhang and E. Zuazua, "A quantitative analysis of Koopman operator methods for system identification and predictions," *Comptes Rendus. Mécanique*, 2023, online first.

[41] Ingo Steinwart and Andreas Christmann, *Support Vector Machines*, 1st ed., ser. Information Science and Statistics. New York, NY: Springer, 2008.

[42] H. Nijmeijer and A. J. van der Schaft, *Nonlinear Dynamic Control Systems*. Springer, 1996.

[43] S. Otto, S. Peitz, and C. Rowley, "Learning bilinear models of actuated koopman generators from partially observed trajectories," *SIAM Journal on Applied Dynamical Systems*, vol. 23, no. 1, pp. 885–923, 2024.

[44] D. Goswami and D. A. Paley, "Bilinearization, Reachability, and Optimal Control of Control-Affine Nonlinear Systems: A Koopman Spectral Approach," *IEEE Transactions on Automatic Control*, vol. 67, no. 6, pp. 2715–2728, 6 2022.

[45] P. Bevanda, M. Beier, A. Capone, S. G. Sosnowski, S. Hirche, and A. Lederer, "Koopman-equivariant gaussian processes," in *The 28th International Conference on Artificial Intelligence and Statistics*, 2025.

[46] C. Williams and M. Seeger, "Using the nyström method to speed up kernel machines," in *Advances in Neural Information Processing Systems*, vol. 13, 2000.

[47] T. E. Ahmad, L. Brogat-Motte, P. Laforgue, and F. d'Alché Buc, "Sketch in, sketch out: Accelerating both learning and inference for structured prediction with kernels," 2023.

[48] Z. Li, D. Meunier, M. Mollenhauer, and A. Gretton, "Optimal rates for regularized conditional mean embedding learning," *Advances in Neural Information Processing Systems*, vol. 35, pp. 4433–4445, 2022.

[49] M. Mollenhauer and P. Koltai, "Nonparametric approximation of conditional expectation operators," 12 2020. [Online]. Available: http://arxiv.org/abs/2012.12917

[50] F. Bach, "Learning Theory from First Principles," Tech. Rep., 2023. [Online]. Available: http://mitpress.mit.edu,

[51] M. Mollenhauer, N. Mücke, and T. J. Sullivan, "Learning linear operators: Infinite-dimensional regression as a well-behaved non-compact inverse problem," 11 2022. [Online]. Available: http://arxiv.org/abs/2211.08875

[52] J.-P. Aubin, *Applied functional analysis*. John Wiley & Sons, 2011.

[53] P. Schmid, K. E. Meyer, and O. Pust, "Dynamic mode decomposition and proper orthogonal decomposition of flow in a lid-driven cylindrical cavity," 08 2009.

[54] S. Klus, I. Schuster, and K. Muandet, "Eigendecompositions of transfer operators in reproducing kernel hilbert spaces," *Journal of Nonlinear Science*, vol. 30, pp. 283–315, 2 2020.

[55] M. O. Williams, C. W. Rowley, and I. G. Kevrekidis, "A kernel-based approach to data-driven koopman spectral analysis," 2014. [Online]. Available: https://arxiv.org/abs/1411.2260

[56] E. Caldarelli, A. Chatalic, A. Colomé, C. Molinari, C. Ocampo-Martinez, C. Torras, and L. Rosasco, "Linear quadratic control of nonlinear systems with koopman operator learning and the nyström method," 3 2024. [Online]. Available: http://arxiv.org/abs/2403.02811

[57] H. W. Engl and R. Ramlau, *Regularization of Inverse Problems*. Springer Berlin Heidelberg, 2015.

[58] Z. Morrison, M. Abudia, J. A. Rosenfeld, and R. Kamalapurkar, "Dynamic mode decomposition of control-affine nonlinear systems using discrete control liouville operators," *IEEE Control Systems Letters*, 2023.

[59] S. Smale and D.-X. Zhou, "Learning theory estimates via integral operators and their approximations," *Constructive approximation*, vol. 26, no. 2, pp. 153–172, 2007.

[60] K. Muandet, K. Fukumizu, B. Sriperumbudur, B. Schölkopf *et al.*, "Kernel mean embedding of distributions: A review and beyond," *Foundations and Trends® in Machine Learning*, vol. 10, no. 1-2, pp. 1–141, 2017.

[61] M. O. Williams, M. S. Hemati, S. T. Dawson, I. G. Kevrekidis, and C. W. Rowley, "Extending data-driven koopman analysis to actuated systems," *IFAC-PapersOnLine*, vol. 49, pp. 704–709, 2016.

[62] J. A. Rosenfeld and R. Kamalapurkar, "Dynamic mode decomposition with control liouville operators," *IEEE Transactions on Automatic Control*, 2024.

[63] A. Rahimi and B. Recht, "Random features for large-scale kernel machines," in *Advances in Neural Information Processing Systems*, J. Platt, D. Koller, Y. Singer, and S. Roweis, Eds., vol. 20, 2007.

[64] A. Chatalic, N. Schreuder, L. Rosasco, and A. Rudi, "Nyström kernel mean embeddings," in *Proceedings of the 39th International Conference on Machine Learning*, ser. Proceedings of Machine Learning Research, vol. 162. PMLR, 17–23 Jul 2022, pp. 3006–3024.

[65] C. Williams and M. Seeger, "The effect of the input density distribution on kernel-based classifiers," in *ICML '00 Proceedings of the Seventeenth International Conference on Machine Learning*, 2000, pp. 1159–1166.

[66] G. Meanti, A. Chatalic, V. Kostic, P. Novelli, M. Pontil, and L. Rosasco, "Estimating koopman operators with sketching to provably learn large scale dynamical systems," in *Advances in Neural Information Processing Systems*, vol. 36, 2023, pp. 77 242–77 276.

[67] J. A. Tropp, A. Yurtsever, M. Udell, and V. Cevher, "Practical sketching algorithms for low-rank matrix approximation," *SIAM Journal on Matrix Analysis and Applications*, vol. 38, no. 4, pp. 1454–1485, 2017.

[68] A. Chatterjee, "An introduction to the proper orthogonal decomposition," *Current Science*, vol. 78, no. 7, pp. 808–817, 2000.

[69] P. J. Schmid, "Dynamic mode decomposition of numerical and experimental data," *Journal of Fluid Mechanics*, vol. 656, p. 5–28, 2010.

[70] A. Falini, "A review on the selection criteria for the truncated svd in data science applications," *Journal of Computational Mathematics and Data Science*, vol. 5, p. 100064, 2022.

[71] P. W. Bearman, "Vortex shedding from oscillating bluff bodies," *Annual Review of Fluid Mechanics*, vol. 16, no. 1, pp. 195–222, 1984.

[72] J. Decuyper, T. De Troyer, K. Tiels, J. Schoukens, and M. Runacres, "A nonlinear model of vortex-induced forces on an oscillating cylinder in a fluid flow," *Journal of Fluids and Structures*, vol. 96, p. 103029, 2020.

[73] G. Beintema, "Data–driven learning of nonlinear dynamic systems: A deep neural state–space approach," Phd Thesis, Eindhoven University of Technology, 2024.

[74] J. H. Hoekstra, B. Cseppento, G. I. Beintema, M. Schoukens, Z. Kollár, and R. Tóth, "Computationally efficient predictive control based on ann state-space models," in *2023 62nd IEEE Conference on Decision and Control (CDC)*, 2023, pp. 6336–6341.

[75] J. Berberich, J. Köhler, M. A. Müller, and F. Allgöwer, "Linear tracking mpc for nonlinear systems—part i: The model-based case," *IEEE transactions on automatic control*, vol. 67, no. 9, pp. 4390–4405, 2022.

[76] V. Nevistić and J. A. Primbs, "Constrained nonlinear optimal control: a converse HJB approach," California Institute of Technology, Tech. Rep., 1996.

[77] G. Turri, V. Kostic, P. Novelli, and M. Pontil, "A randomized algorithm to solve reduced rank operator regression," 12 2023.

APPENDIX

A. Additional Control Koopman Operator Estimators

We can state the result that readily follows as a slight modification of the results in [31], [77]. The RKHS control Koopman operators can be empirically estimated by $\widehat{S}_Z^* \boldsymbol{W} \widehat{S}_+$, where $\boldsymbol{W} \in \mathbb{R}^{n \times n}$ follows:

(1) from the *Kernel Ridge Regression (KRR)* algorithm, provided that $\widehat{G}_\gamma$ minimizes ($\widehat{\mathsf{rRM}}$) without rank constraint, resulting in

$$\widehat{G}_\gamma = \widehat{S}_Z^* (\boldsymbol{K}_Z + \gamma \boldsymbol{I})^{-1} \widehat{S}_+; \tag{KRR}$$

(2) from the *Reduced Rank Regression (RRR)* algorithm, provided that ($\widehat{\mathsf{rRM}}$) minimized under a rank constraint $\widehat{G} \in \{\mathrm{HS}(\mathcal{H}_X, \mathcal{H}_Z) \mid \mathrm{rank}(\widehat{G}) \leq r\}$ so that

$$\widehat{G}_\gamma^{\mathrm{RRR}} \triangleq C_\gamma^{-\frac{1}{2}} [\![C_\gamma^{-\frac{1}{2}} T]\!]_r = \widehat{S}_Z^* \mathsf{U}_r \mathsf{V}_r^\intercal \widehat{S}_+, \tag{RRR}$$

where $C_\gamma \triangleq C_Z + \gamma \mathrm{Id}_{\mathcal{H}_X}$ and $\mathsf{V}_r = \boldsymbol{K}_Z \mathsf{U}_r$ with $\mathsf{U}_r = [\mathsf{u}_1 \mid \cdots \mid \mathsf{u}_r] \in \mathbb{R}^{n \times r}$ is such that (σ_i, u_i) are solutions to the generalized eigenvalue problem $\boldsymbol{K}_{++} \boldsymbol{K}_Z \mathsf{u}_i = \sigma_i^2 (\boldsymbol{K}_Z + \gamma \boldsymbol{I}) \mathsf{u}_i$, s.t. $\mathsf{u}_i^\intercal \boldsymbol{K}_Z (\boldsymbol{K}_Z + \gamma \boldsymbol{I}) \mathsf{u}_i = 1$;

(3) from the *Principal Component Regression (PCR)* algorithm, giving

$$\widehat{G}_\gamma^{\mathrm{PCR}} \triangleq [\![C]\!]_r^\dagger T_Z = \widehat{S}_Z^* \mathsf{U}_r \mathsf{V}_r^\intercal \widehat{S}_+, \tag{PCR}$$

where $[\![\boldsymbol{K}_Z]\!]_r = \mathsf{V}_r \boldsymbol{\Sigma}_r \mathsf{V}_r^*$ is the r-truncated SVD of $\boldsymbol{K}_Z$, and $\mathsf{U}_r = \mathsf{V}_r \boldsymbol{\Sigma}_r^\dagger$.

B. Additional ablation

We expand the numerical study of the Duffing oscillator VIII-A, with an ablation study for even lower levels of regularization $\gamma = 10^{-10}$ and $\gamma = 10^{-11}$. As shown in Figures 13 and 14, our cKOR approach continues to significantly outperform the parametric approaches with its sketched Ny-cKOR version on par or better.

C. Koopman-Based Iterated LPV-MPC

To provide an efficient scheme for controlling the original nonlinear system via the cKOR method provided surrogate models, we propose a model predictive control (MPC) approach that extends the *iterated* LPV scheme of [74] to our control operator setting. The control problem that we want to

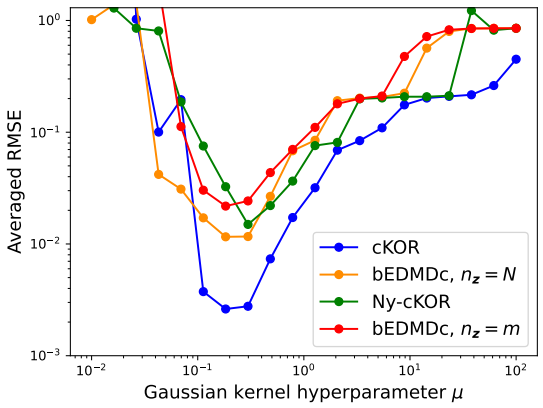


Fig. 13. RMSE of 1-step-prediction for cKOR, Ny-cKOR and bEDMDc models over values of μ and regularizer $\gamma = 10^{-10}$ for the Duffing oscillator example VIII-A.

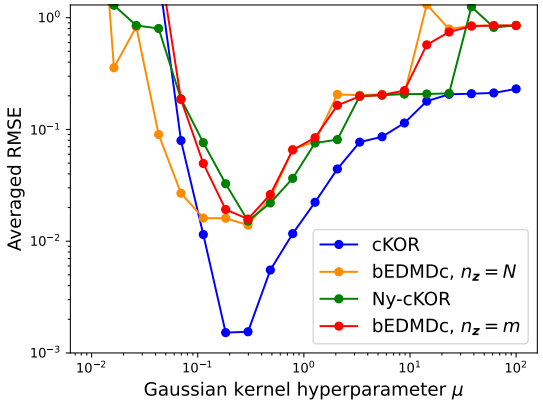


Fig. 14. RMSE of 1-step-prediction for cKOR, Ny-cKOR and bEDMDc models over values of μ and regularizer $\gamma = 10^{-11}$ for the Duffing oscillator example VIII-A.

address is that given a measurement of the state $\boldsymbol{x}(k)$ of the original nonlinear system (NCAS) at time-instant t, based on a cKOR model, solve a predictive control problem on a finite time horizon T with computation cost close to an LTI-MPC to obtain a control sequence $\{\boldsymbol{u}_{i|t}\}_{i=0}^{T-1}$ such that the predicted response of (NCAS) follows a prescribed reference trajectory. Then, $\boldsymbol{u}_{0|t}$ is applied to the system and at the next time-instant $(t+1)$, $\boldsymbol{x}(t+1)$ is measured to start the next control cycle. For our cKOR model, the cKOLPV-MPC optimization problem is

$$\min_{\boldsymbol{u}_{0|t}\cdots\boldsymbol{u}_{T-1|t}} \sum_{i=1}^{T} \left(\|\mathbf{z}_{i|t}-\overline{\mathbf{z}}_{i|t}\|^2_{\boldsymbol{Q}_\mathbf{z}}+\|\boldsymbol{u}_{i|t}-\overline{\boldsymbol{u}}_{i|t}\|^2_{\boldsymbol{R}}\right)+q_{T|t} \quad (43a)$$

$$\text{s.t.} \quad \mathbf{z}_{i+1|t} = \boldsymbol{A}\mathbf{z}_{i|t} + \mathbf{B}(\mathbf{p}_{i|t})\boldsymbol{u}_{i|t} \quad (43b)$$

$$\boldsymbol{x}_{\min} \leq \boldsymbol{C}\mathbf{z}_{j|t} \leq \boldsymbol{x}_{\max} \quad (43c)$$

$$\boldsymbol{u}_{\min} \leq \boldsymbol{u}_{i|t} \leq \boldsymbol{u}_{\max} \quad (43d)$$

where $\boldsymbol{Q}_\mathbf{z} = \boldsymbol{C}\boldsymbol{Q}_x\boldsymbol{C}$, $q_{T|t}=\|\mathbf{z}_{T|t} - \overline{\mathbf{z}}_{T|t}\|^2_{\boldsymbol{Q}_\mathbf{z}(T)}$ and the measured state at time t, i.e., $\boldsymbol{x}_{0|t} = \boldsymbol{x}(t)$, is lifted to determine $\mathbf{z}_{0|t} = \boldsymbol{z}(\boldsymbol{x}_{0|t})$. The variables $\overline{\boldsymbol{x}}_{i|t}$ and $\overline{\boldsymbol{u}}_{i|t}$ correspond to the reference state and input, respectively. In addition, $\boldsymbol{Q} \in \mathbb{R}^{n_x\times n_x}$, $\boldsymbol{R} \in \mathbb{R}^{n_u\times n_u}$ are weighting matrices and $\boldsymbol{Q}_T \in \mathbb{R}^{n_x\times n_x}$ represents the terminal weight matrix. These matrices have to be tuned with respect to user specified performance expectations w.r.t. the tracking problem. The bounds $\boldsymbol{x}_{\min}$, $\boldsymbol{x}_{\max}$, $\boldsymbol{u}_{\min}$ and $\boldsymbol{u}_{\max}$ limit the state and input sequences of the system. Lastly, to efficiently handle the bilinearity of the cKOR model, a scheduling variable taken as $\mathbf{p}_{i|t}$ is introduced so that $\mathbf{B}(\mathbf{p}_{i|t}) = [\boldsymbol{B}_1\mathbf{p}_{i|t} \mid \cdots \mid \boldsymbol{B}_{n_u}\mathbf{p}_{i|t}]$. The core idea is that at any given time-instant t, for a fixed scheduling sequence $\{\mathbf{p}_{i|t}\}_{i=0}^{T-1}$, (43b) is used to formulate a linear MPC problem that can be solved efficiently as a quadratic program (QP). Then, the resulting control sequence $\{\boldsymbol{u}_{i|t}\}_{i=0}^{T-1}$ is used to forward simulate the cKOR model to compute a new sequence $\mathbf{p}_{i|t} = \mathbf{z}(i+t)$ on which a new sequence of control matrices $\mathbf{B}(\mathbf{p}_{i|t})$ is computed in a receding horizon fashion akin to the iterated MPC scheme of [74].

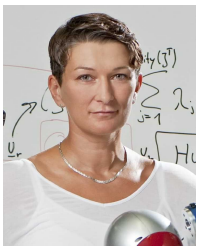
Petar Bevanda received his B.Sc. and M.Sc. degrees in Information Technology and Electrical Engineering from the University of Zagreb, Croatia and the Technical University of Munich (TUM), Germany, respectively, in 2017 and 2019. Since 2020, he has been a PhD student at the Chair of Information-oriented Control, TUM School of Computation, Information and Technology. His current research interests include operator-theoretic machine learning and data-driven control of uncertain systems.

Bas Driessen is a graduate in Mechanical Engineering from Eindhoven University of Technology (TU/e), where he obtained his master's degree in 2023 with a focus on systems and control. His master's thesis was part of an internship at the Chair of Information-oriented Control at TUM School of Computation, Information and Technology, Technical University of Munich (TUM). His research interests include modeling and identification of nonlinear systems, machine learning techniques and predictive control.

Cristi Iacob is a Doctoral Candidate at the Control Systems (CS) Group in the Department of Electrical Engineering. His current research is on modeling and analysis of nonlinear systems using the Koopman and Linear Parameter-Varying (LPV) frameworks, under the supervision of Associate Professor Roland Tóth and Assistant Professor Maarten Schoukens, in the Automated Linear Parameter-Varying Modeling and Control Synthesis for Nonlinear Complex Systems (ARPOCS) project. His main research interests include modeling and identification of nonlinear systems and machine learning techniques.

Stefan Sosnowski received the Dipl.-Ing. degree and Dr.-Ing. degree in electrical engineering from the Technical University of Munich (TUM), Munich, Germany, in 2007 and 2014, respectively. Since 2014, he is a Postdoctoral Fellow with the Chair of Information-oriented Control, at the School of Computation, Information and Technology at TUM. His research interests include bioinspired and underwater robotics, nonlinear control, distributed dynamical systems, and multi-agent systems.

Roland Tóth recieved the Ph.D. degree in electrical engineering with Cum Laude distinction from the Delft Center for Systems and Control (DCSC), Delft University of Technology (TU Delft), Delft, The Netherlands, in 2008. He was a Postdoctoral Research Fellow with TU Delft in 2009 and Berkeley in 2010. He held a position with DCSC, TU Delft in 2011–2012. He is currently a Full Professor with the Control Systems Group, Eindhoven University of Technology, Eindhoven, The Netherlands, and a Senior Researcher with the HUN-REN Institute for Computer Science and Control (SZTAKI), Budapest, Hungary. His research interests include identification and control of linear parameter-varying (LPV) and nonlinear systems, developing machine learning methods with performance and stability guarantees, model predictive control and behavioral system theory with a wide range of applications, including precision mechatronics and autonomous vehicles. Prof. Tóth is a Senior Editor of the IEEE Transactions on Control Systems Technology.



Sandra Hirche holds the TUM Liesel Beckmann Distinguished Professorship and heads the Chair of Information-oriented Control at TUM School of Computation, Information and Technology, Technical University of Munich (TUM), since 2013. She received the diploma engineer degree in Aeronautical and Aerospace Engineering in 2002 from the Technical University Berlin, Germany, and the Doctor of Engineering degree in Electrical and Computer Engineering in 2005 from TUM. Her main research interests include learning, cooperative, and distributed control with application in human-robot interaction, multirobot systems, and general robotics.