Supplementary Material

Petar Bevanda
TU Munich
petar.bevanda@tum.de

Max Beier
TU Munich
max.beier@tum.de

Armin Lederer
TU Munich
armin.lederer@tum.de

Stefan Sosnowski
TU Munich
sosnowski@tum.de

Eyke Hüllermeier
LMU Munich
eyke@ifi.lmu.de

Sandra Hirche
TU Munich
hirche@tum.de

The supplementary material is organized as follows.

- Appendix A contains additional background on non-recurrence and spectral theory of Koopman operators. Additionally, it contains a notation table.
- Proofs of theoretical results are found in Appendix B.
- Finally, Appendix C includes more details on the experimental section, as well as additional experiments.

A Non-recurrence and Koopman Operator Theory

Remark 2 (Operator boundedness)**.** *Consider a forward complete system on a compact set $\mathbb{X}$ and a continuous flow $\boldsymbol{F}^t$. It is well-known that a time-t Koopman operator $\mathcal{K}^t$ is then a contraction semigroup on $C(\mathbb{X})$ [1]. Due to forward completeness of the flow, we therefore obtain a Banach algebra $C(\mathbb{X})$ with a bounded semigroup $\{\mathcal{K}^t\}_{t\geq 0} \in \mathcal{B}(C(\mathbb{X}))$.*

Definition 2 (Non-recurrence)**.** *A non-recurrent domain is one where flow does not intersect itself.*

Non-recurrence is commonly ensured by a choice of the time interval $[0, T]$ so no periodicity is exhibited. Note that it does not mean the system's behavior is not allowed the be periodic, but our perception of it via data does. Effectively this prohibits the multi-valuedness of eigenfunctions – allowing them to define an injective feature map. Thus, non-recurrence is a certain but general condition that bounds the time-horizon T in which it is feasible to completely describe the nonlinear system's flow via an LTI predictor. It makes for a less-restrictive and intuitive condition compared to existing RKHS approaches [2, 3] that rely on the self-adjointness and compactness of the actual Koopman operator which is rarely fulfilled and hard to verify without prior knowledge.

Lemma 2 (Universality of Eigenfunctions)**.** *Consider an quantity of interest $q \in C(\mathbb{X})$, a forward-complete system flow $\boldsymbol{F}^t(\cdot)$ on a non-recurrent domain $\mathbb{X}_0$ (Definition 2) of a compact set $\mathbb{X}$. Then, the output trajectory $y(t) = q(\boldsymbol{x}(t)), \forall t \in [0, T]$ is arbitrarily closely described by the eigenpairs $\{\lambda_j, \phi_j\}_{j\in\mathbb{N}} \subseteq (\mathbb{C}\times C(\mathbb{X}))$ of the Koopman operator semigroup $\{\mathcal{K}^t\}_{t=0}^T$[1] so that $\forall \varepsilon > 0, \exists \bar{D} \in \mathbb{N}$*

$$|q(\boldsymbol{x}(t)) - \textstyle\sum_{j=1}^{\bar{D}} \mathrm{e}^{\lambda_j t}\, \phi_j(\boldsymbol{x}_0)| < \varepsilon, \forall t \in [0, T]. \tag{1}$$

Proof 1 (Lemma 2)**.** *With continuous eigenfunctions for continuous systems proved valid in [5, Lemma 5.1],[6, Theorem 1], the space of continuous functions over a compact set is naturally the space of interest. On a non-recurrent domain, there exist uniquely defined non-trivial eigenfunctions and, by [7, Theorem 3.0.2], the spectrum is rich – with any eigenvalue in the closed complex unit disk legitimate [8]. Further, by [6, Theorem 2], this richness is inherited by the Koopman eigenfunctions — making them universal approximators of continuous functions.*

[1] Note that, compared to "Koopman Mode Decomposition", we let the eigenfunctions absorb the spatial mode coefficients (possible w.l.o.g.) as they correspond to eigenfunctions and not eigenvalues [4, Definition 9].

37th Conference on Neural Information Processing Systems (NeurIPS 2023).

Table 1: Summary of used notation

Notation	Description
$\mathbb{T}$	time interval $[0, T]$
$\mathbb{H}$	collection of points from discretizing the time interval $\mathbb{T}$ at times $\{t_0, \cdots t_{\mathrm{H}}\}$
$\mathbb{X}$	compact state-space set
$\mathbb{X}_0$	compact set of initial conditions that form a non-recurrent domain
$\boldsymbol{x}_T / \boldsymbol{x}_{\mathrm{H}}$	a continuous/discrete time state trajectory
$\mathbb{X}_T / \mathbb{X}_{\mathrm{H}}$	space of continuous/discrete-time *state* trajectories
y_T / y_{H}	a continuous/discrete time output trajectory
$\mathbb{Y}_T / \mathbb{Y}_{\mathrm{H}}$	space of continuous/discrete-time *output* trajectories
$\mathcal{K}^t$	time-t Koopman operator
$M / \hat{M}$	true/learned mode decomposition operator
$K / \boldsymbol{K} / k$	operator/matrix/scalar-valued kernels
λ / μ	continuous/ discrete-time eigenvalue
$K^{\lambda_j} / \boldsymbol{K}^{\mu_j} / k^j$	operator/matrix/base kernel of the j-th Koopman eigenfunction
$\mathcal{H}^k$	RKHS of a scalar base kernel k
$\mathcal{H}^{k^{\mu_j}}$	RKHS of a scalar kernel k^{μ_j}
$\mathcal{H}^{\mu_j}$	RKHS of matrix valued kernel $\boldsymbol{K}^{\mu_j}$ induced by scalar kernel k^{μ_j}
$\mathcal{H}^{\lambda} / \mathcal{H}^{\mu}$	continuous/discrete-time Koopman eigenfunction RKHS $\lambda / \mu \in \mathbb{C}$
$\mathcal{H} / \mathcal{H}^{\Delta t}$	continuous/discrete-time Koopman RKHS
$\mathcal{I}_{\lambda}^{T} / \mathcal{I}_{\mu}^{H}$	invariance transform for time/step length T/H and eigenvalue $\lambda / \mu \in \mathbb{C}$
$\mathbb{D}_{(\cdot)}$	dataset for an estimator $(\cdot)$
$\mathbb{D}_N$	dataset of N *time-continuous* sample trajectories pairs $(\boldsymbol{x}_T^{(i)}, y_T^{(i)})_{i \in [N]}$
$\mathbb{D}_N^{\Delta t}$	dataset of N *time-discrete* sample trajectories pairs $(\boldsymbol{x}_{\mathrm{H}}^{(i)}, y_{\mathrm{H}}^{(i)})_{i \in [N]}$
$\mathcal{B}(\cdot)$	set of bounded operators over a domain
$\overline{\mathbb{B}_r(\mathbf{0})}$	closed ball of radius-r in $\mathbb{C}$
$\boldsymbol{\Gamma}$	extended observability matrix
$\hat{\boldsymbol{\phi}}(\cdot)$	vector-valued function of learned Koopman eigenfunctions
$\mathcal{R}_N(\cdot)$	true forecast risk/generalization error of an estimator
$\hat{\mathcal{R}}_N(\cdot)$	empirical forecast risk of an estimator based on N data samples
$R_N(\cdot)$	true Rademacher complexity of of a hypothesis class based on N samples
$\hat{R}_N(\cdot)$	empirical Rademacher complexity of a hypothesis class based on N samples
$\mathcal{L}(\cdot)$	loss function determining the metric for risk, e.g. squared error

Remark 3 (Choosing the spectral distribution $\lambda \sim \rho(\mu)$)**.** *The choice of our measure of integration might seem arbitrary, and it indeed is. Since we, in general, do not assume knowledge of the spectrum of the Koopman-semigroup, we have to make an approximation. To this end, an educated guess on where the (point-) spectrum might be located is helpful. As elaborated above, the Hille-Yosida-Theorem provides a convenient way to connect the practically attainable growth rates to bounds on the spectrum. Why would sampling spectral features in a set enclosing the spectrum be enough to obtain the spectral decomposition of the Koopman operator? Recalling that the spectral decomposition consists of projections to eigenspaces, we state a well-known result. The Riesz projection operator*

$P_\lambda : \mathcal{C} \mapsto \{g \in \mathcal{C} : \mathcal{K}g = \lambda g\}$ to an eigenspace of $\mathcal{K}$ can be represented by

$$P_\lambda = \frac{1}{2\pi i} \int_{\gamma_\lambda} \frac{\mathrm{d}s}{s - \mathcal{K}},$$

where γ_λ is a Jordan curve enclosing λ and no other point in $\sigma(\mathcal{K})$ [9]. Obviously $\bigcup_{\lambda \in \sigma(\mathcal{K})} \operatorname{range}(P_\lambda) = \mathcal{C}$, iterating on the fact that we can represent the operator T by its spectral components. It becomes apparent that sampling from a set enclosing $\sigma(\lambda)$ can be seen as sampling curves, eventually enclosing sufficient spectral components. And as stated, one can choose arbitrary measures on $\mathbb{C}$ as long as one ensures they enclose the spectrum. The preceding analysis sheds light on the connection of our approach to the Laplace-Stieltjes transform and the spectral pollution occurring in EDMD-type algorithms.

B Proofs of Theoretical Results

Proofs for Section 3 Koopman Kernel Regression

Proof 2 (Lemma 1)**.** *Due to the boundedness of finite-time trajectories of a forward complete system and a continuous $g \in C(\mathbb{X}_0)$ we have well-defined Haar integral invariants [10]*

$$\phi_\lambda(\boldsymbol{x}_T) = \int_{\tau=0}^{T} \mathrm{e}^{-\lambda(\tau - t)}\, \mathcal{K}^\tau g(\boldsymbol{x}(0)) d\tau = \int_{0}^{T} \mathrm{e}^{-\lambda(\tau - t)}\, g(\boldsymbol{F}^\tau(\boldsymbol{x}_0)) d\tau. \tag{2}$$

Then, $\phi_\lambda : \mathbb{X}_0 \mapsto C(\mathbb{X}_0)$ [11, p. 64] is an invariant function for $\{\mathrm{e}^{-\lambda\tau}\, \mathcal{K}^\tau\}_{\tau=0}^{T}$ considering a normalized measure $d\tau(T) = 1$ – fulfilling the Koopman-invariance condition. By simple algebraic manipulation we verify that ϕ_λ indeed has LTI dynamics

$$\begin{aligned} \phi_\lambda(\boldsymbol{x}_T) &= \int_{\tau=0}^{T} \mathrm{e}^{-\lambda(\tau - t)}\, g(\boldsymbol{F}^\tau(\boldsymbol{x}_0)) d\tau \\ &= \mathrm{e}^{\lambda t} \int_{\tau=0}^{T} \mathrm{e}^{-\lambda\tau}\, g(\boldsymbol{F}^\tau(\boldsymbol{x}_0)) d\tau \\ &= \mathrm{e}^{\lambda t}\, \phi_\lambda(\boldsymbol{x}_0). \end{aligned} \tag{3}$$

Proof 3 (Theorem 1)**.** *(i) Due to the one-to-one relationship between kernel functions and RKHS we can examine $\mathcal{H}^\lambda$ by its kernel $K^\lambda(\cdot, \cdot)$. We notice that due to the property that pointwise converging sequences of kernels are again kernels [12, Corollary 4.17]. Showing that K^λ is a kernel thus reduces to showing that the double integral exists. Now, since our continuity assumptions on the system ensure the convergence of the Haar-integrals [11, p. 64], we can conclude that any valid integration scheme [13, Theorem A.1.5] induces a uniformly converging sequence of kernels.*

(ii) We will prove the statement by showing that the universality of the base kernel for continuous functions makes the Koopman eigenfunction RKHS $\mathcal{H}^\lambda$ universal for continuous Koopman-invariant functions at eigenvalue $\lambda \in \mathbb{C}$. It is clear that feature map of the kernel is $\{\mathrm{e}^{-\lambda\tau}\, \mathcal{K}^\tau\}_{\tau=0}^{T}$-invariant, and we only need to prove the completeness part. Let $\mathbb{X}_0$ be a compact subset in $\mathbb{X}$, and $\epsilon > 0$. Then, the non-recurrent domain defined by $\mathbb{X}_T = \cup_{t \in [0,T]} \boldsymbol{F}^t(\mathbb{X}_0)$ under the continuous map $(t, \boldsymbol{x}) \mapsto \boldsymbol{F}^t(\boldsymbol{x})$ is also a compact set. By using a universal RKHS $\mathcal{H}^k$, we know there exists $f \in \mathcal{H}^k$ so that

$$\sup_{\boldsymbol{x} \in \mathbb{X}_T} |f(\boldsymbol{x}) - \phi_\lambda(\boldsymbol{x})| \leq \epsilon.$$

Consider now a $\{\mathrm{e}^{-\lambda\tau}\,\mathcal{K}^\tau\}_{\tau=0}^T$-invariant group-averaged map $f_\lambda(\boldsymbol{x}) = \int_{\tau=0}^T \mathrm{e}^{-\lambda\tau}\, f\left(\boldsymbol{x}(\tau)\right) d\tau$ from the Koopman eigenfunction RKHS $\mathcal{H}^\lambda$ induced by Lemma 1. Then due to

$$
\begin{aligned}
\sup_{\boldsymbol{x}\in\mathbb{X}_0} |f_\lambda(\boldsymbol{x}) - \phi_\lambda(\boldsymbol{x})| &= \sup_{\boldsymbol{x}\in\mathbb{X}_0} \left| \int_{\tau=0}^T \left(\mathrm{e}^{-\lambda\tau}\, f\left(\boldsymbol{x}(\tau)\right) - \mathrm{e}^{-\lambda\tau}\, \phi_\lambda\left(\boldsymbol{x}(\tau)\right)\right) d\tau \right| \\
\text{(triangle inequality)} \quad &\leq \sup_{\boldsymbol{x}\in\mathbb{X}_0} \int_{\tau=0}^T \left| \left(\mathrm{e}^{-\lambda\tau}\, f\left(\boldsymbol{x}(\tau)\right) - \mathrm{e}^{-\lambda\tau}\, \phi_\lambda\left(\boldsymbol{x}(\tau)\right)\right) \right| d\tau \\
&\leq \int_{\tau=0}^T \sup_{\boldsymbol{x}\in\mathbb{X}_0} \left| \left(\mathrm{e}^{-\lambda\tau}\, f\left(\boldsymbol{x}(\tau)\right) - \mathrm{e}^{-\lambda\tau}\, \phi_\lambda\left(\boldsymbol{x}(\tau)\right)\right) \right| d\tau \\
\text{(Cauchy–Schwarz inequality)} \quad &\leq \int_{\tau=0}^T \left|\mathrm{e}^{-\lambda\tau}\right| \sup_{\boldsymbol{x}\in\mathbb{X}_0} \left| f\left(\boldsymbol{x}(\tau)\right) - \phi_\lambda\left(\boldsymbol{x}(\tau)\right)\right| d\tau \\
&\leq \sup_{\tau'\in[0,T]} \left|\mathrm{e}^{-\lambda\tau'}\right| \int_{\tau=0}^T \sup_{\boldsymbol{x}\in\mathbb{X}_T} \left| f\left(\boldsymbol{x}\right) - \phi_\lambda\left(\boldsymbol{x}\right)\right| d\tau \\
&= \max\{1, \left|\mathrm{e}^{-\lambda T}\right|\} T\epsilon,
\end{aligned}
$$

we can approximate any Koopman eigenfunction ϕ_λ with a Koopman-invariant function f_λ to arbitrary accuracy.

(iii) With the knowledge of an explicit LTI feature representation from Lemma 1, we show that $\mathcal{H}^\lambda$ satisfies Koopman-invariance along sampled trajectories $\{\boldsymbol{x}_T^{(i)}\}_{i=1}^N$. For representing an open eigenfunction over an initial condition, we choose an RKHS $\mathcal{H}^k$ of a universal kernel $k(\cdot,\cdot)\colon \mathbb{X}\times\mathbb{X} \mapsto \mathbb{R}$. As a consequence of Mercer's theorem [14], there exists a feature map $\boldsymbol{\xi}\colon \mathbb{R}^d \mapsto \mathcal{H}^k$ for every kernel $k(\cdot,\cdot)$ such that

$$
k(\cdot,\cdot) = \langle \boldsymbol{\xi}(\cdot), \boldsymbol{\xi}(\cdot)\rangle_{\mathcal{H}^k}. \tag{4}
$$

Due to universality of $k(\cdot,\cdot)$ and continuity of eigenfunctions [5], there exists a parameter vector $\boldsymbol{\theta}$ so that

$$
g(\boldsymbol{x}_T^{(i)}(0)) = \langle \boldsymbol{\theta}, \boldsymbol{\xi}(\boldsymbol{x}_T^{(i)}(0))\rangle_{\mathcal{H}^k}, \quad \forall i = 1,\ldots,N. \tag{5}
$$

To enforce Lemma 1 at data points we utilize an RKHS $\mathcal{H}^\lambda$ induced by $\mathcal{I}_\lambda^T : \mathcal{H}^k \to \mathcal{H}^\lambda$. Due to universality for arbitrary continuous Koopman eigenfunctions by (ii), there exists a parameter vector $\boldsymbol{\alpha}$ so that

$$
f_\lambda(\boldsymbol{x}_T^{(i)}) = \langle \boldsymbol{\alpha}, \mathcal{I}_\lambda^T \boldsymbol{\xi}(\boldsymbol{x}_T^{(i)}(0))\rangle_{\mathcal{H}^\lambda}, \quad \forall i = 1,\ldots,N. \tag{6}
$$

From (6) we recognize a modified feature map $\boldsymbol{\psi}(\cdot) = \mathcal{I}_\lambda^T \boldsymbol{\xi}(\cdot)$, representing the eigenfunction flow at $\boldsymbol{x}_T^{(i)}, i = 1,\ldots,N$, $\forall t \in [0,T]$

$$
f_\lambda(\boldsymbol{x}_T) = \langle \boldsymbol{\alpha}, \boldsymbol{\psi}(\boldsymbol{x}_T^{(i)})\rangle_{\mathcal{H}^\lambda}, \quad \forall i = 1,\ldots,N, \tag{7}
$$

inducing a kernel

$$
K^\lambda(\cdot,\cdot) = \langle \boldsymbol{\psi}(\cdot), \boldsymbol{\psi}(\cdot)\rangle_{\mathcal{H}^\lambda}. \tag{8}
$$

By exploiting inner product properties, we recognize

$$
K^\lambda(\cdot,\cdot) = \langle \mathcal{I}_\lambda^T \boldsymbol{\xi}(\cdot), \mathcal{I}_\lambda^T \boldsymbol{\xi}(\cdot)\rangle_{\mathcal{H}^\lambda}, \tag{9}
$$

leading to

$$
K^\lambda(\boldsymbol{x}_T, \boldsymbol{x}_T') = \mathcal{I}_\lambda^T (\mathcal{I}_\lambda^T)^* \langle \boldsymbol{\xi}(\boldsymbol{x}_T(0)), \boldsymbol{\xi}(\boldsymbol{x}_T'(0))\rangle_{\mathcal{H}^k} = \mathcal{I}_\lambda^T k(\boldsymbol{x}_T(0), \boldsymbol{x}_T'(0)) \mathcal{I}_{\lambda^*}^{T'}. \tag{10}
$$

Finally, by applying the operators to the kernel, we obtain the induced "Koopman kernel"

$$
K^\lambda(\boldsymbol{x}_T, \boldsymbol{x}_T') = \int_{\tau=0}^T \int_{\tau'=0}^T \frac{k\left(\boldsymbol{x}_T(\tau), \boldsymbol{x}_T'(\tau')\right))}{\mathrm{e}^{\lambda(\tau-t)}\, \mathrm{e}^{\lambda^*(\tau'-t)}} d\tau d\tau'. \tag{11}
$$

fulfilling Lemma 1 along sampled trajectories $\boldsymbol{x}_T^{(i)}, i,\ldots,N$.

Proof 4 (Proposition 1). *(i) We show that $\mathcal{H}$ is an RKHS by showing it is associated with a kernel which is the limit of a pointwise converging sequence of kernels [12, Corrollary 4.17]. Since K^λ is a finite sum, it is bounded by virtue of its elements being bounded, which is due to Theorem 1.*

(ii) Universality of $\mathcal{H}$ is guaranteed by using eigenspace universality [6, Theorem 2] and applying Theorem 1 (ii) component-wise. Our goal is to represent a function in terms of an LTI predictor, the mode composition of the Koopman operator. Due to Proposition 1, we know the exact mode decomposition M is countable so the contribution of neglected eigenspaces can be made arbitrarily small by choosing $\bar{D}$ large enough.

$$\begin{aligned}
\|y_T - \hat{M}(\boldsymbol{x}_T)\|_{\mathbb{Y}_T} &= \|M(\boldsymbol{x}_T) - \hat{M}(\boldsymbol{x}_T)\|_{\mathbb{Y}_T} \\
&= \|\mathbf{1}^\top[\phi_{\lambda_1} \cdots \phi_{\lambda_{\bar{D}}}](\boldsymbol{x}_T) - \mathbf{1}^\top[\hat{\phi}_{\lambda_1} \cdots \hat{\phi}_{\lambda_D} \cdots](\boldsymbol{x}_T)\|_{\mathbb{Y}_T} \\
&= \|\phi_{\lambda_1} - \hat{\phi}_{\lambda_1} + \cdots + \phi_{\lambda_{\bar{D}}} - \hat{\phi}_{\lambda_{\bar{D}}} + \sum_{j=\bar{D}+1}^{\infty} \phi_{\lambda_j}\|_{\mathbb{Y}_T} \\
&\leq \|\phi_{\lambda_1} - \hat{\phi}_{\lambda_1}\|_{\mathbb{Y}_T} + \cdots + \|\phi_{\lambda_{\bar{D}}} - \hat{\phi}_{\lambda_{\bar{D}}}\|_{\mathbb{Y}_T} + \delta \\
&\overset{\text{Proposition 1 (ii)}}{\leq} \epsilon_1 + \cdots + \epsilon_{\bar{D}} + \delta
\end{aligned}$$

Now choosing $\bar{D}$ such that $\delta < \epsilon$ and $\epsilon_i = \frac{\epsilon - \delta}{\bar{D}}$, yields the assertion.

Proof 5 (Corollary 1). *(i) By considering the integral equation (11) at H regular intervals Δt so that $H = T/\Delta t$ with $\forall t \in \{t_k\}_{k=0}^{H}$ the integrals are replaced by sums. Due to considering normalized measures of $d\tau(T)$ and $d\tau'(T)$ in (11), each sum is normalized by the number of elements $(H+1)$. All properties from Theorem 1 transfer straightforwardly using the same arguments as in Proof 3.*

(ii) The construction of the kernel matrix sum directly follows directly follows the direct Hilbert space sum

$$\tilde{\mathcal{H}}^{\Delta t} = \mathcal{H}^{\mu_1} \oplus \cdots \oplus \mathcal{H}^{\mu_{\bar{D}}} \quad \text{so that} \quad \mathcal{H}^{\Delta t} = \operatorname{range}(\mathcal{S}) := \{f_1 + \ldots + f_{\bar{D}} : f_1 \in \mathcal{H}^{\mu_1}, \ldots, f_{\bar{D}} \in \mathcal{H}^{\mu_{\bar{D}}}\} \tag{12}$$

All properties straightforwardly transfer from Proposition 1 using the same arguments as in Proof 4.

Proof 6 (Proposition 2). *It is easily recognizable that the time-discretization of problem the functional regressino problem reads*

$$\min_{\underline{\boldsymbol{\beta}}^\top = [\boldsymbol{\beta}_1 \cdots \boldsymbol{\beta}_N]} \textstyle\sum_{i=1}^{N} \|y_{\mathrm{H}}^{(i)} - \boldsymbol{K}(\boldsymbol{x}_{\mathrm{H}}^{(i)}, \boldsymbol{X}_{\mathrm{H}})\boldsymbol{\beta}_i\|_{\mathbb{Y}_{\mathrm{H}}}^2 + \gamma \boldsymbol{\beta}_i^\top \boldsymbol{K}(\boldsymbol{x}_{\mathrm{H}}^{(i)}, \boldsymbol{x}_{\mathrm{H}}^{(i)})\boldsymbol{\beta}_i. \tag{13}$$

with $\underline{\boldsymbol{\beta}}$ the unique solution to the system of linear equations

$$(\boldsymbol{K}(\boldsymbol{X}_{\mathrm{H}}, \boldsymbol{X}_{\mathrm{H}}) + \gamma \boldsymbol{I}_{H+1} \otimes \boldsymbol{I}_N) \underbrace{[\boldsymbol{\beta}_1^\top, \ldots, \boldsymbol{\beta}_N^\top]^\top}_{\underline{\boldsymbol{\beta}}} = \underbrace{[y_{\mathrm{H}}^{(1)^\top}, \ldots, y_{\mathrm{H}}^{(N)^\top}]^\top}_{\boldsymbol{y}_{\mathrm{H}}}, \tag{14}$$

Due to being a particular case linear coregionalization models [15, 16], it follows that the approximations $\hat{\phi}_j(\cdot)$ of Koopman eigenfunctions over trajectory samples are uniquely defined by

$$\hat{\phi}_j(\boldsymbol{x}_{\mathrm{H}}) = \sum_{i=1}^{N} \left(k^{\mu_j}\left(\boldsymbol{x}_{\mathrm{H}}, \boldsymbol{x}_{\mathrm{H}}^{(i)}\right) \otimes \boldsymbol{\mu}_j^{*\top} \right) \boldsymbol{\beta}_i = k^{\mu_j}_{\boldsymbol{X}_{\mathrm{H}}\boldsymbol{X}_{\mathrm{H}}} \left(\boldsymbol{I}_N \otimes \boldsymbol{\mu}_j^{*\top} \right) \underline{\boldsymbol{\beta}}. \tag{15}$$

As a consequence of a non-recurrent domain, the time-discrete invariance transformation is a bijection at time-instances of the trajectory. Therefore, a base kernel RKHS $\mathcal{H}^{k^j}$ is isometric to $\mathcal{H}^{k^{\mu_j}}$ with isometry $\mathcal{I}_{\mu_j}^H$, it is guaranteed $\forall \boldsymbol{x}_{\mathrm{H}} \in \mathbb{D}_N^{\Delta t} \mid \boldsymbol{x}_0 \equiv \boldsymbol{x}_{\mathrm{H}}(0)$

$$\hat{\phi}_j(\boldsymbol{x}_0) = \hat{\phi}_j(\boldsymbol{x}_{\mathrm{H}}), \tag{16a}$$

$$k^j_{\boldsymbol{X}_0\boldsymbol{X}_0}\boldsymbol{\alpha}_j = k^{\mu_j}_{\boldsymbol{X}_{\mathrm{H}}\boldsymbol{X}_{\mathrm{H}}} \left(\boldsymbol{I}_N \otimes \boldsymbol{\mu}_j^{*\top} \right) \underline{\boldsymbol{\beta}}. \tag{16b}$$

Then via $\boldsymbol{\alpha}_j = k^{-1}_{\boldsymbol{X}_0\boldsymbol{X}_0} k^{\mu_j}_{\boldsymbol{X}_{\mathrm{H}}\boldsymbol{X}_{\mathrm{H}}} \left(\boldsymbol{I}_N \otimes \boldsymbol{\mu}_j^{*\top} \right) \underline{\boldsymbol{\beta}}$ eigenfunctions are uniquely determined as

$$\hat{\boldsymbol{\phi}}(\boldsymbol{x}_0) = \left[k^j_{\boldsymbol{x}_0\boldsymbol{X}_0}\boldsymbol{\alpha}_j \right]_{j=1}^{\bar{D}}, \tag{17}$$

concluding the proof.

Proof 7 (Proposition 3). *Due to [7, Theorem 3.0.2], we consider, w.l.o.g., a dense set* $\{\mu_j\}_{j=1}^{\infty}$ *in* $\overline{\mathbb{B}_1(\mathbf{0})}$ *and a finite-rank kernel* $\tilde{\boldsymbol{K}} = \sum_{j=1}^{D} \boldsymbol{K}^{\mu_j}(\boldsymbol{x}_\mathrm{H}, \boldsymbol{x}_\mathrm{H}')$. *As the "oracle" kernel* $\boldsymbol{K} = \int_{\mu \sim \rho(\overline{\mathbb{B}_1(\mathbf{0})})} \boldsymbol{K}^{\mu}(\boldsymbol{x}_\mathrm{H}, \boldsymbol{x}_\mathrm{H}')\,\mathrm{d}\mu$ *is an operator norm limit of compact Riemann sums* $\tilde{\boldsymbol{K}}$ *on a Hilbert space* $\mathbb{Y}_\mathrm{H}$, *it is a compact operator. Thus, by [17, Theorem II (p. 374)],* $\tilde{\boldsymbol{K}} \to \boldsymbol{K}$ *uniformly as* $D \to \infty$.

Proof 8 (Theorem 2). *Consider a universal Koopman kernel* $\boldsymbol{K}$. *Consider the base kernel is Mercer and recall the properties of the invariance transformation from Proof 5: the matrix-valued kernel* $\boldsymbol{K}$ *is trace-class as* $\mathcal{I}_\mu^H \mathcal{I}_\mu^{H*}$ *is a bounded self-adjoint operator [18] and the base kernel is Mercer [14]. With Proposition 1, the universal consistency is immediate via [19]. Thus, as* $N \to \infty$, *the mode decomposition is consistent* $\|M - \hat{M}\|_{\mathbb{Y}_\mathrm{H}} \to 0$ *and the same immediately follows for individual eigenfunctions as the universality of summand RKHSs is unaffected so* $\|\phi_{\mu_j} - \hat{\phi}_{\mu_j}\|_{\mathbb{Y}_\mathrm{H}} \to 0$, $j = 1, \ldots, \overline{D}$.

Proofs for Section 4 Generalization Gap: Uniform Bounds We use the seminal result of [20], which we will restate here for completeness.

Theorem 4 (Rademacher Generalization Risk Bound, [20] – Theorem 8, 11). *Consider a loss function* $\mathcal{L} : \mathcal{Y} \times \mathcal{A} \to [0, 1]$. *Let* $\mathcal{F}$ *be a class of functions with signature* $\mathcal{X} \to \mathcal{A}$ *and let* $\{X_i, Y_i\}_{i=1}^N$ *be independently selected according to the probability measure P. then, for any integer* n *and any* $\delta \in (0, 1)$, *with probability at least* $1 - \delta$ *over samples of length* n, *every* $f \in \mathcal{F}$ *satisfies*

$$\mathbb{E}[\mathcal{L}(Y, f(X))] \leq \hat{\mathbb{E}}^N[\mathcal{L}(Y, f(X))] + 2L(\mathcal{L}_0)R_N(\mathcal{F}) + \sqrt{\frac{8 \log \frac{2}{\delta}}{N}},$$

where $\mathcal{L}_0(y, a) = \mathcal{L}(y, a) - \tilde{\mathcal{L}}(y, 0)$.

To apply it to our use-case, we need to quantify the Rademacher complexities of our hypothesis space for which we make the following assumption.

Assumption 2 (Bounded RKHS Norm). *The unknown function* M *has a bounded norm in the RKHS* $\mathcal{H}^{\Delta t}$ *attached to the Koopman kernel* $\boldsymbol{K}(\cdot, \cdot)$, *i.e.,* $\|M\|_{\mathcal{H}^{\Delta t}} \leq B$ *for some* $B \in \mathbb{R}_+$.

An extension of classical results for operator-valued Rademacher complexities:

Lemma 3 (Rademacher Complexities of the Koopman Kernel). *Consider the, Mercer, Koopman kernel* $\boldsymbol{K}$ *and* $\mathcal{H}^{\Delta t}$ *its RKHS as defined in Proof 5 and* $T_{\boldsymbol{K}} g = \int_{\mathbb{X}_\mathrm{H}} \boldsymbol{K}(\cdot, \boldsymbol{x}_\mathrm{H}) g(\boldsymbol{x}_\mathrm{H})\,\mathrm{d}\tilde{\boldsymbol{x}}_\mathrm{H}$ *the corresponding integral operator on* $L^2(\mathbb{X}_\mathrm{H})$. *Then under Assumption 2, the Rademacher complexities of* $\mathcal{H}^{\Delta t}$ *are upper bounded by*

$$\text{Asymptotic: } R_N(\mathcal{H}^{\Delta t}) \leq \tfrac{B}{\sqrt{N}} \sqrt{\operatorname{trace}\left(T_{\boldsymbol{K}}\right)} \quad \text{Non-Asymptotic: } \hat{R}_N(\mathcal{H}^{\Delta t}) \leq \tfrac{B}{N} \sqrt{\operatorname{trace}\left(T_{\boldsymbol{K}}^N\right)}.$$

Proof 9 (Lemma 3). *We derive an upper bound on the Rademacher complexities of the Koopman kernel using a procedure similar to the one described in [20, Lemma 22]. Let* X_i *be random element of* $(\mathbb{X}_\mathrm{H}, \rho_\mathcal{D})$ *and* σ *a vector of independent uniform random functions on* $\{-1, 1\}$, *then the* n*-th Rademacher complexity of* $\mathcal{F}$ *is defined as*

$$R_N(\mathcal{F}) = \mathbb{E}_{\sigma, \rho_\mathcal{D}} \sup_{f \in \mathcal{F}} \frac{1}{N} \sum_{i=1}^{N} |\langle \sigma_i, f(X_i) \rangle| \overset{\text{scalar}}{=} \mathbb{E}_{\sigma, \rho_\mathcal{D}} \sup_{f \in \mathcal{F}} \frac{1}{N} \sum_{i=1}^{N} \sigma_i f(X_i).$$

The empirical case $\hat{R}_n$ *is similar to the expectation of* σ. *Now consider the Rademacher complexities of the RKHS* $\mathcal{H}^{\Delta t}$ *corresponding to the Koopman kernel for some fixed* D, *with respect to initial*

conditions $\boldsymbol{x}_{\mathrm{H}}^{(i)}$ drawn from $(\mathbb{X}_{\mathrm{H}}, \rho_{\mathcal{D}})$.

$$
\begin{aligned}
R_N(\mathcal{H}_N^{\Delta t}) &= \mathbb{E}_{\sigma,\rho_{\mathcal{D}}} \sup_{M \in \mathcal{H}_N^{\Delta t}} \frac{1}{N} \sum_{i=1}^{N} |\langle \sigma_i, M(\boldsymbol{x}_{\mathrm{H}}^{(i)}) \rangle| \\
&\leq \\
R_N(\mathcal{H}^{\Delta t}) &= \mathbb{E}_{\sigma,\rho_{\mathcal{D}}} \sup_{M \in \mathcal{H}^{\Delta t}} \frac{1}{N} \sum_{i=1}^{N} |\langle \sigma_i, M(\boldsymbol{x}_{\mathrm{H}}^{(i)}) \rangle| && \textit{Pre-RKHS property} \\
&\leq \mathbb{E}_{\sigma,\rho_{\mathcal{D}}} \sup_{M \in \mathcal{H}^{\Delta t}} \frac{1}{N} \sum_{i=1}^{N} \|\sigma_i\|_2 \|M(\boldsymbol{x}_{\mathrm{H}}^{(i)})\|_2 && \textit{Hölder's inequality} \\
&= \mathbb{E}_{\rho_{\mathcal{D}}} \sup_{M \in \mathcal{H}^{\Delta t}} \frac{1}{N} \sum_{i=1}^{N} \|M(\boldsymbol{x}_{\mathrm{H}}^{(i)})\|_2 && \textit{property of Rademacher functions} \\
&\leq \mathbb{E}_{\rho_{\mathcal{D}}} \sup_{\|\underline{\boldsymbol{\beta}}\| \leq B} \frac{1}{N} \sum_{i=1}^{N} \|\boldsymbol{K}(\boldsymbol{x}_{\mathrm{H}}^{(i)}, \cdot)\boldsymbol{\beta}\|_2 && \textit{by construction} \\
&\leq \mathbb{E}_{\rho_{\mathcal{D}}} \frac{1}{N} \sum_{i=1}^{N} B \|\boldsymbol{K}(\boldsymbol{x}_{\mathrm{H}}^{(i)}, \cdot)\|_2 && \textit{operator norm} \\
&= \mathbb{E}_{\rho_{\mathcal{D}}} \frac{B}{N} \sum_{i=1}^{N} \sqrt{\boldsymbol{K}(\boldsymbol{x}_{\mathrm{H}}^{(i)}, \boldsymbol{x}_{\mathrm{H}}^{(i)})} && \textit{reproducing property}
\end{aligned}
$$

By applying concavity and the respective definition, it follows that

$$
R_N(\mathcal{H}^{\Delta t}) \leq \frac{B}{\sqrt{N}} \sqrt{\frac{1}{N} \sum_{i=1}^{N} \mathbb{E}_{\rho_{\mathcal{D}}} \boldsymbol{K}(\boldsymbol{x}_{\mathrm{H}}^{(i)}, \boldsymbol{x}_{\mathrm{H}}^{(i)})} = \frac{B}{\sqrt{N}} \sqrt{\operatorname{trace}\left(T_{\boldsymbol{K}}\right)}
$$

and

$$
\hat{R}_N(\mathcal{H}^{\Delta t}) \leq \frac{B}{N} \sum_{i=1}^{N} \sqrt{\boldsymbol{K}(\boldsymbol{x}_{\mathrm{H}}^{(i)}, \boldsymbol{x}_{\mathrm{H}}^{(i)})} \leq \frac{B}{N} \sqrt{\operatorname{trace}\left(T_{\boldsymbol{K}}^N\right)}.
$$

Note that the different exponent in n stems from the different definitions of the operator and matrix trace*.*

Apart from the data density dependencies, the complexity of the hypothesis space is captured by the trace of the integral operator, the Gramian, iterating on a well-known property of RKHS methods. Naturally, this provides little insight asymptotically as the trace of an operator is not immediately assessable. Treatment of the trace in the asymptotic case is provided in the following result on the generalization gap of KKR, which we are now ready to state.

Theorem 3 (Generalization Gap of KKR)**.** *Let $\mathbb{D}_N^{\Delta t} = \{\boldsymbol{x}_{\mathrm{H}}^{(i)}, y_{\mathrm{H}}^{(i)}\}_{i=1}^N$ be a dataset of time-discrete trajectories consistent with a Lipschitz system on a non-recurrent domain. Then the generalization gap of a model $\hat{M}$ from Proposition 2 under Assumption 2 is, with probability $1-\delta$, upper bounded by*

$$
|\mathcal{R}(\hat{M}) - \hat{\mathcal{R}}_N(\hat{M})| \leq 4RB\sqrt{\frac{\kappa H^2}{N}} + \sqrt{\frac{8 \log \frac{2}{\delta}}{N}} \in \mathcal{O}\left(\frac{H}{\sqrt{N}}\right), \tag{18}
$$

where R is an upper bound on the loss in the domain, and κ the supremum of the base kernel.

Proof 10 (Theorem 3)**.** *The statements follow by combining Theorem 4 with approximations of the Rademacher complexities of the Koopman kernel RKHS provided in Lemma 3. In the asymptotic case,*

the behaviour of $\operatorname{trace}(T_{\boldsymbol{K}})$ *is of interest. We employ the following upper bound.*

$$
\begin{aligned}
\operatorname{trace}(T_{\boldsymbol{K}}) &= \sum_i \langle T_{\boldsymbol{K}} e_i, e_i \rangle && \textit{by definition} \\
&= \sum_i \left\langle T_{\boldsymbol{K}}^{\frac{1}{2}} e_i, T_{\boldsymbol{K}}^{\frac{1}{2}\,\star} e_i \right\rangle && \textit{trace-class property} \\
&= \int_{\mathbb{X}} \langle \boldsymbol{K}(\cdot, \boldsymbol{x}_{\mathrm{H}}), \boldsymbol{K}(\cdot, \boldsymbol{x}_{\mathrm{H}}) \rangle \, \mathrm{d}\boldsymbol{x}_{\mathrm{H}} && \textit{kernel trick} \\
&= \int_{\mathbb{X}} \boldsymbol{K}(\boldsymbol{x}_{\mathrm{H}}, \boldsymbol{x}_{\mathrm{H}}) \, \mathrm{d}\boldsymbol{x}_{\mathrm{H}} && \textit{reproducing property} \\
&= \int_{\mathbb{X}} \int_{\rho_\mu} \boldsymbol{K}^{\mu}(\boldsymbol{x}_{\mathrm{H}}, \boldsymbol{x}_{\mathrm{H}}) \, \mathrm{d}\mu \, \mathrm{d}\boldsymbol{x}_{\mathrm{H}} && \textit{Koopman kernel} \\
&= \int_{\mathbb{X}} \int_{\rho_\mu} \boldsymbol{C}(\mu, H) \boldsymbol{K}_0{}^{\mu}(\boldsymbol{x}_{\mathrm{H}}, \boldsymbol{x}_{\mathrm{H}}) \, \mathrm{d}\mu \, \mathrm{d}\boldsymbol{x}_{\mathrm{H}} && \textit{Koopman kernel flow} \\
&\leq \|\boldsymbol{C}(\mu, H)\| \int_{\mathbb{X}} \int_{\rho_\mu} \boldsymbol{K}^{\mu}(\boldsymbol{x}_{\mathrm{H}}, \boldsymbol{x}_{\mathrm{H}}) \, \mathrm{d}\mu \, \mathrm{d}\boldsymbol{x}_{\mathrm{H}} && \textit{Fubini's Theorem} \\
&\leq \|\boldsymbol{C}(\mu, H)\| \sup_{\boldsymbol{x}_{\mathrm{H}}} [\boldsymbol{K}_0^{\mu}] H \int_{\mathbb{X}} \int_{\rho_\mu} \mathrm{d}\boldsymbol{x}_{\mathrm{H}} \, \mathrm{d}\mu && \textit{Gershgorin Circle Theorem} \\
&= \|\boldsymbol{C}(\mu, H)\| \kappa H \int_{\mathbb{X}} \int_{\rho_\mu} \mathrm{d}\boldsymbol{x} \, \mathrm{d}\mu && \textit{bounded kernel} \\
&= \|\boldsymbol{C}(\mu, H)\| \kappa H && \textit{appropriate normalization} \\
&\leq 1 H \kappa H = \kappa H^2 && \textit{Gershgorin Circle Theorem (again)}
\end{aligned}
$$

Where $\boldsymbol{K}^{\mu} = \boldsymbol{C}(\mu, H) \boldsymbol{K}_0{}^{\mu}$ *is the decomposition of the eigenfunction kernel into an evaluation at a point* in space $\boldsymbol{K}_0{}^{\mu} = \boldsymbol{K}^{\mu}|_{t=0}$ *and its flow* in time $\boldsymbol{C}(\mu, H) = \mu^k \otimes \mu^{k\,\star} \in \mathbb{C}^{H \times H}$ *defined by the outer product of the eigenfunction flow. Consequently, the last inequality follows from the fact that exponential frequencies do not diverge within a finite number of steps* H.

The last ingredient we need is an approximation of the Lipschitz constant $L(\mathcal{L}_0)$. *Consider the Representation-Error* $\|\boldsymbol{y}_T - \hat{M}(\boldsymbol{x}_T)\| \leq R$. *On our non-recurrent domain of finite time* $\boldsymbol{y}_T$ *does not diverge, neither does* $\hat{M}(\boldsymbol{x}_T)$, *since we solve a regularized problem. This entails the boundedness of* $\mathcal{L}$ *by* R. *Thus, the squared error loss is Lipschitz with constant* $L = \sup_{\boldsymbol{x}} \frac{\partial}{\partial \boldsymbol{x}} \mathcal{L}(\boldsymbol{x}) = 2R$.

We can now combine the preceding investigations with Theorem 4 and obtain our claim immediately.

C Numerical Evaluation Details and Additional Experiments

All of the experiments were performed on a machine with 2TB of RAM, 8 NVIDIA Tesla P100 16GB GPUs and 4 AMD EPYC 7542 CPUs.

The comparisons to PCR (EDMD) and RRR are done utilizing code accompanying [2] available at https://github.com/csml-iit-ucl/kooplearn[2]. Signature kernels implementation is that of Sig-PDEs accompanying [21], available at https://github.com/crispitagorico/sigkernel[3]. For forecasting with Sig-PDE we fit a ridge regressor from observation time-delays and times to their successor. The prediction is then concatenated to the history and used to forecast subsequent steps. To ensure that Sig-PDE forecasts the same times in $\{0, \ldots, H\Delta t\}$ we simulate the systems backwards in time and train Sig-PDE with observations from the interval $\{-l\Delta t, \ldots, H\Delta t\}$.

[2] last accessed version `"0.1.24"` at https://github.com/csml-iit-ucl/kooplearn/tree/legacy_kooplearn from April 25, 2023

[3] last accessed version from July 25, 2023

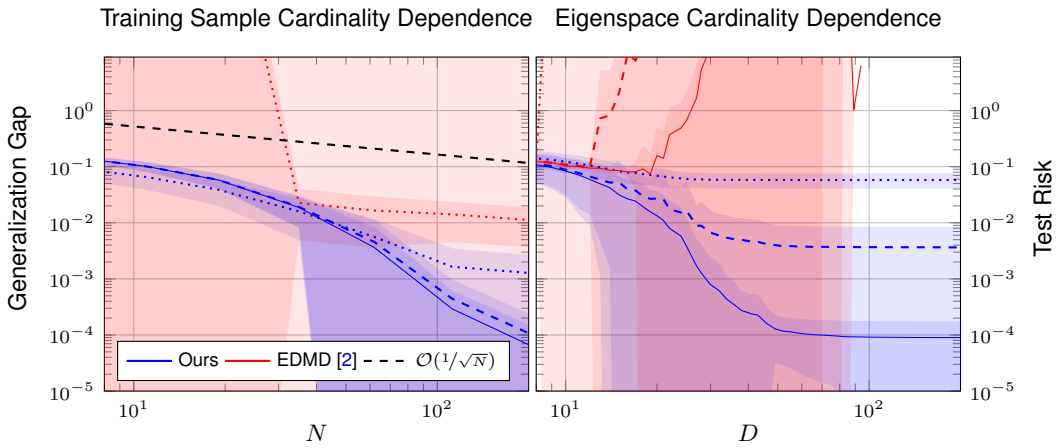

Figure 1: Forecasting risks for the bi-stable system over a time-horizon $H = 14$. **Left:** Forecast generalization gap for $D \in \{10 : \cdots, 41 : - -, 400 : —\}$ is depicted with a growing number of data points. **Right:** Test risk behavior with an increasing amount of eigenspaces is shown for $N \in \{19 : \cdots, 62 : - -, 200 : —\}$, demonstrating the benefits of KKR.

C.1 Numerical Evaluation Details

Normalizing the invariance transform We normalize the invariance transformation of each eigenvalue by the norm of its pullback $\| \mathrm{e}^{-\lambda t} \|_{\mathbb{T}} / \, \|\mu^h\|_{\mathbb{H}}$. Normalizing increases numerical stability significantly as for discrete-time eigenvalues close to the origin the pullback μ^{-k} go to infinity. Beyond mere numerical convenience, this also provides intuition on what the invariance transformation does. Consider the aforementioned case $\mu \to 0$, then the eigenfunction decays infinitesimally fast: the invariance transformation becomes an indicator at the final time $\delta_T(t)$.

Details on the bi-stable system experiment We chose $N = 50$ datapoints. For the base kernel we utilize the radial basis function (RBF) kernel $k(\boldsymbol{x}, \boldsymbol{x}') = \mathrm{e}^{\frac{1}{2\ell^2}\|\boldsymbol{x}-\boldsymbol{x}'\|^2}$ with a length scale of $\ell = 0.05$, covering the whole state space, while allowing for sufficient distinction of trajectories due the time-horizon $H = 14$ fulfilling our non-recurrence assumption. We trained models for EDMD and KKR with predictor rank D in a range from 1 to 100 and chose the best performing for each method. Unsurprisingly, KKR performs best with 100 eigenfunctions while EDMD attains its minimizer at 10.

Van der Pol oscillator experiment detail We utilize RBF kernels with a length scale of $\ell = 0.1$.

C.2 Additional Experiments

Eigenspace and sample cardinality dependence To provide more intuition on how our method, and as a baseline EDMD, performs dependent on the number of samples and eigenfunctions used, we provide parameterized versions of the experiments from the main text. Bite that the bi-stable system experiment is here run with parameters $a = 4$, $b = -16$. Figure 1 depicts these dependencies for the bi-stable system, while Figure 2 displays the same experiments for the Van der Pol oscillator. We observe that KKR admits the sane property of increased excess and test performance with increasing cardinality of eigenspaces D. It also becomes clear that, due to limited data, increase in the number of eigenfunctions has, at some point, diminished returns for the test risk of KKR. Nevertheless, additional eigenfunctions do not deteriorate the test risk, a salient feature or our approach compared to EDMD that might yield worse performance on test data – as predicted by [2].

Validation of other theoretical results Using Monte-Carlo-Integration, we verify the convergence of the discrete-time Koopman kernel in the misspecified case by Figure 3. We sample eigenvalues from the uniform distribution on the complex unit disk. We use the kernel with $D = 2 \times 10^5$ as a baseline and average the difference of the operator-valued kernel to the baseline with the Frobenius norm. Results are averaged over $N = 5$ different points over 20 (i.i.d.) runs each with time-horizon $H = 14$.

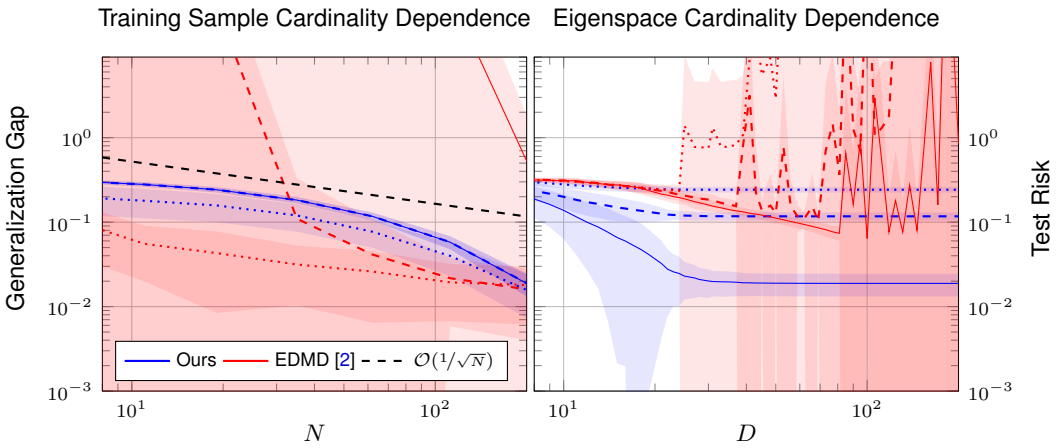

Figure 2: Forecasting risks for the Van der Pol oscillator over a time-horizon $H = 14$. **Left:** Forecast generalization gap for $D \in \{10 : \cdots, 50 : - -, 200 : —\}$ is depicted with a growing number of data points. **Right:** Test risk behavior with an increasing amount of eigenspaces is shown for $N \in \{19 : \cdots, 62 : - -, 200 : —\}$, demonstrating the benefits of KKR.

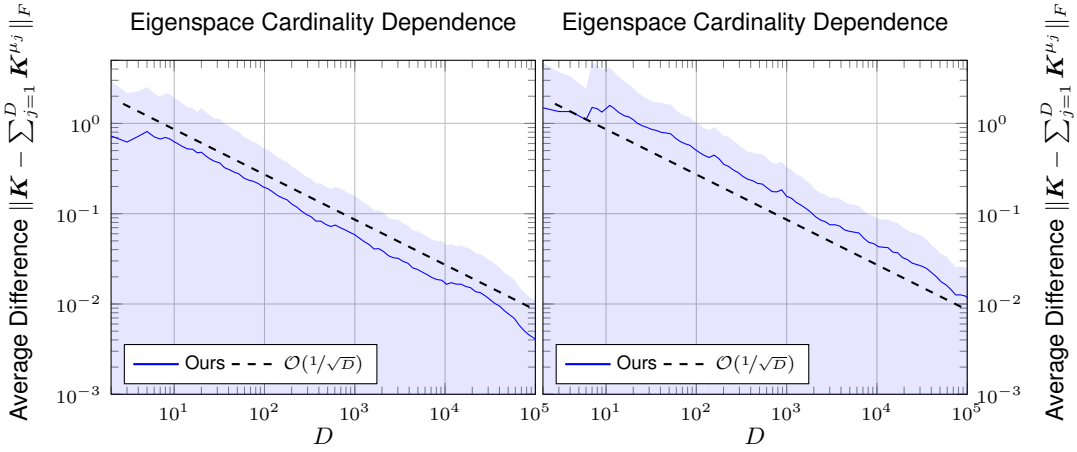

Figure 3: Norm difference of the sampled kernel to the specified kernel. **Left:** Norm difference of the kernel for the Van der Pol oscillator is depicted with a growing number of eigenvalues. **Right:** Norm difference of the kernel for the bi-stable system is depicted with a growing number of eigenvalues.

Kármán vortex street In fluid dynamics, a Kármán vortex street is a phenomenon that is observed when a laminar flow is disturbed by a solid object. We consider a cylinder. After a settling phase, the transient, periodically oscillating vortices behind the cylinder eventuate. This phenomenon occurs, for example, in the airflow behind a car or a wind turbine. Therefore, predicting the effect of vortex streets on velocity fields is highly relevant for engineers in the aero- and hydro-dynamic design of systems since the frequency of oscillation might cause undesirable resonance. Fluid dynamics simulations solving some variation of the Navier-Stokes equations, usually by discretizing space into a grid, are employed to predict the aforementioned effects. However, integrating these simulations in complex multi-physics simulations is challenging due to their relatively high computational complexity – making fluid simulation a bottleneck. Thus, surrogate modelling of the effect of interest through a faster-to-evaluate model is of great interest. Nevertheless, as the states of a fluid simulation are usually velocities or other quantities at each grid point, the data available to train surrogate models is high-dimensional and, thus, often challenging to handle.

To demonstrate that our method is capable of performing well with high dimensional data in the context described above, we employ it to obtain a simplified representation – an *LTI predictor* – of the measurements of a sensor in a Kármán vortex street under variation of the initial condition. The

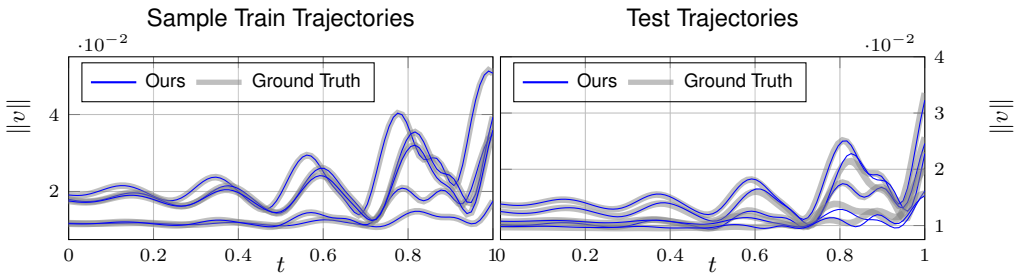

Figure 4: Observable trajectories of the simulated cylinder flow and the surrogate model **Left:** Samples from the training data are depicted. **Right:** The test data is depicted.

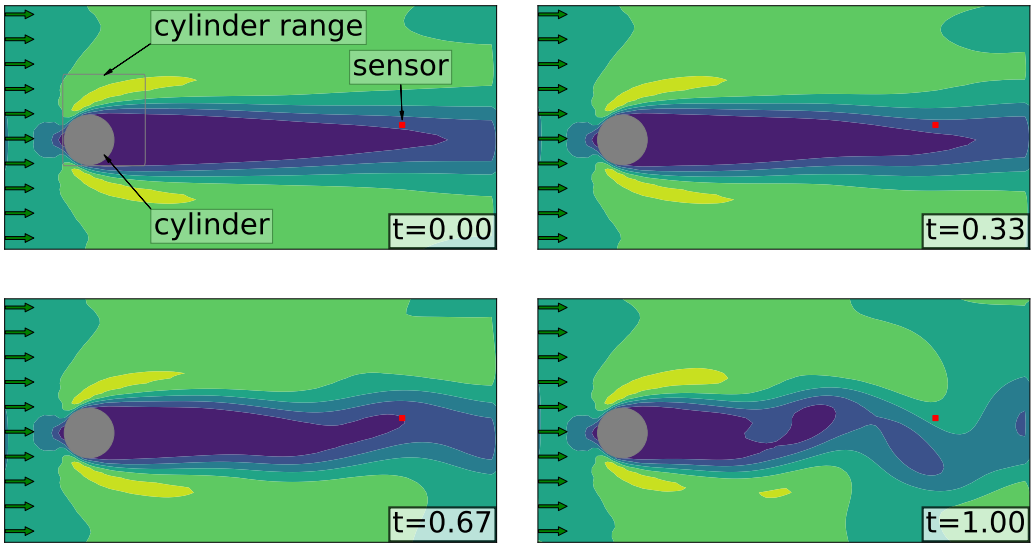

Figure 5: Velocity magnitudes in a developing Kármán vortex street behind a cylinder at different times. Yellow color indicates high and blue low magnitude.

variation is a deviation in the cylinder placement. The setup is depicted in Figure 5. To obtain the ground truth, we employ a solver based on the Lattice-Boltzmann Method [22] from an MIT-licensed implementation available at https://github.com/Ceyron/machine-learning-and-simulation/tree/main/english/simulation_scripts. We specify a Reynolds number of 40, a 100×50 grid and an inlet velocity at $(0, y)$ of $0.05^{m}/_{s}$ in x-direction. The cylinder position is varied by up to three grid points in each direction around $(20, 25)$, amounting to 49 different initial conditions, for which sample trajectories are computed. We randomly split those into 44 training and five testing samples. Simulation yields our state – the velocity magnitudes at each grid point $d = 100 \times 50 = 5000$ – over horizon length $H = 99$. Therefore, a trajectory can be interpreted as a sequence of images. A sample trajectory can be found next to this document in the supplemental. We place a virtual sensor at (80, 25), such that the corresponding velocity magnitude is our observable. Using the knowledge that the Kármán vortex street admits stable periodic behaviour, we select Koopman operator eigenvalues λ that are purely imaginary, for the stable periodic manifold, or purely decaying, for the transient regime [5, 23]: $\mu = \mathrm{e}^{\lambda \Delta t}$, where $\lambda \sim \rho_\lambda = \mathrm{uniform}\left(\{\pm aj, -a | 0 \leq a \leq 1\}\right)$. We fit a KKR model with $D = 500$ and an RBF base kernel with length scale $\ell = 30$. The model enables us to forecast the observable using an image of the velocity magnitudes – a 5000 dimensional vector – as input. In Figure 4, our model's prediction is compared to ground truth. We observe that training trajectories are accurately reconstructed, with good performance on test data, despite the low number of training samples $N = 45$. Notably, reproducing the dataset using KKR takes ≈ 0.05 seconds (average over 1000 calls), while simulating the ground tooth takes ≈ 1 second per run (average over 49 runs), both using one GPU unit – demonstrating suitability for surrogate models.

References

[1] H. Kreidler, "Compact operator semigroups applied to dynamical systems," *Semigroup Forum*, vol. 97, no. 3, pp. 523–547, 12 2018.

[2] 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*, 2022, pp. 4017–4031.

[3] V. Kostic, K. Lounici, P. Novelli, and M. Pontil, "Koopman Operator Learning: Sharp Spectral Rates and Spurious Eigenvalues," 2 2023. [Online]. Available: http://arxiv.org/abs/2302.02004

[4] M. Budišić, R. Mohr, and I. Mezić, "Applied Koopmanism," *Chaos*, vol. 22, no. 4, 10 2012.

[5] I. Mezić, "Spectrum of the Koopman Operator, Spectral Expansions in Functional Spaces, and State-Space Geometry," *Journal of Nonlinear Science*, vol. 30, no. 5, pp. 2091–2145, 2020.

[6] M. Korda and I. Mezic, "Optimal Construction of Koopman Eigenfunctions for Prediction and Control," *IEEE Transactions on Automatic Control*, vol. 65, no. 12, pp. 5114–5129, 12 2020.

[7] K. Küster, "The Koopman Linearization of Dynamical Systems," 2015. [Online]. Available: https://homepages.laas.fr/henrion/mfo16/kari-kuester.pdf

[8] M. Ikeda, I. Ishikawa, and C. Schlosser, "Koopman and Perron–Frobenius operators on reproducing kernel Banach spaces," *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 32, no. 12, p. 123143, 12 2022.

[9] N. Dunford, "Spectral Theory. I Convergence to Projections," *Transactions of the American Mathematical Society*, vol. 54, no. 2, p. 185, 9 1943.

[10] B. Haasdonk and H. Burkhardt, "Invariant kernel functions for pattern analysis and machine learning," *Machine Learning*, vol. 68, no. 1, pp. 35–61, 7 2007.

[11] L. Nachbin, *The Haar integral*. Huntington, N.Y: R. E. Krieger Pub. Co, 1976.

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

[13] D. Werner, *Funktionalanalysis*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2018.

[14] J. Mercer, "Functions of positive and negative type, and their connection the theory of integral equations," *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, vol. 209, no. 441-458, pp. 415–446, 1 1909.

[15] A. Lederer, A. Capone, T. Beckers, J. Umlauft, and S. Hirche, "The Impact of Data on the Stability of Learning-Based Control," in *Proceedings of the 3rd Conference on Learning for Dynamics and Control*, 2021, pp. 623–635.

[16] M. A. Álvarez, L. Rosasco, and N. D. Lawrence, "Kernels for vector-valued functions: A review," pp. 195–266, 2011.

[17] N. Aronszajn, "Theory of Reproducing Kernels," *Transactions of the American Mathematical Society*, vol. 68, no. 3, p. 337, 1950.

[18] C. Carmeli, E. De Vito, and A. Toigo, "Vector Valued Reproducing Kernel Hilbert Spaces of Integrable Functions and Mercer Theorem," *Analysis and Applications*, vol. 4, no. 4, pp. 377–408, 2006.

[19] A. Caponnetto, C. A. Micchelli, M. Pontil, and Y. Ying, "Universal Multi-Task Kernels," *Journal of Machine Learning Research*, vol. 9, no. 52, pp. 1615–1646, 2008.

[20] P. L. Bartlett and S. Mendelson, "Rademacher and Gaussian Complexities: Risk Bounds and Structural Results," *Journal of Machine Learning Research*, vol. 3, pp. 463–482, 3 2002.

[21] C. Salvi, T. Cass, J. Foster, T. Lyons, and W. Yang, "The Signature Kernel Is the Solution of a Goursat PDE," *SIAM Journal on Mathematics of Data Science*, vol. 3, no. 3, pp. 873–899, 1 2021.

[22] T. Krüger, H. Kusumaatmaja, A. Kuzmin, O. Shardt, G. Silva, and E. M. Viggen, *The Lattice Boltzmann Method*. Springer International Publishing, 2017.

[23] A. Mauroy, I. Mezić, and Y. Susuki, *The Koopman Operator in Systems and Control*, ser. Lecture Notes in Control and Information Sciences. Cham: Springer International Publishing, 2020, vol. 484.