Noname manuscript No.
(will be inserted by the editor)

Discrete-Time Dynamical Systems Characterization via Invariance and Approximation of Koopman Operators and Operator-Valued Kernels

Soonyong Yang · Andrew J. Kurdila · Andrea L'Afflitto · Rushikesh Kamalapurkar · Sai Tej Paruchuri · Joel Rosenfeld · Haoran Wang

Received: date / Accepted: date

Abstract This paper presents approaches based on native space theory and Koopman operators to characterize the dynamics of nonlinear, discrete-time dynamical models employing measured data only. Given approximation schemes of the plant dynamics based on Koopman operators contained in vector-valued reproducing kernel Hilbert spaces (vRKHSs), we deduce rates of convergence for these schemes. In particular, we present a necessary and sufficient condition for Koopman invariance of observables in vRKHSs that are defined via generic non-diagonal operator-valued kernels, and develop sufficient conditions to guarantee the Koopman invariance for vRKHSs defined in terms of a class of diagonal operator-valued kernels. Principles of inverse problems are leveraged to derive error bounds for approximations of the Koopman operator that include both a deterministic sampling error and an approximation error term.

S. Yang · A.J. Kurdila · H. Wang
Department of Mechanical Engineering, Virginia Tech, Blacksburg, VA 24061, USA
E-mail: soonyong@vt.edu, kurdila@vt.edu, haoran9@vt.edu

A. L'Afflitto
Grado Department of Industrial and Systems Engineering, Virginia Tech, Blacksburg, VA 24060, USA
E-mail: a.lafflitto@vt.edu

R. Kamalapurkar
Department of Mechanical and Aerospace Engineering, University of Florida, Gainesville, FL 32611, USA
E-mail: rkamalapurkar@ufl.edu

S.T. Paruchuri
Department of Mechanical Engineering and Mechanics, Lehigh University, Bethlehem, PA 18015, USA
E-mail: saitejp@lehigh.edu

J. Rosenfeld
Department of Mathematics and Statistics, University of South Florida, Tampa, FL 33620, USA
E-mail: rosenfeldj@usf.edu

The deterministic sampling error arises since imprecisely measured samples are used to approximate the Koopman operator. This work is the first to present overall bounds in the deterministic setting that explicitly account for the sampling error, which, in general, increases with the reduced dimension. Numerical examples illustrate the proposed results.

Keywords Koopman theory · inverse problems · Koopman invariance · approximation

Mathematics Subject Classification (2020) 46E22 · 47A58 · 47B32

1 Introduction and Motivation

This paper studies the now-classical setting in which Koopman operators arise in the analysis of deterministic discrete-time nonlinear systems. Consider the deterministic discrete-time system

$$x_{i+1} = f(x_i), \tag{1}$$

$$y_{i+1} = g(x_{i+1}) \triangleq (U_f g)(x_i), \tag{2}$$

where $i \in \mathbb{N}$ denotes the index of the time series, $x_i \in \mathbb{X} \triangleq \mathbb{R}^n$ denotes the *system's state* at the step i, $y_i \in \mathbb{Y} \triangleq \mathbb{R}^m$ denotes the *observation value*, $f : \mathbb{X} \to \mathbb{X}$ denotes the *propagation function*, $g : \mathbb{X} \to \mathbb{Y}$ denotes the *observable function*, $U_f : H \to P$ denotes a *Koopman operator*, and H and P are suitably chosen spaces of functions. The space H is referred to as the *space of observable functions*, or just the *space of observables*. We typically analyze this system in terms of a bounded set $\Omega \subset \mathbb{X}$ that is positively invariant for (1), that is, $f(\Omega) \subseteq \Omega \subset \mathbb{X}$. In the proposed analysis, the set Ω is quite generic. However, worthy of mention are the proposed numerical examples in which $\Omega \triangleq \mathcal{M} \subset \mathbb{X}$ and $\mathcal{M}$ is a compact, connected, smooth submanifold regularly embedded in $\mathbb{R}^n$.

Assuming that that the propagation function $f : \Omega \to \Omega$ is unknown and that only imprecisely measured output samples $\{(x_i, y_i^\delta)\}_{i=1}^M$ that approximate the exact samples $\{(x_i, y_i)\}_{i=1}^M \triangleq (x_i, g(x_i))\}_{i=1}^M$ along the trajectories of (1) are available, we seek to construct approximations of the Koopman operator U_f to build estimates $\hat{y}_{m+1}$ of the future output from the current state x_m at any index $m \in \mathbb{N}$. In this paper, we exploit vector-valued reproducing kernel Hilbert spaces (vRKHSs) and assume that $H = \boldsymbol{\mathcal{H}}$, where $\boldsymbol{\mathcal{H}}$ denotes a vRKHS, to construct approximations of the Koopman operator U_f. We study how $\hat{y}_{m+1}$ can be a good approximation of the true outputs y_{m+1} for any choice of the observable function $g \in \boldsymbol{\mathcal{H}}$. Assuming that the vRKHS $\boldsymbol{\mathcal{H}}$ is approximated by the vRKHS $\boldsymbol{\mathcal{H}}_N$ defined in terms of N kernel centers, we build forecasts $\hat{y}_{m+1}$ that depend on the number of samples M in the training set and the number of centers N used to construct the approximating subspace $\boldsymbol{\mathcal{H}}_N \subseteq \boldsymbol{\mathcal{H}}$. Key results of this paper are deterministic, finite sample, error bounds that depend on both N and M, which can be used to gauge the

accuracy of these forecasts. The dependency on N is expressed explicitly in terms of the fill distance or minimal separation of the centers.

The error analysis of the Koopman approximations in this paper is based on principles of inverse problems and, hence, is entirely deterministic. The noisy samples in this paper should be interpreted as the true sample values with some perturbation or error. We do not imply that the noisy samples are generated by some stochastic measurement process. However, the finite-sample bounds derived in this paper still hold if the samples are generated by a discrete stochastic process. In particular, if the samples are randomly generated consistent with the assumption that $g \circ f$ is a Gaussian process, then the final finite sample error bounds in this paper are still applicable.

2 Novelty of the Proposed Results

Methods to approximate the Koopman operator U_f have been examined in many references, such as, for instance, [2–4, 7, 9, 10, 16–18, 25, 27, 31, 32, 34, 36, 38, 43, 46, 47, 50, 56], to name a few, under a variety of hypotheses. When formulating these approximation problems, the choice of the space of observables H and the range $P = \mathcal{R}(U_f)$ can be crucial. Common choices for the space of observables H include the spaces of Lebesgue square integrable functions $L^2(\Omega, \mathbb{R})$, continuous functions $C(\Omega, \mathbb{R})$, or scalar-valued RKHSs $\mathcal{H}(\Omega, \mathbb{R})$. For these choices, it is perhaps most common to study the approximation problem with $P = L^2(\Omega, \mathbb{R})$.

This paper differs from previous works for a variety of reasons. To the authors' knowledge, the most recent existing work that studies approximation of Koopman operators acting on vRKHS generated by operator-valued kernels is stochastic in nature. That is, the Koopman operators are defined for a generally nonlinear stochastic system. For example, the authors in [37, 39, 44] derive error bounds via approximations of the kernel mean embedding (KME) technique, which is inherently a stochastic formulation. In practice, these techniques require the collection $\{(X_i, Y_i)\}_{i=1}^M$ of independently and identically distributed (IID) observations of the stationary process determining the stochastic dynamics. The resulting final error estimates yield stochastic error bounds that hold "in high probability." Errors are expressed in terms of the norm of an abstract interpolation space derived from the eigenfunctions of a covariance operator, and are rather abstract. The approach in this paper is purely deterministic, as stated in the definition of nonlinear recursion that defines the dynamics in Equations 1 and 2. Samples to be used in calculations can be generated along the path of the deterministic evolution law; they are dependent rather than IID samples. This means that it is easy to keep or delete samples collected along a deterministic trajectory when convenient, without regard to the fact that such deletions in KME would violate the requirement above for IID samples. By extending the doubling trick to the setting of vRKHS generated by general, non-diagonal operator-valued kernels, the resulting error bounds based on the power function hold without regard to whether the

samples originate from IID samples of a random process or from dependent iterates of a deterministic recursion. Moreover, the error bounds hold in the original norm of the vRKHS, for all functions that satisfy a common regularity condition popularized in approximation methods in scalar-valued RKHSs. Finally, the bounds derived in this paper are strict, deterministic, upper bounds on the worst-case error. In contrast, the bounds generated by KME only hold with high probability, and, although improbably, for some realizations of measurements, the KME approximation errors will be large.

Additional elements of novelty of the proposed results are the following. The proposed analysis explicitly uses certain pullback vRKHSs to derive rates of convergence. Only the recent efforts by the authors in [47, 49] make systematic use of pullback spaces in deriving error bounds. However, in [47, 49], the analysis is limited to scalar-valued RKHS, and not vRKHS as in this paper. The analysis via pullback spaces in this paper also includes a derivation of necessary and sufficient conditions for Koopman invariance in vRKHSs defined by operator-valued kernels and a description of a standard constructive procedure for generating Koopman-invariant spaces based on infinite direct sums of pullback spaces. Lastly, this paper derives rates of convergence using principles of inverse methods. These results enable the derivation of error bounds that are explicit in the contributions of the sampling error and the approximation error. The sampling error only arises because the sample values are assumed to be noisy in this paper. The contribution of the sampling error, which we show can play a vital role when using noisy measurements to make forecasts, was unaccounted for in the cited references on deterministic approximations of Koopman operators.

The work in this paper can be understood as a concrete means of deriving error bounds in data-driven frameworks that build on the classical approaches on inverse problems, such as those described in Section 3.4.1 of [33] or Sections 17.3 and 17.4 of [26], by introducing recently derived techniques [57, 58] based on power functions for vRKHS. The approaches introduced in this paper also provide practical generalizations of the results in [40–42], which routinely focus on the applications in a scalar-valued RKHS, ordinarily over intervals. The approaches introduced in [57, 58] provide for a more general theory and associated development of algorithms in a vRKHS setting.

It is also noteworthy that this paper extends to vRKHSs the very recent analysis in [29], which establishes sufficient conditions for Koopman invariance of certain scalar-valued RKHSs equivalent to Sobolev spaces that are generated by scalar-valued Wendland kernels. In this paper, we show that the general procedure outlined in [29] can be extended to a broader class of scalar-valued RKHSs and vRKHSs defined by operator-valued kernels. For example, we show that the sufficient conditions in [29] also guarantee that diagonal operator-valued kernels defined in terms of scalar-valued Wendland, Sobolev-Matérn, Poisson, or Abel kernels are Koopman-invariant. Notably, if applied to diagonal operator-valued kernels, the proposed results are a relatively direct extension of the results in [29]. However, in the general context of non-diagonal operator-valued kernels, the proposed results have no counterpart in [29]. Ad-

ditional key differences with [29] are the following. Theorems 5.1 and 5.2 and Corollary 5.4 of [29] rely on problem data that consists of *exact sample values* of the function at various centers. These results can serve as starting points to address the question of how the approximation of Koopman operators depends on perturbations to the problem data, which is addressed in this paper. The description of the sensitivity to noise of approximations of the Koopman operator is one of the central concepts driving the analysis in this paper. We show how the contributions of the sample error and approximation error affect each other. This result is attained by resorting to the theory of inverse methods. In essence, the bounds derived in Theorems 5.1 and 5.2 and Corollary 5.4 of [29] only correspond, or only could be used, to assess the approximation error. In [29], however, there is no discussion on the sample error due to imprecision in the problem data, which is addressed in this paper.

3 Notation

In this paper, $\mathbb{N}$ denotes the *set of positive integers*, $\mathbb{N}_0$ denotes the *set of nonnegative integers*, $\mathbb{R}$ denotes the *set of real numbers*, $\mathbb{R}^n$ denotes the *set of n-valued real vectors*, and $\mathbb{R}^{m\times n}$ denotes the *set of m-by-n-valued matrices*. We write $\|\cdot\|_2$ and $\|\cdot\|_\infty$ for the classical norms on the normed vector spaces $(\mathbb{R}^n, \|\cdot\|_{\ell^2})$ and $(\mathbb{R}^n, \|\cdot\|_{\ell^\infty})$, respectively. The *inner product* over $\mathbb{R}^n$ is denoted by $\langle\cdot,\cdot\rangle_{\mathbb{R}^n}$ and is such that $\langle v, v\rangle_{\mathbb{R}^n} = \|v\|_2^2$ for any $v \in \mathbb{R}^n$. For $1 \le p, q \le \infty$, the (p,q)*-induced norm on* $\mathbb{R}^{m\times n}$ is denoted by $\|\cdot\|_{p,q}$. In general, blackboard bold fonts, such as $\mathbb{X}$ or $\mathbb{Y}$, represent generic vector spaces equipped with some consistent norm. The generic *inner product on the vector space* $\mathbb{Y}$ is denoted by $\langle\cdot,\cdot,\rangle_{\mathbb{Y}}$.

We denote the fact that two norms $\|\cdot\|_a, \|\cdot\|_b$ on the vector space $\mathbb{V}$ are *equivalent* by writing $\|\cdot\|_a \approx \|\cdot\|_b$, which means that there are two constants $C_1, C_2 > 0$ such that $C_1\|v\|_a \le \|v\|_b \le C_2\|v\|_a$ for all $v \in \mathbb{V}$. Also, this paper often must carefully distinguish between function spaces that contain functions that take values in $\mathbb{R}$ and those that take values in $\mathbb{R}^m$ for some $m > 1$. We refer to the former as *scalar-valued function spaces* and the latter as *vector-valued function spaces*.

For two Banach spaces U, V, we denote by $\mathcal{L}(U, V)$ the *Banach space of bounded linear operators* T equipped with the usual induced operator norm $\|T\| \triangleq \sup_{u\in U\setminus\{0\}} \|Tu\|_V/\|u\|_U$. If $U = V$, then we write $\mathcal{L}(U)$ for $\mathcal{L}(U, V)$. The *range space* and *nullspace* of T are denoted by $\mathcal{R}(T)$ and $\mathcal{N}(T)$, respectively. For normed vector spaces U, V, we say that U is *continuously embedded* in V, that is, $U \overset{\mathcal{I}}{\hookrightarrow} V$, if $U \subseteq V$ and the canonical injection $\mathcal{I} : U \to V$ that satisfies $\mathcal{I} : u \in U \mapsto \mathcal{I}(u) = u \in V$ is bounded and linear.

We recall that if U, V are normed vector spaces, the mapping $T : U \to V$ is an *isometry* if $\|Tu\|_V = \|u\|_U$ for all $u \in U$. The mapping T is a *partial isometry* if there is a closed subspace $U_I \subset U$ such that $T|_{U_I} : U_I \to V$ is an isometry and $T|_{U_I^\perp} = 0$. The space U_I is called the *initial space*, and the range

$T(U_I) = V_F \subseteq V$ is called the *final space* of the partial isometry; see [54, Sec. 4.6] for a discussion on the properties of partial isometries.

Several standard function spaces are used in this paper. The symbol $C_b \triangleq C_b(\Omega, \mathbb{R}^m)$ denotes the *collection of bounded continuous functions* over the subset $\Omega \subseteq \mathbb{X}$ that take values in $\mathbb{R}^m$, which is equipped with the usual uniform norm

$$\|f\|_{C_b} \triangleq \sup_{x\in\Omega} \|f(x)\|_2 \approx \sup_{x\in\Omega} \|f(x)\|_\infty. \tag{3}$$

For each integer multi-index $\alpha \triangleq [\alpha_1, \ldots, \alpha_n]^{\mathrm{T}} \in \mathbb{N}_0^n$ and each function $f : \Omega \to \mathbb{R}^m$ with $f = [f_1, \ldots, f_m]^{\mathrm{T}}$, we define the *partial derivative* $D^\alpha f \triangleq \{D^\alpha f_1, \ldots, D^\alpha f_m\}$ entry-wise as

$$(D^\alpha f)_i \triangleq D^\alpha f_i \triangleq \frac{\partial^{|\alpha|} f_i}{\partial x^{\alpha_1} \cdots x^{\alpha_n}}, \tag{4}$$

where $|\alpha| = \sum_{i=1}^n \alpha_i$ denotes the *length of the integer multi-index*. For any $k \in \mathbb{N}_0$, we denote by $C_b^k \triangleq C_b^k(\Omega, \mathbb{Y})$ the set of k-times continuously differentiable functions from $\Omega \subseteq \mathbb{R}^n$ to $\mathbb{R}^m$ with norm

$$\|f\|_{C_b^k} \triangleq \max_{0\le|\alpha|\le m} \|D^\alpha f\|_{C_b}. \tag{5}$$

The *Lebesgue space* $L^2(\Omega, \mathbb{R}^m)$ is the collection of square integrable functions with norm

$$\|f\|_{L^2(\Omega,\mathbb{R}^m)} \triangleq \sqrt{\int_\Omega \|f(x)\|_2^2 \mathrm{d}x}. \tag{6}$$

We denote the $\mathbb{R}^m$-valued *Sobolev spaces over* $\mathbb{R}^n$ *with smoothness index* $s > 0$ by $\boldsymbol{\mathcal{W}}^s(\mathbb{R}^n) \triangleq \boldsymbol{\mathcal{W}}^s(\mathbb{R}^n, \mathbb{R}^m)$. This space consist of those functions in $\boldsymbol{f} \in L^2(\mathbb{R}^n, \mathbb{R}^m)$ with bounded norm, that is, such that $\|\boldsymbol{f}\|_{\boldsymbol{\mathcal{W}}^s(\mathbb{R}^n,\mathbb{R}^m)} < \infty$, where the norm is defined in terms of the inner product

$$\langle f, g\rangle_{\boldsymbol{\mathcal{W}}^s(\mathbb{R}^n,\mathbb{R}^m)} \triangleq \frac{1}{(2\pi)^{n/2}} \int_{\mathbb{R}^n} \left(1 + \|\omega\|_2^2\right)^s \hat{\boldsymbol{f}}(\omega) \cdot \overline{\hat{\boldsymbol{g}}(\omega)} \mathrm{d}\omega \tag{7}$$

for any $\boldsymbol{f}, \boldsymbol{g} \in \boldsymbol{\mathcal{W}}^s(\mathbb{R}^n, \mathbb{R}^m)$, and $\hat{(\cdot)}$ denotes the *Fourier transform* of its argument. With this norm, note that $\boldsymbol{\mathcal{W}}^s(\mathbb{R}^n, \mathbb{R}^m) = (\mathcal{W}^s(\mathbb{R}^n))^m \triangleq \mathcal{W}^s(\mathbb{R}^n) \times \cdots \times \mathcal{W}^s(\mathbb{R}^n)$, where $\mathcal{W}^s(\mathbb{R}^n) \triangleq \mathcal{W}(\mathbb{R}^n, \mathbb{R})$ denotes the scalar-valued Sobolev space with positive smoothness indexes $s > 0$.

We also use the Sobolev spaces $\mathcal{W}^s(\Omega, \mathbb{R})$ and $\boldsymbol{\mathcal{W}}^s(\Omega, \mathbb{R}^m)$ defined over a bounded domain $\Omega \subset \mathbb{R}^n$ that has a Lipschitz boundary. These spaces are defined either by restricting the integral in (7) to Ω or by restricting functions in $\boldsymbol{\mathcal{W}}^s(\mathbb{R}^n, \mathbb{R}^m)$. Both approaches yield equivalent definitions when Ω is an open connected set having a Lipschitz boundary; see [1]. Finally, we use the Sobolev spaces $\boldsymbol{\mathcal{W}}^s(\mathcal{M}, \mathbb{R})$ and $\boldsymbol{\mathcal{W}}^s(\mathcal{M}, \mathbb{R}^m)$, which are defined, for example, in [14, 53].

4 Elements of Native Space Theory

This paper makes use of several basic properties of vRKHSs found in [35, Ch. 3] and [5, 45] that are defined in terms of generally non-diagonal operator-valued kernels over a bounded subset $\Omega \subset \mathbb{X}$. In particular, the notions of symmetric, positive, strictly positive, and admissible operator-valued kernels $\mathcal{K} : \Omega \times \Omega \to \mathcal{L}(\mathbb{Y})$ can be found in Definitions 3.15–3.17 of [35], respectively. It is worthwhile recalling the *generalized Grammian matrix*

$$\mathbb{K}_N \triangleq \begin{bmatrix} \mathcal{K}(\xi_1, \xi_1) & \cdots & \mathcal{K}(\xi_1, \xi_N) \\ \vdots & \ddots & \vdots \\ \mathcal{K}(\xi_N, \xi_1) & \cdots & \mathcal{K}(\xi_N, \xi_N) \end{bmatrix} \in \mathbb{R}^{mN \times mN} \tag{8}$$

generated by the *set of centers* $\Xi_N \triangleq \{\xi_i\}_{i=1}^N \subset \Omega$. The operator-valued kernel $\mathcal{K}$ is of *strictly positive kind* if the associated generalized Grammian matrix $\mathbb{K}_N$ is positive-definite for any set of distinct centers Ξ_N. For brevity, in this paper, we assume that the operator-valued kernel $\mathcal{K}$ is of strictly positive kind. Results under weaker assumptions can be produced by proceeding similarly.

These positivity conditions on the Grammian matrix can be used to define a partial ordering on operator kernels. For two symmetric positive semidefinite matrices $A, B \in \mathbb{R}^{n \times n}$, $B \leq A$, per definition, means that $\langle (A - B)x, x \rangle_{\mathbb{R}^n} \geq 0$ for all $x \in \mathbb{R}^n$. Similarly, if $\mathcal{K}$ and $\mathcal{P}$ are two admissible operator kernels that map $\Omega \times \Omega$ into $\mathcal{L}(\mathbb{Y})$, per definition, $\mathcal{P} \preceq \mathcal{K}$ means that the underlying Grammian matrices, $\mathbb{P}_N$ and $\mathbb{K}_N$, respectively, are such that $\mathbb{P}_N \leq \mathbb{K}_N$ for every collection of N centers $\{\xi_i\}_{i=1}^N \subset \Omega$.

Given the set of centers $\Xi_N \subset \Omega$, we define the *minimal separation radius*

$$r_{\Xi_N} \triangleq \frac{1}{2} \min_{\xi_i, \xi_j \in \Xi_N, i \neq j} \|\xi_i - \xi_j\|_{\mathbb{X}} \tag{9}$$

and the *fill distance*

$$h_{\Xi_N, \Omega} \triangleq \sup_{x \in \Omega} \min_{\xi_i \in \Xi_N} \|x - \xi_i\|_{\mathbb{X}} \tag{10}$$

of the centers Ξ_N in the set Ω. The fill distance is a better measure of the uniformity of separation among the centers. If there exist $C_1, C_2 > 0$ such that

$$C_1 r_{\Xi_N} \leq h_{\Xi_N, \Omega} \leq C_2 r_{\Xi_N} \qquad \text{for all } N \in \mathbb{N}, \tag{11}$$

then, per definition, the sample distribution is *quasi-uniform* in Ω. If (11) is verified for all $N > 0$ sufficiently large, then, per definition, the minimal separation distance is *asymptotically equivalent* to the fill distance $h_{\Xi_N, \Omega}$.

As shown in [35, Th. 3.18], an admissible operator kernel $\mathcal{K} : \Omega \times \Omega \to \mathcal{L}(\mathbb{Y})$ can be employed to construct a vRKHS $\boldsymbol{\mathcal{H}}(\Omega, \mathbb{Y})$ of functions over Ω that take values in $\mathbb{Y}$; if clear from the context, in the following, we omit the domain and the codomain underlying a vRKHS and write, for instance, $\boldsymbol{\mathcal{H}}$ for $\boldsymbol{\mathcal{H}}(\Omega, \mathbb{Y})$. Given $\mathcal{K}$, we define $\mathcal{K}_x \triangleq \mathcal{K}(\cdot, x)$ for all $x \in \Omega$ and the finite span $\mathring{\boldsymbol{\mathcal{H}}} \triangleq$

span $\{\mathcal{K}_x y : x \in \Omega, y \in \mathbb{Y}\}$, which consists of the finite linear combinations of terms in the form $\mathcal{K}_x y$ for some $x \in \Omega$ and some $y \in \mathbb{Y}$. By definition, the *norm* of $g_N \triangleq \sum_{i=1}^{N} \alpha_{N,i} \mathcal{K}_{x_{N,i}} y_{N,i} \in \mathring{\boldsymbol{\mathcal{H}}}$, where $\{x_{N,i}\}_{i=1}^{N} \subset \Omega$, $\{y_{N,i}\}_{i=1}^{N} \subset \mathbb{Y}$, and $\{\alpha_{N,i}\}_{i=1}^{N} \subset \mathbb{R}$, is defined as

$$\|g_N\|_{\boldsymbol{\mathcal{H}}} \triangleq \sqrt{\sum_{i,j=1}^{N} \alpha_{N,i} \alpha_{N,j} \left\langle \mathcal{K}(x_{N,i}, x_{N,j}) y_{N,j}, y_{N,i} \right\rangle_{\mathbb{Y}}}. \tag{12}$$

Thus, we define the vRKHS

$$\boldsymbol{\mathcal{H}}(\Omega, \mathbb{Y}) \triangleq \overline{\mathring{\boldsymbol{\mathcal{H}}}}(\Omega, \mathbb{Y}) = \overline{\operatorname{span}\{\mathcal{K}_x y : x \in \Omega, y \in \mathbb{Y}\}}, \tag{13}$$

where the closure is taken with respect to the norm $\|\cdot\|_{\boldsymbol{\mathcal{H}}}$. Hence, $g \in \boldsymbol{\mathcal{H}}$ if and only if there exist a sequence of centers $\{x_{N,i}\}_{i=1}^{N} \subset \mathbb{X}$, directions $\{y_{N,i}\}_{i=1}^{N} \in \mathbb{Y}$, and real coefficients $\{\alpha_{N,i}\}_{i=1}^{N} \in \mathbb{R}$ such that

$$\lim_{N \to \infty} \left\| g - \sum_{i=1}^{N} \alpha_{N,i} \mathcal{K}_{x_{N,i}} y_{N,i} \right\|_{\boldsymbol{\mathcal{H}}} = 0. \tag{14}$$

This construction ensures that the vRKHS $\boldsymbol{\mathcal{H}}$ satisfies the *reproducing property* [35, Def. 3.13]

$$\langle g(x), y \rangle_{\mathbb{Y}} = \langle E_x g, y \rangle_{\mathbb{Y}} = \langle g, \mathcal{K}_x y \rangle_{\boldsymbol{\mathcal{H}}} \quad \text{for all } (x, y) \in \Omega \times \mathbb{Y}, \tag{15}$$

where, for any $g \in \boldsymbol{\mathcal{H}}$, the mapping $E_x : g \mapsto g(x)$ denotes the *evaluation operator* at $x \in \Omega$ [35, Def. 3.13]. The reproducing property implies that $E_x^* = (E_x)^* = \mathcal{K}_x$ for all $x \in \Omega$.

This paper makes use of a few important spaces that are defined in terms of a subset $S \subset \Omega$, using the kernel $\mathcal{K} : \Omega \times \Omega \to \mathcal{L}(\mathbb{Y})$ that is defined on all of Ω. We define the closed subspace $\boldsymbol{\mathcal{H}}_S \subset \boldsymbol{\mathcal{H}}$ that is *generated by the subset* $S \subset \Omega$ as

$$\boldsymbol{\mathcal{H}}_S(\Omega, \mathbb{Y}) \triangleq \overline{\operatorname{span}\{\mathcal{K}_x y : x \in S, y \in \mathbb{Y}\}}, \tag{16}$$

where the closure is taken with respect to the norm that $\boldsymbol{\mathcal{H}}_S$ inherits as a subset of $\boldsymbol{\mathcal{H}}$; for details, see [35, Sec. 3.6.2]. We also define the vector space

$$\boldsymbol{\mathcal{Z}}_S \triangleq \{g \in \boldsymbol{\mathcal{H}} : g(x) = 0 \text{ for all } x \in S\}. \tag{17}$$

We emphasize that both $\boldsymbol{\mathcal{H}}_S$ and $\boldsymbol{\mathcal{Z}}_S$ contain functions defined on all of Ω, which are not functions restricted to $S \subseteq \Omega$.

In this paper, we assume that any operator-valued kernel $\mathcal{K}$ is bounded on the diagonal [35, Def. 3.18], that is,

$$\|\mathcal{K}(x, x)\|_{\mathcal{L}(\mathbb{Y})} \leq \bar{\mathfrak{K}} \qquad \text{for all } x \in \Omega \tag{18}$$

and for some $\bar{\mathfrak{K}} > 0$. Lastly, we assume that the mapping $(x, y) \mapsto \|\mathcal{K}(x, y)\|_{\mathcal{L}(\mathbb{Y})}$ is continuous.

5 Koopman-Invariant Spaces

5.1 Some Properties of Pullback vRKHSs

In this paper, we choose the space of observables H as the vRKHS $\boldsymbol{\mathcal{H}}(\Omega, \mathbb{Y})$ generated by the symmetric admissible operator-valued kernel $\mathcal{K} : \Omega \times \Omega \to \mathcal{L}(\mathbb{Y})$. Then, we consider the *pullback kernel*

$$\mathcal{P}(x_1, x_2) \triangleq \mathcal{K}(f(x_1), f(x_2)) \quad \text{for all } x_1, x_2 \in \Omega, \tag{19}$$

and define the *pullback space* as the vRKHS

$$\boldsymbol{\mathcal{P}} \triangleq \mathring{\boldsymbol{\mathcal{P}}} = \overline{\text{span}\{\mathcal{P}_x y : x \in \Omega, y \in \mathbb{Y}\}}, \tag{20}$$

where $\mathcal{P}_x \triangleq \mathcal{P}(\cdot, x)$ and the norm $\|\cdot\|_{\boldsymbol{\mathcal{P}}}$ over $\mathring{\boldsymbol{\mathcal{P}}}$ is defined by proceeding as in (12)-(14) for the generic RKHS $\boldsymbol{\mathcal{H}}$.

Since $\mathcal{K}$ is a symmetric admissible operator kernel, $\mathcal{P} : \Omega \times \Omega \to \mathcal{L}(\mathbb{Y})$ is a symmetric admissible operator kernel. The generalized Grammian matrix $\mathbb{P}_N$ underlying $\mathcal{P}$ is positive semidefinite for all choices of centers in Ξ_N.

Remark 1 Although $\mathcal{K}$ is of strictly positive type, and, hence, $\mathbb{K}_N$ is symmetric positive definite for any set of distinct centers Ξ_N, in general, the kernel $\mathcal{P}$ induces a generalized Grammian $\mathbb{P}_N$ that is only positive semidefinite.

The properties of pullback spaces associated with scalar-valued RKHS have been studied systematically in numerous standard references, such as, for instance, [45, Sec. 5.4.1] or [51, Sec. 2.2.2]. The next theorem extends the analysis in [5, Prop. 7] to the vector-valued case. This result is obtained by introducing the mapping $\Gamma_f : \mathcal{P}_x y \mapsto \mathcal{K}_{f(x)} y$ and showing that $\Gamma_f = U_f^*$. This result generalizes the approach used in the proof of Theorem 5.7 of [45] for the scalar-valued case. The following theorem plays an essential role in the analysis of the noise-free and noisy Koopman operator equations studied later in the paper.

Theorem 1 *Suppose that $f : \Omega \to \Omega$ is continuous. Let $\mathcal{K}(\cdot, \cdot) : \Omega \times \Omega \to \mathcal{L}(\mathbb{Y})$ be a continuous admissible operator kernel that generates the vRKHS $\boldsymbol{\mathcal{H}}(\Omega, \mathbb{Y})$, and let $\mathcal{P} : \Omega \times \Omega \to \mathcal{L}(\mathbb{Y})$ be the associated pullback kernel that defines the pullback RKHS $\boldsymbol{\mathcal{P}}(\Omega, \mathbb{Y})$. Then,*

1. *It holds that*

$$\boldsymbol{\mathcal{P}} = \mathcal{R}(U_f) \triangleq \{U_f g : g \in \boldsymbol{\mathcal{H}}\}, \tag{21}$$

and

$$\|p\|_{\boldsymbol{\mathcal{P}}} = \inf\{\|g\|_{\boldsymbol{\mathcal{H}}} : p = U_f g\} \quad \textit{for all } p \in \boldsymbol{\mathcal{P}}. \tag{22}$$

2. *The Koopman operator $U_f \in \mathcal{L}(\boldsymbol{\mathcal{H}}, \boldsymbol{\mathcal{P}})$ is bounded with $\|U_f\|_{\mathcal{L}(\boldsymbol{\mathcal{H}}, \boldsymbol{\mathcal{P}})} = 1$, and $\mathcal{N}(U_f) = \boldsymbol{\mathcal{Z}}_{f(\Omega)}$. Furthermore, $\boldsymbol{\mathcal{H}} = \boldsymbol{\mathcal{H}}_{f(\Omega)} \oplus \boldsymbol{\mathcal{Z}}_{f(\Omega)}$, where $\oplus$ denotes the* direct sum *and*

$$\boldsymbol{\mathcal{H}}_{f(\Omega)} \triangleq \overline{\text{span}\{\mathcal{K}_{f(x)} y : x \in \Omega, y \in \mathbb{Y}\}}. \tag{23}$$

3. *The Koopman operator $U_f : \boldsymbol{\mathcal{H}} \to \boldsymbol{\mathcal{P}}$ is a partial isometry with initial space $\boldsymbol{\mathcal{H}}_{f(\Omega)}$ and final space $\boldsymbol{\mathcal{P}}$. The adjoint $U_f^* : \boldsymbol{\mathcal{P}} \to \boldsymbol{\mathcal{H}}$ of U_f is a partial isometry with initial space $\boldsymbol{\mathcal{P}}$ and final space $\boldsymbol{\mathcal{H}}_{f(\Omega)}$.*

Proof: For brevity, we only outline the proof of the main steps here. The proofs of points 1) and 2) follow from using [5, Prop. 1, 7] in light of the separability of $\boldsymbol{\mathcal{H}}$. Indeed, $\mathbb{Y} \triangleq \mathbb{R}^m$ is finite-dimensional, and, hence, separable, and continuously reproducing kernels over compact sets generate RKHSs that are separable. The fact that $U_f : \boldsymbol{\mathcal{H}} \to \boldsymbol{\mathcal{P}}$ is a partial isometry also follows from [5, Prop. 1, 7]. In the following we prove the remainder of Point 3), which can be interpreted as the extension of the approach for scalar-valued kernels in Theorem 5.7 of [45] to the vRKHS setting. let $\boldsymbol{\mathcal{P}}$ and $\|\cdot\|_{\boldsymbol{\mathcal{P}}}$ be given as in (21) and (22), respectively, and consider the finite linear spans

$$\mathring{\boldsymbol{\mathcal{H}}} \triangleq \operatorname{span}\left\{\mathcal{K}_x y : x \in \Omega, y \in \mathbb{Y}\right\}, \tag{24}$$

$$\mathring{\boldsymbol{\mathcal{P}}} \triangleq \operatorname{span}\left\{\mathcal{P}_x y : x \in \Omega, y \in \mathbb{Y}\right\}. \tag{25}$$

For any $f \in \mathring{\boldsymbol{\mathcal{H}}}$ and for any $p_N = \sum_{k=1}^N \alpha_k \mathcal{P}_{x_k} y_k$, where $\{x_k\}_{k=1}^N \subset \Omega$, $\{y_k\}_{k=1}^N \subset \mathbb{Y}$, and $\{\alpha_k\}_{k=1}^N \subset \mathbb{R}$, define the operator $\mathit{\Gamma}_f : \mathring{\boldsymbol{\mathcal{P}}} \to \mathring{\boldsymbol{\mathcal{H}}}$ such that

$$\mathit{\Gamma}_f \left(\sum_{k=1}^N \alpha_i \mathcal{P}_{x_k} y_k \right) \triangleq \sum_{k=1}^N \alpha_k \mathcal{K}_{f(x_k)} y_k; \tag{26}$$

note that $\mathit{\Gamma}_f : \mathcal{P}_x y \mapsto \mathcal{K}_{f(x)} y$. The linear operator $\mathit{\Gamma}_f$ is an isometry since

$$\begin{aligned} \|\mathit{\Gamma}_f p_N\|_{\boldsymbol{\mathcal{H}}}^2 &= \left\langle \mathit{\Gamma}_f \sum_{k=1}^N \alpha_k \mathcal{P}_{x_k} y_k, \mathit{\Gamma}_f \sum_{\ell=1}^N \alpha_k \mathcal{P}_{x_\ell} y_\ell \right\rangle_{\boldsymbol{\mathcal{H}}} \\ &= \left\langle \sum_{k=1}^N \alpha_k \mathcal{K}_{f(x_k)} y_k, \sum_{\ell=1}^N \alpha_\ell \mathcal{K}_{f(x_\ell)} y_\ell \right\rangle_{\boldsymbol{\mathcal{P}}} \\ &= \|p_N\|_{\boldsymbol{\mathcal{P}}}^2, \quad \text{for any } p_N \in \mathring{\boldsymbol{\mathcal{P}}}. \end{aligned} \tag{27}$$

Thus, $\mathit{\Gamma}_f \in \mathcal{L}(\mathring{\boldsymbol{\mathcal{P}}}, \mathring{\boldsymbol{\mathcal{H}}})$ and $\|\mathit{\Gamma}_f\|_{\mathcal{L}(\mathring{\boldsymbol{\mathcal{P}}}, \mathring{\boldsymbol{\mathcal{H}}})} = 1$. We employ the classical construction of an extension by continuity, such as those discussed in [35, p. 32] and [11], to obtain $\mathit{\Gamma}_f : \boldsymbol{\mathcal{P}} \to \boldsymbol{\mathcal{H}}$ that is a bounded linear operator with $\|\mathit{\Gamma}_f\|_{\mathcal{L}(\boldsymbol{\mathcal{P}}, \boldsymbol{\mathcal{H}})} = 1$. Now, let $p_N = \sum_{k=1}^N \alpha_k \mathcal{P}_{x_k} y_k \in \mathring{\boldsymbol{\mathcal{P}}}$ and $g_M = \sum_{i=1}^M \beta_i \mathcal{K}_{f(x_i)} y_i \in \mathring{\boldsymbol{\mathcal{H}}}$, where $\{\alpha_k\}_{k=1}^N \subset \mathbb{R}$, $\{\beta_k\}_{k=1}^M \subset \mathbb{R}$, $\{x_k\}_{k=1}^{\max\{N,M\}} \subset \Omega$, and $\{y_k\}_{k=1}^{\max\{N,M\}} \subset \mathbb{Y}$. It holds that $\langle \mathit{\Gamma}_f p_N, g_M \rangle_{\boldsymbol{\mathcal{H}}} = \langle p_N, U_f g_M \rangle_{\boldsymbol{\mathcal{P}}}$. Taking the limits for $N, M \to \infty$, we observe that the extension by continuity $\mathit{\Gamma}_f : \boldsymbol{\mathcal{P}} \to \boldsymbol{\mathcal{H}}$ is the adjoint of the Koopman operator U_f, that is, $\mathit{\Gamma}_f \equiv U_f^*$, which concludes the proof. ■

5.2 A Necessary and Sufficient Condition for Koopman Invariance of the Pullback Space

In the approximation of the Koopman operator U_f, it is often critical to ensure that the RKHS $\boldsymbol{\mathcal{H}}$ is Koopman-invariant. We are interested in spaces $\boldsymbol{\mathcal{H}}$ such that if $g \in \boldsymbol{\mathcal{H}}$, then $U_f g \in \boldsymbol{\mathcal{H}}$ for all $g \in \boldsymbol{\mathcal{H}}$. One way to guarantee Koopman invariance is to show that

$$\boldsymbol{\mathcal{P}} = U_f(\boldsymbol{\mathcal{H}}) \overset{\mathcal{I}}{\hookrightarrow} \boldsymbol{\mathcal{H}}, \tag{28}$$

where $\mathcal{I}$ denotes a canonical injection such that $\mathcal{I} : g \mapsto \mathcal{I}g = g$ for all $g \in \boldsymbol{\mathcal{H}}$.

Theorem 2 *Let $f : \Omega \to \Omega$, let the vRKHS $\boldsymbol{\mathcal{H}}(\Omega, \mathbb{Y})$ denote the space of observables, and let $\boldsymbol{\mathcal{P}}(\Omega, \mathbb{Y})$ denote the pullback space of $\boldsymbol{\mathcal{H}}$ under the mapping f. The Koopman invariance property* (28) *is verified if and only if there exists a constant $c > 0$ such that $\mathcal{P} \preceq c^2\mathcal{K}$. In this case, $\|\mathcal{I}\| \leq c$ and $\|\mathcal{I}g\|_{\boldsymbol{\mathcal{H}}} \leq c\|g\|_{\boldsymbol{\mathcal{P}}}$.*

Proof: The result follows from [45, Th. 6.25]. Let $\mathcal{K}_1, \mathcal{K}_2 : \Omega \times \Omega \to \mathbb{Y}$ be admissible operator kernels that define the vRKHSs $\boldsymbol{\mathcal{H}}_1$ and $\boldsymbol{\mathcal{H}}_2$, respectively. The vRKHS $\boldsymbol{\mathcal{H}}_1$ is continuously embedded into $\boldsymbol{\mathcal{H}}_2$, that is, $\boldsymbol{\mathcal{H}}_1 \overset{\mathcal{I}}{\hookrightarrow} \boldsymbol{\mathcal{H}}_2$ if and only if there exists a constant $c > 0$ such that $\mathcal{K}_1 \preceq c^2\mathcal{K}_2$ [45, Th. 6.25]. In the case at hand, we choose $\mathcal{K}_1 \triangleq \mathcal{P}$ and $\mathcal{K}_2 \triangleq \mathcal{K}$. Per the definition of the induced norm, it holds that $\|\mathcal{I}\| \leq c$, which concludes the proof. ■

To the authors' knowledge, the necessary and sufficient condition in Theorem 2 is the first of its kind in the literature on approximations of Koopman operators in vRKHSs. It follows from this result that Koopman invariance necessitates the semidefinite (Loewner) order on kernels. Theorem 2 corrects the pointwise order incorrectly applied by some of the authors in [49]. The following example provides a relatively simple application of Theorem 2.

Example 1 Let $\boldsymbol{\mathcal{H}}$ be the Bargmann-Fock space $\boldsymbol{\mathcal{H}} \triangleq F^2(\mathbb{C}) = \{f(z) = \sum_{m=0}^{\infty} a_m z^m : \sum_{m=0}^{\infty} |a_m|^2 m! < \infty\}$, where $\mathcal{K}(z, w) \triangleq e^{z\bar{w}}$, and, for $g \in F^2(\mathbb{C})$, $\|g\|^2_{F^2(\mathbb{C})} \triangleq \sum_{m=0}^{\infty} |a_m|^2$. In [6], it is proven that the only Koopman operators satisfying the Koopman invariance property are those that correspond to $f(z) = az + b$ with $a, b \in \mathbb{C}$ and $|a| \leq 1$. For this example, let $b = 0$ and $|a| = 1$.

For $g \in F^2(\mathbb{C})$, it holds that $U_f g(z) = g(az)$ and

$$\|U_f g\|^2_{F^2(\mathbb{C})} = \sum_{m=0}^{\infty} |a|^{2m} |a_m|^2 = \|g\|_{F^2(\mathbb{C})}.$$

In particular, U_f is invertible. Thus, if $g \in \boldsymbol{\mathcal{P}}$, then $\|g\|_{\boldsymbol{\mathcal{P}}} = \inf\{\|\tilde{g}\|_{F^2(\mathbb{C})} : U_f \tilde{g} = g\} = \|U_f^{-1} g\|_{F^2(\mathbb{C})} = \|g\|_{F^2(\mathbb{C})}$. Therefore, the inclusion map $\mathcal{I} : \boldsymbol{\mathcal{P}} \to F^2(\mathbb{C})$ is bounded with norm 1. In this setting, $\mathcal{P}(z, w) = \mathcal{K}(az, aw) = e^{(az)\overline{aw}} = e^{z\bar{w}}$.

5.3 Constructing a Koopman-Invariant vRKHS

In the following, we introduce a general strategy, based on countable Hilbert sums of vRKHSs, to generate a Koopman-invariant vRKHS. This result is based on the new necessary and sufficient condition in Theorem 2.

We begin by noting that we can apply the analysis summarized in Theorem 1 recursively by setting

$$\boldsymbol{\mathcal{P}}_0 \triangleq \boldsymbol{\mathcal{H}}, \tag{29a}$$

$$\boldsymbol{\mathcal{P}}_i \triangleq U_f(\boldsymbol{\mathcal{P}}_{i-1}) \quad \text{for all } i \in \mathbb{N}. \tag{29b}$$

With this convention, it holds that $U_f : \boldsymbol{\mathcal{P}}_{i-1} \to \boldsymbol{\mathcal{P}}_i$, $U_f \in \mathcal{L}(\boldsymbol{\mathcal{P}}_{i-1}, \boldsymbol{\mathcal{P}}_i)$, and $\|U_f\|_{\mathcal{L}(\boldsymbol{\mathcal{P}}_{i-1},\boldsymbol{\mathcal{P}}_i)} \leq 1$. Furthermore, the reproducing kernel $\mathcal{P}_i$ of $\boldsymbol{\mathcal{P}}_i$ is given by the recursion

$$\mathcal{P}_0(x_1, x_2) \triangleq \mathcal{K}(x_1, x_2), \tag{30a}$$

$$\begin{aligned}\mathcal{P}_1(x_1, x_2) &\triangleq \mathcal{P}_0(f(x_1), f(x_2)) \\ &\triangleq \mathcal{K}(f(x_1), f(x_2)),\end{aligned} \tag{30b}$$

$$\mathcal{P}_i(x_1, x_2) \triangleq \mathcal{K}(f^i(x_1), f^i(x_2)), \quad \text{for any } i \in \mathbb{N}. \tag{30c}$$

In general, the spaces $\boldsymbol{\mathcal{P}}_i$ are neither nested nor ordered in any easily discernible way for a general function f. However, the Koopman operator U_f maps $\boldsymbol{\mathcal{P}}_{i-1}$ to $\boldsymbol{\mathcal{P}}_i$. In the following, we construct a vRKHS $\boldsymbol{\mathcal{P}}_{\boldsymbol{w}}$ such that

$$\boldsymbol{\mathcal{P}}_i \subset \boldsymbol{\mathcal{P}}_{\boldsymbol{w}} \quad \text{for all } i \in \mathbb{N}_0. \tag{31}$$

For any $i \in \mathbb{N}_0$, let the *weight* $w_i \geq 0$ be such that $\sum_{i=0}^{\infty} w_i^2 < \infty$, and define the *weighted pullback space* $\boldsymbol{\mathcal{P}}_{w_i}$ as the vRKHS generated by

$$\mathcal{P}_{w_i}(x_1, x_2) \triangleq w_i^2 \mathcal{P}_i(x_1, x_2), \quad \text{for any } i \in \mathbb{N}_0. \tag{32}$$

Thus, the countable family of vRKHSs $\{\boldsymbol{\mathcal{P}}_{w_i}\}_{i=0}^{\infty}$ can be employed to construct the container vRKHS $\boldsymbol{\mathcal{P}}_{\boldsymbol{w}}$ as the countably infinite direct sum $\boldsymbol{\mathcal{P}}_{\boldsymbol{w}} \triangleq \bigoplus_{i=0}^{\infty} \boldsymbol{\mathcal{P}}_{w_i}$, where $\boldsymbol{w} \triangleq \{w_i\}_{i=0}^{\infty}$. By proceeding as in [19], we define

$$\mathring{\boldsymbol{\mathcal{P}}}_{\boldsymbol{w}} \triangleq \operatorname{span}\left\{ g = \sum_{i=1}^{N} p_{w_i} : p_{w_i} \in \boldsymbol{\mathcal{P}}_{w_i}, \text{ for } 1 \leq i \leq N \right\}, \tag{33}$$

the candidate norm over $\mathring{\boldsymbol{\mathcal{P}}}_{\boldsymbol{w}}$ as

$$\|g\|_{\boldsymbol{\mathcal{P}}_{\boldsymbol{w}}} \triangleq \inf\left\{ \sqrt{\sum_{i=0}^{N} \|p_{w_i}\|_{\boldsymbol{\mathcal{P}}_{w_i}}^2} : g = \sum_{i=0}^{N} p_{w_i}, p_{w_i} \in \boldsymbol{\mathcal{P}}_{w_i} \right\}, \tag{34}$$

and, as shown in (13), $\boldsymbol{\mathcal{P}}_{\boldsymbol{w}}$ as

$$\boldsymbol{\mathcal{P}}_{\boldsymbol{w}} = \overline{\mathring{\boldsymbol{\mathcal{P}}}_{\boldsymbol{w}}}, \tag{35}$$

where the closure of the finite span $\mathring{\boldsymbol{\mathcal{P}}}_{\boldsymbol{w}}$ is taken with respect to the norm in (34).

Theorem 3 *Suppose that the operator kernel $\mathcal{K}(x_1, x_2) \in \mathcal{L}(\mathbb{Y})$ is bounded on the diagonal as in* (18) *by a constant $\bar{\mathfrak{K}} > 0$. The normed vector space* (35) *is a vRKHS with operator-valued kernel*

$$\mathcal{P}_{\boldsymbol{w}}(x_1, x_2) \triangleq \sum_{i=0}^{\infty} \mathcal{P}_{w_i}(x_1, x_2) = \sum_{i=0}^{\infty} w_i^2 \mathcal{P}_i(x_1, x_2). \tag{36}$$

The series in (36) *converges in the strong operator topology. If we set $w_i \triangleq 2^{-\alpha i}$ for some $\alpha > 0$, then the space $\boldsymbol{\mathcal{P}}_{\boldsymbol{w}}$ is Koopman invariant with*

$$U_f(\boldsymbol{\mathcal{P}}_{\boldsymbol{w}}) \hookrightarrow \boldsymbol{\mathcal{P}}_{\boldsymbol{w}}. \tag{37}$$

Proof: First, we show that

$$\sum_{i=0}^{\infty} \langle \mathcal{P}_{w_i}(x, x)y, y \rangle_{\mathbb{Y}} < \infty \quad \text{for all } x \in \Omega \text{ and } y \in \mathbb{Y}. \tag{38}$$

Since $\mathcal{K}(x_1, x_2)$ is bounded on the diagonal by the constant $\bar{\mathfrak{K}}$, the kernel $\mathcal{P}_i$ of pullback space $\boldsymbol{\mathcal{P}}_i$, for each $i \in \mathbb{N}_0$, is bounded on the diagonal, that is,

$$\|\mathcal{P}_i(x, x)\| = \|\mathcal{K}(f^i(x), f^i(x))\| \leq \bar{\mathfrak{K}}, \quad \text{for all } x \in \Omega \tag{39}$$

and each $i \geq 0$. Thus, (38) is verified since

$$\sum_{i=0}^{\infty} w_i^2 \langle \mathcal{P}_i(x, x)y, y \rangle_{\mathbb{Y}} \leq \sum_{i=0}^{\infty} w_i^2 \|\mathcal{P}_i(x, x)\| \|y\|_{\mathbb{Y}}^2 \leq \left(\sum_{i=0}^{\infty} w_i^2 \right) \bar{\mathfrak{K}} \|y\|_{\mathbb{Y}}^2, \tag{40}$$

and $\sum_{i=0}^{\infty} w_i^2$ is convergent by assumption. Now Proposition 5 of [5] guarantees that

$$\mathcal{P}_{\boldsymbol{w}}(x_1, x_2) \triangleq \sum_{i=0}^{\infty} w_i^2 \mathcal{P}_i(x_1, x_2) \tag{41}$$

is the operator kernel for $\boldsymbol{\mathcal{P}}_{\boldsymbol{w}}$, where the series in (41) converges in the strong operator topology.

We now prove the Koopman invariance property. Since $\mathcal{P}_{\boldsymbol{w}}$ is the admissible operator kernel that defines $\boldsymbol{\mathcal{P}}_{\boldsymbol{w}}$, we can apply the same arguments as in the first part of this proof to argue that $\boldsymbol{\mathcal{Q}}_{\boldsymbol{w}} \triangleq U_f(\boldsymbol{\mathcal{P}}_{\boldsymbol{w}})$ is a vRKHS with reproducing operator kernel

$$\begin{aligned} \mathcal{Q}_{\boldsymbol{w}}(x_1, x_2) &\triangleq \mathcal{P}_{\boldsymbol{w}}(f(x_1), f(x_2)) \\ &= \sum_{i=0}^{\infty} w_i^2 \mathcal{K}(f^{i+1}(x_1), f^{i+1}(x_2)) \\ &= \sum_{j=1}^{\infty} \left(\frac{w_{j-1}}{w_j} \right)^2 w_j^2 \mathcal{K}(f^j(x_1), f^j(x_2)) \end{aligned}$$

$$= \left(\frac{1}{2}\right)^{-\alpha} \sum_{j=1}^{\infty} w_j^2 \mathcal{K}(f^j(x_1), f^j(x_2))$$
$$= \left(\frac{1}{2}\right)^{-\alpha} (\mathcal{P}_{\boldsymbol{w}}(x_1, x_2) - \mathcal{K}(x_1, x_2)) \tag{42}$$

for all $x_1, x_2 \in \Omega$. If follows from (42) that

$$\left(\frac{1}{2}\right)^{-\alpha} \mathcal{K}(x_1, x_2) = \left(\frac{1}{2}\right)^{-\alpha} \mathcal{P}_{\boldsymbol{w}}(x_1, x_2) - \mathcal{Q}_{\boldsymbol{w}}(x_1, x_2) \tag{43}$$

for all $x_1, x_2 \in \Omega$. For any set of distinct centers $\{x_i\}_{i=1}^N \subset \Omega$, coefficients $\{\alpha_i\}_{i=1}^N \subset \mathbb{R}$, and directions $\{y_i\}_{i=1}^N \subset \mathbb{Y}$, it holds that

$$\left(\frac{1}{2}\right)^{-\alpha} \sum_{i,j} \alpha_i \alpha_j \left\langle \mathcal{K}(x_i, x_j) y_j, y_i \right\rangle_{\mathbb{Y}} = \left(\frac{1}{2}\right)^{-\alpha} \sum_{i,j} \alpha_i \alpha_j \left\langle \mathcal{P}_{\boldsymbol{w}}(x_i, x_j) y_j, y_i \right\rangle_{\mathbb{Y}} - \sum_{i,j} \alpha_i \alpha_j \left\langle \mathcal{Q}_{\boldsymbol{w}}(x_i, x_j) y_j, y_i \right\rangle_{\mathbb{Y}}. \tag{44}$$

For any choice of centers, the left-hand side of (44) is nonnegative since, by assumption, the generalized Grammian matrix $\mathbb{K}_N$ is positive definite. Hence, the partial ordering on the operator kernels

$$\mathcal{Q}_{\boldsymbol{w}} \leq \left(\frac{1}{2}\right)^{-\alpha} \mathcal{P}_{\boldsymbol{w}}. \tag{45}$$

is proven. Applying Theorem 6.25 of [45], we deduce that

$$U_f(\boldsymbol{\mathcal{P}}_{\boldsymbol{w}}) = \boldsymbol{\mathcal{Q}}_{\boldsymbol{w}} \hookrightarrow \boldsymbol{\mathcal{P}}_{\boldsymbol{w}}, \tag{46}$$

and, hence, $\boldsymbol{\mathcal{P}}_{\boldsymbol{w}}$ is Koopman-invariant. ■

5.4 Alternative Sufficient Condition for Koopman Invariance

Section 5.2 established necessary and sufficient conditions for the Koopman invariance condition (28) and Section 5.3 outlined the construction of a vRKHS and the associated operator kernel that are invariant according to this characterization. However, the hypotheses of Theorem 2 may be difficult to verify in problems of practical interest, especially if we employ classical kernel functions such as the Wendland, Sobolev-Matérn, Poisson, and Abel kernels. In this section, we describe sufficient conditions to ensure a different Koopman invariance condition, namely

$$U_f \in \mathcal{L}(\boldsymbol{\mathcal{W}}^s(\Omega, \mathbb{Y})) \tag{47}$$

that holds if the vRKHS is a type of vector-valued Sobolev space, that is, $\boldsymbol{\mathcal{H}} = \boldsymbol{\mathcal{W}}^s(\Omega, \mathbb{Y})$. The proposed analysis relies on Theorem 4.2 of [29], which gives sufficient conditions for Koopman invariance of certain scalar-valued Sobolev spaces $\mathcal{W}^s(\Omega)$ and extends the recent work in [29] for scalar-valued native spaces $\mathcal{H} = \mathcal{H}(\Omega, \mathbb{R})$ generated by the Wendland scalar-valued kernel [55].

Theorem 4 *Consider the dynamic equation* (1) *and assume that*

1. $f \in C_b^k(\mathbb{R}^n, \mathbb{R}^n)$ *for some integer* $k > n/2$*;*
2. f *is a* C^1*-diffeomorphism whose Jacobian matrix* $J(x) \triangleq [\partial f_i / \partial x_j(x)] \in \mathbb{R}^{n \times n}$ *is bounded below in the sense that*

$$\inf_{x \in \Omega} |\det(J(x))| > 0, \quad \textit{for all } x \in \Omega. \tag{48}$$

Then, (47) *is verified for all positive real smoothness indices* $0 < s \leq k$*.*

Proof: This result is a transcription to vector-valued Sobolev spaces $\boldsymbol{\mathcal{W}}^s(\mathbb{R}^n, \mathbb{R}^m)$ of the same result for scalar-valued Sobolev spaces $\mathcal{W}^s(\mathbb{R}^n, \mathbb{R})$ presented in [29, Th. 4.2], noting that $\boldsymbol{\mathcal{W}}^s(\mathbb{R}^n, \mathbb{R}^m) = (\mathcal{W}^s(\mathbb{R}^n, \mathbb{R}))^m$. ■

In the following, we apply Theorem 4 to vRKHSs defined in terms of Wendland, Sobolev-Matérn, Poisson, or Abel diagonal kernels, for instance. We begin by reviewing some properties of popular scalar-valued native spaces that are generated by radial kernels. To this goal, recall that the kernel $\mathfrak{K} : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ is *radial* if it can be written as

$$\mathfrak{K}(x, y) \triangleq \Phi(x - y) \quad \text{for all } x, y \in \Omega \tag{49}$$

for some real-valued positive definite function $\Phi \in C(\mathbb{R}^n) \bigcap L^1(\mathbb{R}^n, \mathbb{R})$; see [55, Th. 10.12] for a discussion. The Fourier transform $\hat{\Phi}(\cdot)$ of $\Phi(\cdot)$ satisfies an *algebraic decay condition* with smoothness index $s > 0$ whenever there are two positive constants $C_1, C_2 > 0$ such that

$$C_1(1 + \|\omega\|_2^2)^{-s} \leq \hat{\Phi}(\omega) \leq C_2(1 + \|w\|_2^2)^{-s}, \qquad \text{for all } \omega \in \mathbb{R}^n. \tag{50}$$

Finally, we recall two theorems from [55] that hold for scalar-valued kernels. The first of these theorems gives sufficient conditions for the scalar-valued RKHS $\mathcal{H}(\mathbb{R}^n)$ of functions defined on all of $\mathbb{R}^n$ to be a scalar-valued Sobolev space $\mathcal{W}^s(\mathbb{R}^n)$. The second of these results shows how the restriction of these functions to a sufficiently regular domain $\Omega \subset \mathbb{R}^n$ generates a scalar-valued native space $\mathcal{H}(\Omega) \equiv \mathcal{W}^s(\Omega)$.

Theorem 5 ([55, Cor. 10.13]) *Let* $\mathfrak{K}(x, y) \triangleq \Phi(x - y)$ *be a radial kernel defined in terms of the radial function* $\Phi \in C(\mathbb{R}^n) \bigcap L^1(\mathbb{R}^n, \mathbb{R})$*, whose Fourier transform verifies the algebraic decay condition* (50) *for some smoothness index* $s > n/2$*. Then, the scalar-valued native space* $\mathcal{H}(\mathbb{R}^n, \mathbb{R})$ *defined by* $\mathfrak{K}$ *coincides with the scalar-valued Sobolev space* $\mathcal{W}^s(\mathbb{R}^n, \mathbb{R})$*.*

Theorem 6 ([55, Cor. 10.48]) *Let the kernel* $\mathfrak{K} : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ *define the scalar-valued RKHS* $\mathcal{H}(\mathbb{R}^n)$ *in terms of the radial function* Φ *as in Theorem 5 and suppose that its Fourier transform satisfies the algebraic decay condition in* (50) *for some integer* $s > n/2$*. If* $\Omega \subset \mathbb{R}^n$ *has a Lipschitz boundary, then*

$$\mathcal{H}(\Omega) \equiv \mathcal{W}^s(\Omega), \tag{51}$$

where the scalar-valued RKHS $\mathcal{H}(\Omega)$ *is generated by the reproducing kernel defined in terms of the restriction of* $\mathfrak{K}(\cdot, \cdot)$ *to* $\Omega \times \Omega$*.*

These two theorems can be ported to the framework of [29] and obtain the following result.

Theorem 7 *Let $\mathcal{K} \triangleq \mathfrak{K} I_m$ be a diagonal operator-valued kernel that induces the vRKHS $\boldsymbol{\mathcal{H}}(\mathbb{R}^n, \mathbb{R}^m) = \mathcal{H}^m(\mathbb{R}^n, \mathbb{R}))$, where the scalar-valued kernel $\mathfrak{K}$ that generates $\mathcal{H}(\mathbb{R}^n)$ satisfies the hypotheses of Theorem 4. If the bounded domain $\Omega \subset \mathbb{R}^n$ has a Lipschitz boundary, then the vector-valued vRKHS $\boldsymbol{\mathcal{W}}^s(\Omega, \mathbb{R}^m)$ satisfies the Koopman invariance property* (47).

Example 2 (Wendland Kernels) Let $\Phi_{n,s}$ designate the scalar-valued Wendland kernel in [55] for dimension n and integer smoothness $s \geq 1$. Then, it follows from Theorem 4.1 of [29] that, if $\Omega \subset \mathbb{R}^n$ has a Lipschitz boundary, then the scalar-valued RKHSs $\mathcal{H}_{\Phi_{n,s}}(\Omega)$ generated by the Wendland kernels $\Phi_{n,s}$ are equivalent to the Sobolev spaces

$$\mathcal{H}_{\Phi_{n,k}}(\Omega) \equiv \mathcal{W}^{\sigma_{n,k}}(\Omega), \tag{52}$$

where $\sigma_{n,s} \triangleq (n+1)/2 + s$. Corollary 4.4 of [29] guarantees that, if we additionally require that $f \in C_b^{\lceil \sigma_{n,s} \rceil}$, then

$$U_f \in \mathcal{L}(\mathcal{W}^{\sigma_{n,s}}(\Omega)) \equiv \mathcal{L}(\mathcal{H}_{\Phi_{n,s}}(\Omega)). \tag{53}$$

Lastly, it follows from Theorem 4 that Koopman invariance holds for the vector-valued vRKHS $\boldsymbol{\mathcal{H}}^{\sigma_{n,k}}(\Omega, \mathbb{Y})$ defined in terms of diagonal operator-valued kernels construct from these scalar-valued kernels, that is,

$$U_f \in \mathcal{L}(\boldsymbol{\mathcal{W}}^{\sigma_{n,s}}(\Omega, \mathbb{Y})) \equiv \mathcal{L}(\boldsymbol{\mathcal{H}}^{\sigma_{n,s}}(\Omega, \mathbb{Y})). \tag{54}$$

Example 3 (Sobolev-Matérn Kernels) Sobolev-Matérn kernels are defined in terms of a radial function

$$\Phi_k(r) \triangleq c_s \|r\|_2^{s-n/2} K_{s-n/2}(\|r\|_2), \tag{55}$$

where K_ν denotes the modified Bessel function of the second kind of order ν and c_s denotes a normalization constant that varies among different authors. This kernel has a Fourier transform that satisfies the algebraic decay condition (50) for the smoothness parameter s. As a consequence, the Sobolev-Matérn kernel for $s > n/2$ generates a scalar-valued native space that is equivalent to the scalar-valued Sobolev space $\mathcal{W}^s(\mathbb{R}^n)$. Furthermore, it follows from Theorem 7 that the vRKHS $\boldsymbol{\mathcal{H}}^s(\Omega, \mathbb{Y})$ induced by a diagonal kernel constructed from the scalar-valued kernel Φ_k is Koopman invariant, that is,

$$U_f \in \mathcal{L}(\boldsymbol{\mathcal{W}}^s(\Omega, \mathbb{Y})) \equiv \mathcal{L}(\boldsymbol{\mathcal{H}}^s(\Omega, \mathbb{Y})). \tag{56}$$

Example 4 (Poisson Kernels) Let $\Omega \triangleq [0,1]^n$. The *Poisson reproducing kernel* is given by [52]

$$\mathfrak{K}_P(x,y) \triangleq 2^n \sum_{k_1,\ldots,k_n=1}^{\infty} \frac{\prod_{i=1}^n \sin(\pi k_i x_i) \sin(\pi k_i y_i)}{\pi^2 (k_1^2 + \cdots + k_n^2)}, \tag{57}$$

where $x = [x_1, \ldots, x_n]^{\mathrm{T}} \in \mathbb{R}^n$ and $y = [y_1, \ldots, y_n]^{\mathrm{T}} \in \mathbb{R}^n$. The scalar-valued RKHS generated by this kernel is the Sobolev space $\mathcal{W}^1(\Omega, \mathbb{R})$. This kernel can be understood as the Greens functions for Poisson's equation on $\Omega = [0,1]^n$ subject to zero Dirichlet boundary conditions. Theorem 4 gives sufficient conditions such that the vRKHS $\boldsymbol{\mathcal{H}}_P(\Omega, \mathbb{Y}) = \boldsymbol{\mathcal{W}}^1(\Omega, \mathbb{Y})$ defined in terms of the associated diagonal operator kernel is invariant under the Koopman operator, that is,

$$U_f \in \mathcal{L}(\boldsymbol{\mathcal{W}}^1(\Omega, \mathbb{Y})) \equiv \mathcal{L}(\boldsymbol{\mathcal{H}}_P). \tag{58}$$

Example 5 (Abel Kernels) The *Abel kernel* is defined as

$$\mathfrak{K}_A(x, y) \triangleq C_A e^{-\|x-y\|/\sigma} \quad \text{for all } x, y \in \mathbb{R}^n, \tag{59}$$

where the hyperparameters $\sigma > 0$ and $C_A > 0$ are normalization constants. This reproducing kernel defines the scalar-valued native space $\mathcal{H}_A(\mathbb{R}^n)$ that is equivalent to the Sobolev space [8]

$$\mathcal{H}_A(\mathbb{R}^n) \approx \mathcal{W}^{(n+1)/2}(\mathbb{R}^n) \quad \text{for any } n \in \mathbb{N}. \tag{60}$$

If the domain Ω is sufficiently regular, then $\mathcal{H}_A(\Omega) \approx \mathcal{W}^{(n+1)/2}(\Omega)$, and it follows from Theorem 4 that

$$U_f \in \mathcal{L}(\mathcal{W}^{(n+1)/2}(\Omega, \mathbb{Y})) \approx \mathcal{L}(\mathcal{H}_A(\Omega, \mathbb{Y})). \tag{61}$$

6 Using Noise-Free and Noisy Observations

In this section, we extend the error analysis based on principles of inverse problems, which was introduced in [47, 49] for scalar-valued kernels, to the setting of vRKHSs. Remarkably, the results discussed in this section and the remainder of this paper also apply to vRKHSs that can not be generated by diagonal kernels. The overall philosophy is well known in the study of Galerkin approximations of inverse problems, and the reader is referred to [12, 26, 33] for a detailed account of the general strategy.

6.1 Petrov-Galerkin Approximations

In this paper, we are interested in Galerkin approximations of the solution $g \in \boldsymbol{\mathcal{H}}$ of the *noise-free Koopman operator equation*

$$U_f g = p \in \boldsymbol{\mathcal{P}} \hookrightarrow \boldsymbol{\mathcal{W}} \tag{62}$$

and of the solution $g^\delta \in \boldsymbol{\mathcal{H}}$ of the associated *noisy Koopman operator equation*

$$U_f g^\delta = p^\delta \in \boldsymbol{\mathcal{W}}, \tag{63}$$

where p^δ denotes a *perturbation of* p in (62) due to uncertainty, and δ denotes a *measure of the uncertainty*. In the following, we refer to uncertainties as *noise*.

However, uncertainties considered in this paper do not need to be stochastic. Given the current x, $p(\cdot)$ predicts the next noise-free output so that $y_{i+1} = p(x_i) = (U_f g)(x_i)$. The noisy data p^δ is defined similarly so that $y_{i+1}^\delta = p^\delta(x_i)$. The introduction of the space $\boldsymbol{\mathcal{W}}$ is a standard step in the analysis of Galerkin approximations arising in inverse problems. This space allows formalizing the fact that, while the noise-free data p is an element of the (perhaps small) pullback space $\boldsymbol{\mathcal{P}}$, the noisy data p^δ may reside in some larger space $\boldsymbol{\mathcal{W}}$. Theorem 1 enables the analysis of the noise-free operator equation (62). A solution of the noise-free equation exists for any $p \in \boldsymbol{\mathcal{P}}$. Indeed, if $\mathcal{N}(U_f) = \boldsymbol{\mathcal{Z}}_{f(\Omega)} = \{0\}$, which is verified if f is a function onto Ω, then the solution of (62) is unique. Otherwise, we choose the minimum-norm solution $g \in \boldsymbol{\mathcal{H}}_{f(\Omega)}$ among all the solutions $g \in \boldsymbol{\mathcal{H}}$ of (62). If $\boldsymbol{\mathcal{W}}$ is chosen such that $U_f : \boldsymbol{\mathcal{H}} \to \boldsymbol{\mathcal{W}}$ is compact, then the noisy operator equation is ill-posed [12, 26, 33].

A typical approach to approximating the solution to the inverse problem introduces two finite-dimensional subspaces $\boldsymbol{\mathcal{H}}_N \subseteq \boldsymbol{\mathcal{H}}$ and $\boldsymbol{\mathcal{W}}_N \subseteq \boldsymbol{\mathcal{W}}$ and two associated $\boldsymbol{\mathcal{H}}$- and $\boldsymbol{\mathcal{W}}$-orthogonal projection operators $\boldsymbol{\Pi}_{\boldsymbol{\mathcal{H}}_N} : \boldsymbol{\mathcal{H}} \to \boldsymbol{\mathcal{H}}_N$ and $\boldsymbol{\Pi}_{\boldsymbol{\mathcal{W}}_N} : \boldsymbol{\mathcal{W}} \to \boldsymbol{\mathcal{W}}_N$, respectively, and subsequently defines the *Galerkin approximation operator*

$$G_N \triangleq \left(\boldsymbol{\Pi}_{\boldsymbol{\mathcal{W}}_N} U_f|_{\boldsymbol{\mathcal{H}}_N}\right)^{-1} \boldsymbol{\Pi}_{\boldsymbol{\mathcal{W}}_N}. \tag{64}$$

Thus, we define the Galerkin approximation g_N of the solution g of the noise-free operator equation (62), and correspondingly the approximation g_N^δ of the solution g^δ of the noisy operator equation (63) to be $g_N \triangleq G_N p$ and $g_N^\delta \triangleq G_N p^\delta$, respectively. The Galerkin method defined by G_N is convergent if $G_N U_f g \to g$ in $\boldsymbol{\mathcal{H}}$ for any $g \in \boldsymbol{\mathcal{H}}$.

Let $d(N) \triangleq \dim(\boldsymbol{\mathcal{H}}_N) = \dim(\boldsymbol{\mathcal{W}}_N)$, define the bases

$$\boldsymbol{\mathcal{H}}_N \triangleq \operatorname{span}\{\phi_i\}_{i=1}^{d(N)}, \tag{65}$$

$$\boldsymbol{\mathcal{W}}_N \triangleq \operatorname{span}\{\psi_i\}_{i=1}^{d(N)}, \tag{66}$$

and define the realizations $g_N \triangleq \sum_{i=1}^{d(N)} \phi_i \theta_i$ and $g_N^\delta \triangleq \sum_{i=1}^{d(N)} \phi_i \theta_i^\delta$, where, for all $i \in \{1, \ldots, d(N)\}$, $\phi_i \in \boldsymbol{\mathcal{H}}$, $\psi_i \in \boldsymbol{\mathcal{W}}$, and $\theta_i, \theta_i^\delta \in \mathbb{R}$. The matrix realizations for these approximations are

$$\mathbb{U}_N \boldsymbol{\Theta}_N = \mathbb{P}_N, \tag{67}$$

$$\mathbb{U}_N \boldsymbol{\Theta}_N^\delta = \mathbb{P}_N^\delta, \tag{68}$$

where

$$\mathbb{U}_N \triangleq \left[\langle \psi_i, U_f \phi_j \rangle_{\boldsymbol{\mathcal{W}}}\right]_{(i,j)} \in \mathbb{R}^{d(N) \times d(N)}, \tag{69a}$$

$$\boldsymbol{\Theta}_N \triangleq \left[\theta_i\right]_{(i,1)} \in \mathbb{R}^{d(N)}, \tag{69b}$$

$$\boldsymbol{\Theta}_N^\delta \triangleq \left[\theta_i^\delta\right]_{(i,1)} \in \mathbb{R}^{d(N)}, \tag{69c}$$

$$\mathbb{P}_N \triangleq \left[\langle p, \psi_i \rangle_{\boldsymbol{\mathcal{W}}}\right]_{(i,1)} \in \mathbb{R}^{d(N)}, \tag{69d}$$

$$\mathbb{P}_N^\delta \triangleq \left[\langle p^\delta, \psi_i \rangle_{\boldsymbol{\mathcal{W}}}\right]_{(i,1)} \in \mathbb{R}^{d(N)}, \tag{69e}$$

and $[\cdot]_{(i,j)}$ denotes a matrix by specifying the generic element on its i-th row and j-th column. The general study of Petrov-Galerkin approximations, such as the one proposed here, is a vast topic [26, 33], which gave rise to many variants of individual algorithms generated by different choices of the original operator equation, the bases $\{\phi_i\}_{i=1}^{d(N)}$ and $\{\psi_i\}_{i=1}^{d(N)}$, and the container space $\boldsymbol{\mathcal{W}}$.

The proposed analysis of the vRKHS is based on the following result, which is one of the simpler variants among the available host [26, 33] on Petrov-Galerkin approximations.

Theorem 8 ([33, Th. 17.6]) *Consider the matrix realizations* (67) *and* (68)*, and assume that* $\|\mathbb{P}_N - \mathbb{P}_N^\delta\|_2 \leq \delta$ *for some constant* $\delta > 0$*. Then, the error in a convergent Petrov-Galerkin approximation of the noisy operator equation* (63) *is such that*

$$\|g - g_N^\delta\|_{\boldsymbol{\mathcal{H}}} \leq \underbrace{\frac{1}{\sqrt{s_{\min}(\mathbb{U}_N) s_{\min}(\boldsymbol{\Phi}_N)}}}_{\text{sampling error}} \delta + C \underbrace{\inf_{h_N \in \boldsymbol{\mathcal{H}}_N} \|g - h_N\|_{\boldsymbol{\mathcal{H}}}}_{\text{approximation error}}, \tag{70}$$

where $C > 0$ *is constant,* $s_{\min}(\cdot)$ *denotes the smallest singular value of its argument, and* $\boldsymbol{\Phi}_N \triangleq [\langle \phi_i, \phi_j \rangle_{\boldsymbol{\mathcal{H}}}]_{(i,j)}$ *denotes a* Grammian matrix.

Remark 2 It is known from the theory of inverse methods that there are two distinct contributions to the error bounds on the fidelity of Petrov-Galerkin approximations. The spaces $\boldsymbol{\mathcal{H}}_N$ are designed so that the approximation error decreases as $d(N) \to \infty$. However, the sampling error, in general for a compact operator in the original equations, grows as $d(N) \to \infty$. The contribution of the sampling error, which magnifies the noise δ, is not ordinarily discussed in references on deterministic approximations of Koopman operators. These terms are analogous to the variance and bias contributions to the error in a stochastic setting in statistical and machine learning theory, [35, p. 282], or in the study of distribution-free regression [20].

In this paper, we apply Theorem 8 to two cases of interest. In the first case, we choose the collections of distinct centers $\boldsymbol{\eta}_N \triangleq \{\eta_1, \ldots, \eta_N\} \subset \Omega$ and $\Xi_N \triangleq \{\xi_1, \ldots, \xi_N\} \subset \Omega$, where $\xi_i \triangleq f(\eta_i)$. These sets of centers generate bases and subspaces of approximation defined by specializing (65) and (66) as

$$\boldsymbol{\mathcal{H}}_N \triangleq \operatorname{span}\{\mathcal{K}_{\xi_i} e_j : \xi_i \in \Xi_N, 1 \leq j \leq m\}, \tag{71}$$

$$\boldsymbol{\mathcal{W}}_N \triangleq \boldsymbol{\mathcal{P}}_N \triangleq \operatorname{span}\{\boldsymbol{\mathcal{P}}_{\eta_i} e_j : \eta_i \in \boldsymbol{\eta}_N, 1 \leq j \leq m\}, \tag{72}$$

respectively, where e_i denotes the *i-th canonical basis of* $\mathbb{R}^m$, that is, the i-th column of the identity matrix $I_m \in \mathbb{R}^{m \times m}$. In this case, since the generalized Grammian matrix is positive definite for distinct centers, the dimension $d(N) = mN$ and the matrix realizations become

$$\mathbb{U}_N \triangleq [\mathcal{K}(f(\eta_i), f(\eta_j))]_{(i,j)} \in \mathbb{R}^{mN \times mN}, \tag{73a}$$

$$\mathbb{P}_N \triangleq [p(\eta_i)]_{(i,1)} \in \mathbb{R}^{mN}, \tag{73b}$$

$$\mathbb{P}_N^\delta \triangleq \left[p^\delta(\eta_i)\right]_{(i,1)} \in \mathbb{R}^{mN}. \tag{73c}$$

Note that the matrices $\mathbb{U}_N$ and vector $\mathbb{P}_N^\delta$ can be constructed using (some subset of) the samples $\{(x_i, y_i^\delta)\}_{i=1}^M \approx \{(x_i, p(x_i))\}_{i=1}^M$.

Remark 3 This approach can be interpreted as a kind of collocation method for the Koopman operator. See, for example, [26, Sec. 3.4.1] and [33, Sect 17.3, 17.4], which treat a few cases of using collocation for integral operators of the first kind when the domain Ω is an interval.

In this paper, we also apply Theorem 8 to the case whereby $\boldsymbol{\mathcal{W}} \triangleq L^2(\Omega, \mathbb{Y})$, $\boldsymbol{\mathcal{H}}_N$ is given by (71), and

$$\boldsymbol{\mathcal{W}}_N \triangleq \operatorname{span}\{U_f \mathcal{K}_{\xi_i} e_j = \mathcal{K}(f(\cdot), \xi_i) e_j : \xi_i \in \Xi_N, 1 \le j \le m\}. \tag{74}$$

In this case, $d(N) = mN$ and

$$\mathbb{U}_N \triangleq \left[\int_\Omega \mathcal{K}(\xi_i, f(s))\mathcal{K}(f(s), \xi_j)\mathrm{d}s\right]_{(i,j)} \in \mathbb{R}^{mN\times mN}, \tag{75a}$$

$$\mathbb{P}_N \triangleq \left[\int_\Omega \mathcal{K}(\xi_i, f(s))p(s)\mathrm{d}s\right]_{i,1} \in \mathbb{R}^{mN}, \tag{75b}$$

$$\mathbb{P}_N^\delta \triangleq \left[\int_\Omega \mathcal{K}(\xi_i, f(s))p^\delta(s)\mathrm{d}s\right]_{(i,1)} \in \mathbb{R}^{mN}. \tag{75c}$$

In general, the integrals in (75) can not be computed without explicit knowledge of the unknown function f. To overcome this limitation, we approximate these matrices as follows. Let $\{\Omega_k\}_{k=1}^{M_q}$ denote a partition on Ω and let $\{q_k\}_{k=1}^{M_q}$ denote a set of quadrature points with $q_k \in \Omega_k$ and $k \in \{1, \dots M_q\}$. In this case, we approximate (75) as

$$\mathbb{U}_N \approx \left[\sum_{k=1}^{M_q} \mathcal{K}(\xi_i, f(q_k))\mathcal{K}(f(q_k), \xi_j)|\Omega_k|\right]_{(i,j)}, \tag{76a}$$

$$\mathbb{P}_N \approx \left[\sum_{k=1}^{M_q} \mathcal{K}(\xi_i, f(q_k))p^\delta(q_k)|\Omega_k|\right]_{(i,1)}, \tag{76b}$$

$$\mathbb{P}_N^\delta \approx \left[\sum_{k=1}^{M_q} \mathcal{K}(\xi_i, f(q_k))p^\delta(q_k)|\Omega_k|\right]_{(i,1)}, \tag{76c}$$

where $|\Omega_k|$ denotes the Lebesgue measure of the bounded set Ω_k. To obtain data-driven algorithms, the pairs of quadrature points and values $\{(q_i, f(q_i))\}_{i=1}^{M_q}$ are approximated by pruning the samples $\{(x_i, y_i^\delta)\}_{i=1}^M$.

Remark 4 This case should be compared to the analysis in [33, Sec. 17.3], which treats the approximation of an integral operator that acts on a scalar-valued RKHS over the interval $\Omega = [0, 1]$. In this paper, we study the case whereby the operator is the Koopman operator, the domain Ω is a general bounded domain or manifold $\mathcal{M}$, and the error analysis is based on the doubling trick and matrix-valued power function derived in the next section.

6.2 Power Functions and The Doubling Trick in vRKHS

This section proposes an upper bound on the approximation error term in (70), which has not appeared yet in the literature on the deterministic approximation of Koopman operators. This upper bound extends the strategy used in the proof of [55, Th. 11.23] for scalar-valued kernels to the vRKHS setting; see Section 11.4 of [55]. This technique, referred to as the "doubling trick" in the scalar-valued RKHS setting, describes a regularity condition that enables significant improvement in the convergence rates of approximations constructed from scalar-valued kernels. In the setting of scalar-valued RKHS, the doubling trick continues to be the foundation for estimates with high convergence rates [24].

Since $\Pi_{\boldsymbol{\mathcal{H}}_N} : \boldsymbol{\mathcal{H}} \to \boldsymbol{\mathcal{H}}_N$ is the $\boldsymbol{\mathcal{H}}$-orthogonal projection onto the vRKHS $\boldsymbol{\mathcal{H}}_N \subset \boldsymbol{\mathcal{H}}$, the pointwise bound

$$\begin{aligned} \|E_x(I - \Pi_{\boldsymbol{\mathcal{H}}_N})f\|_2 &\leq \sqrt{m}\|E_x(I - \Pi_{\boldsymbol{\mathcal{H}}_N})f\|_\infty \\ &\leq \sqrt{m}\bar{P}_{\boldsymbol{\mathcal{H}}_N}(x)\|(I - \Pi_{\boldsymbol{\mathcal{H}}_N})f\|_{\boldsymbol{\mathcal{H}}} \end{aligned} \tag{77}$$

follows from [35, Th. 3.22] for all $f \in \boldsymbol{\mathcal{H}}(\Omega, \mathbb{Y})$, where

$$\bar{P}_{\boldsymbol{\mathcal{H}}_N}(x) \triangleq \max_{1 \leq i \leq m} \sqrt{|\mathcal{K}_{ii}(x, x) - \mathcal{K}_{N,ii}(x, x)|} \tag{78}$$

denotes the *power function* [35, Def. 3.19], $\mathcal{K}_N$ denotes the operator-valued kernel that defines $\boldsymbol{\mathcal{H}}_N$ [35, p. 103], and A_{ij} denotes the element of the generic matrix $A \in \mathbb{R}^{n \times m}$ on the i-th row and j-th column. It should be noted that alternative forms of the bound in (77) are documented in the proof of [35, Th. 3.22] using the Euclidean equi-induced matrix norm instead of the supremum norm. We use (77) because it avoids the calculation of eigenvalues required to compute $\|\mathcal{K}(x, x) - \mathcal{K}_N(x, x)\|_{2,2}$ and it allows a straightforward implementation at the modest expense of introducing the factor $\sqrt{m}$ associated with the norm equivalence of $(\mathbb{Y}, \|\cdot\|_2)$ and $(\mathbb{Y}, \|\cdot\|_\infty)$. An immediate consequence of (77) is that

$$\|(I - \Pi_{\boldsymbol{\mathcal{H}}_N})f\|_{L^2(\Omega,\mathbb{Y})} \leq \sqrt{m}|\Omega| \sup_{\xi \in \Omega} \bar{P}_{\boldsymbol{\mathcal{H}}_N}(\xi)\|(I - \Pi_{\boldsymbol{\mathcal{H}}_N})f\|_{\boldsymbol{\mathcal{H}}} \tag{79}$$

for all $f \in \boldsymbol{\mathcal{H}}(\Omega, \mathbb{Y})$.

Next, we study how the pointwise error bounds (77) and (79), expressed in terms of the power function, can extend the "doubling trick" [55, Th. 11.23]

in the context of scalar-valued native spaces to vector-valued native spaces. To this goal, define the integral operator

$$(Lv)(x) \triangleq \int_{\Omega} \mathcal{K}(x,s)v(s)\mathrm{d}s. \tag{80}$$

This operator is bounded and linear from $L^2(\Omega,\mathbb{Y})$ to $\boldsymbol{\mathcal{H}}$. Indeed, since we assume that the mapping $(x,y)\mapsto\|\mathcal{K}(x,y)\|$ is continuous and $\mathcal{K}$ is uniformly bounded on the diagonal, it follows from [35, Th. 3.21] that the embedding $\boldsymbol{\mathcal{H}} \overset{\mathcal{J}}{\hookrightarrow} C(\Omega,\mathbb{Y})$ is compact, and, hence, so is the embedding $\boldsymbol{\mathcal{H}} \overset{\mathcal{I}}{\hookrightarrow} L^2(\Omega,\mathbb{Y})$.

Theorem 9 *Consider the integral operator* (80)*. If* $\boldsymbol{\mathcal{H}} \overset{\mathcal{I}}{\hookrightarrow} L^2(\Omega,\mathbb{Y})$*, then* $L = \mathcal{I}^*$ *and*

$$\langle Lv,h\rangle_{\boldsymbol{\mathcal{H}}} = \langle v,\mathcal{I}h\rangle_{L^2(\Omega,\mathbb{Y})} = \langle v,h\rangle_{L^2(\Omega,\mathbb{Y})} \tag{81}$$

for all $v\in L^2(\Omega,\mathbb{Y})$ *and* $h\in\boldsymbol{\mathcal{H}}\subset L^2(\Omega,\mathbb{Y})$*.*

Proof: For any $v\in L^2(\Omega,\mathbb{Y})$ and $h\in\boldsymbol{\mathcal{H}}\subset L^2(\Omega,\mathbb{Y})$, it holds that

$$\begin{aligned}
\langle Lv,h\rangle_{\boldsymbol{\mathcal{H}}} &= \left\langle \int_{\Omega}\mathcal{K}(\cdot,s)v(s)\mathrm{d}s, h(\cdot)\right\rangle_{\boldsymbol{\mathcal{H}}} \\
&= \int_{\Omega}\langle \mathcal{K}(\cdot,s)v(s), h(\cdot)\rangle_{\boldsymbol{\mathcal{H}}}\,\mathrm{d}s \\
&= \int_{\Omega}\langle v(s), E_s h(\cdot)\rangle_{\mathbb{Y}}\,\mathrm{d}s \\
&= \langle v,\mathcal{I}h\rangle_{L^2(\Omega,\mathbb{Y})} \\
&= \langle v,h\rangle_{L^2(\Omega,\mathbb{Y})}\,,
\end{aligned} \tag{82}$$

which concludes the proof. ■

Remark 5 To apply Theorem 9, the set $\Omega\subset\mathbb{X}\triangleq\mathbb{R}^n$ underlying (80) must be Lebesgue measurable. For instance, Ω can be chosen as a bounded domain, that is, a bounded, open, connected set. Regularity of the boundary of Ω is not required. Of particular interest is the case whereby Theroem 9 is applied with Ω being a compact, connected, smooth manifold $\mathcal{M}$ that is regularly embedded in $\mathbb{X}\triangleq\mathbb{R}^n$. In this case, we employ the operator kernel $\mathcal{K}(m_1,m_2)\in\mathcal{L}(\mathbb{Y})$ defined for all $m_1,m_2\in\mathcal{M}$, the integral operator (80) is redefined as

$$(Lv)(x) \triangleq \int_{\mathcal{M}}\mathcal{K}(x,m)v(m)\mu(\mathrm{d}m) \quad \text{for all } x\in\mathcal{M}, \tag{83}$$

and Theorem 9 proves that $\boldsymbol{\mathcal{H}}(\mathcal{M},\mathbb{Y}) \overset{\mathcal{I}}{\hookrightarrow} L^2_{\mu}(\mathcal{M},\mathbb{Y})$. This result is essential for the numerical examples discussed in Section 7 below.

The next result provides an upper bound on the error made by projecting f onto the finite-dimensional RKHS $\boldsymbol{\mathcal{H}}_N$.

Theorem 10 (The "Doubling Trick" in vRKHSs) *If f verifies the regularity condition $f = Lv$ for some $v \in L^2(\Omega, \mathbb{Y})$, then*

$$\|(I - \Pi_{\boldsymbol{\mathcal{H}}_N})f\|_{\boldsymbol{\mathcal{H}}} \leq \sqrt{m}|\Omega| \sup_{\xi \in \Omega} \bar{P}_{\boldsymbol{\mathcal{H}}_N}(\xi)\|v\|_{L^2(\Omega,\mathbb{Y})}. \tag{84}$$

Proof: It follows from [55, Lemma 10.24] that

$$\langle (I - \Pi_{\boldsymbol{\mathcal{H}}_N})f, (I - \Pi_{\boldsymbol{\mathcal{H}}_N})f\rangle_{\boldsymbol{\mathcal{H}}} = \langle (I - \Pi_{\boldsymbol{\mathcal{H}}_N})f, f\rangle_{\boldsymbol{\mathcal{H}}}\,. \tag{85}$$

Theorem 9 then implies that

$$\begin{aligned}\|(I - \Pi_{\boldsymbol{\mathcal{H}}_N})f\|^2_{\boldsymbol{\mathcal{H}}} &= \langle (I - \Pi_{\boldsymbol{\mathcal{H}}_N})f, (I - \Pi_{\boldsymbol{\mathcal{H}}_N})f\rangle_{\boldsymbol{\mathcal{H}}} \\ &= \langle (I - \Pi_{\boldsymbol{\mathcal{H}}_N})f, Lv\rangle_{\boldsymbol{\mathcal{H}}} \\ &= \langle (I - \Pi_{\boldsymbol{\mathcal{H}}_N})f, v\rangle_{L^2(\Omega,\mathbb{Y})}\,, \\ &\leq \|(I - \Pi_{\boldsymbol{\mathcal{H}}_N})f\|_{L^2(\Omega,\mathbb{Y})}\|v\|_{L^2(\Omega,\mathbb{Y})},\end{aligned} \tag{86}$$

and the result follows substituting the bound in (79) into (86). ■

Every vRKHS satisfies the pointwise bound in (77). The "doubling trick" provides sufficient conditions for an improved bound to be verified. In general, approximations of smoother functions converge faster than approximations of less smooth functions do. The integral operator L in the regularity condition maps a function in $L^2_\nu(\Omega, \mathbb{Y})$, which may not even be continuous, into a smoother function. Thus, since the integral operator underlies a reproducing kernel, it smoothens functions in $L^2_\nu(\Omega, \mathbb{Y})$ that are contained in the vRKHS generated by the kernel.

The next result provides an upper bound on approximating the solution of the noise-free Koopman operator equation (62) with the solution of the noisy Koopman operator equation (63).

Theorem 11 *Assume that the unknown function $g \in \boldsymbol{\mathcal{H}}$ satisfies the regularity condition in Theorem 10. Then, there exists a constant $C > 0$ such that*

$$\|g - g^\delta_N\|_{\boldsymbol{\mathcal{H}}} \leq \underbrace{\frac{1}{\sqrt{s_{\min}(\mathbb{U}_N)s_{\min}(\mathbb{K}_N)}}\delta}_{\text{sampling error}} + \underbrace{C\,\sqrt{m}|\Omega| \sup_{\xi \in \Omega} \bar{P}_{\boldsymbol{\mathcal{H}}_N}(\xi)\|v\|_{L^2(\Omega,\mathbb{Y})}}_{\text{approximation error}}, \tag{87}$$

where $\mathbb{K}_N$ denotes the generalized Grammian matrix of $\boldsymbol{\mathcal{H}}_N$, and $\mathbb{U}_N$ is defined as in (73a) *or, alternatively,* (75a).

Proof: The result is a direct consequence of Theorem 8 and 10. ■

Remark 6 In general, the sampling error grows as $d(N) \to \infty$. This occurs, for example, if the Koopman operator $U_f : \boldsymbol{\mathcal{H}} \to \boldsymbol{\mathcal{W}}$ is compact; see, for example, [26, Th. 2.2]. This qualitative behavior is reflected in the dependence of (87) on the minimum singular value of the generalized Grammians $\mathbb{K}_N$

and $\mathbb{U}_N$. Further intuition can be based on a careful analysis of the scalar-valued case when $m = 1$, which has been studied extensively. For a wide collection of common kernels, it is known that the smallest singular value of $\mathbb{K}_N$ is bounded from below by a function of the minimal separation r_{Ξ_N} between the centers raised to some power s that depends on the smoothness of the choice of kernel [55, Cor. 12.6–12.8]. Then, it is to be expected that $s_{\min}(\mathbb{K}_N) \sim r_{\Xi_N}^s \to 0$ as $d(N) \to \infty$.

The error bound (87) is expressed in terms of the operator-valued power function $\overline{P}_{\boldsymbol{\mathcal{H}}_N}$. This expression holds for any general, possibly non-diagonal, operator-valued kernel. However, as discussed in [35, Ch. 3], if the operator-valued kernel is diagonal, that is, $\mathcal{K} = \mathfrak{K} I_m$, where $\mathfrak{K}(\cdot,\cdot)$ denotes a scalar-valued Mercer kernel and I_m denotes the identity matrix in $\mathbb{R}^{m\times m}$, then this bound can be improved. In this case, the power function in (87) can be bounded by

$$\overline{\mathcal{P}}_{\mathcal{H}_N}(x) \leq \sqrt{\mathcal{N}(h_{\Xi_N,\Omega})} \tag{88}$$

where $\mathcal{N}(\cdot)$ denotes a known function and $h_{\Xi_N,\Omega}$ denotes the fill distance defined in (10). For details, see [35, Sec. 3.8.2].

6.3 Post-Projection of Galerkin Approximations

The error analysis outlined in Section 6.2 is quite general and can be applied to the study of Petrov-Galerkin approximations of the noise-free and noisy, but deterministic, Koopman operator equations in a host of situations. The Galerkin approximations can be computed using the collection of noisy observations $\{(x_i, y_i^\delta)\}_{i=1}^M$. The bounds on the Galerkin approximation errors can be used to infer bounds on the approximations of the Koopman operators since

$$\|U_f g - U_f g_N^\delta\|_{\boldsymbol{\mathcal{P}}} \leq \|U_f\| \|g - g_N^\delta\|_{\boldsymbol{\mathcal{H}}}. \tag{89}$$

Furthermore, the results in Section 6.2 do not rely on any form of Koopman invariance.

While the computations of g_N^δ can be performed in a "data-dependent way" using only the observations along a trajectory, if we want to use the expression

$$\hat{y}_N = U_f g_N^\delta = \sum_{i=1}^N \mathcal{K}(f(\cdot), \xi_i)\theta_i^\delta \tag{90}$$

for creating forecasts, which was the original intent, then explicit knowledge of f would be needed. To overcome this problem, a "post-projection" step has been studied in several works by the authors such as [48,49]. If $\boldsymbol{\mathcal{P}} \hookrightarrow \boldsymbol{\mathcal{H}}$, that is, if $\boldsymbol{\mathcal{H}}$ is Koopman invariant, then

$$\hat{y}_N \triangleq \mathit{\Pi}_{\boldsymbol{\mathcal{H}}_N} U_f g_N^\delta \tag{91}$$

can be constructed *without* explicit knowledge of f. The collection of samples $\{(x_i, y_i^\delta)\}_{i=1}^M$ is sufficient to calculate the output prediction. In fact, as noted in [48], the matrix realizations of (91) arise in the popular extended dynamic model decomposition (EDMD) method [27, 28, 30].

This post-projection step does not affect the asymptotic behavior of the error bounds. Indeed, if $\boldsymbol{\mathcal{P}} \overset{\mathcal{I}}{\hookrightarrow} \boldsymbol{\mathcal{H}}$, then

$$\|U_f g - \mathit{\Pi}_{\boldsymbol{\mathcal{H}}_N} U_f g_N^\delta\|_{\boldsymbol{\mathcal{H}}} \leq \|U_f g - U_f g_N^\delta\|_{\boldsymbol{\mathcal{H}}} + \|(I - \mathit{\Pi}_{\boldsymbol{\mathcal{H}}_N}) U_f g_N^\delta\|_{\boldsymbol{\mathcal{H}}}. \tag{92}$$

The first term on the right-hand side of (92) can be bounded as in Theorem 11 since $\|U_f g - U_f g_N^\delta\|_{\boldsymbol{\mathcal{H}}} \leq \|\mathcal{I}\| \|U_f\| \|g - g_N^\delta\|_{\boldsymbol{\mathcal{H}}}$. This term captures the ability of $\hat{y}_{m+1} = U_f g_N^\delta$ to capture the true value $y_{m+1} = U_f g$ and it depends explicitly on the number of centers N used to define the vRKHS $\boldsymbol{\mathcal{H}}_N$ that approximates $\boldsymbol{\mathcal{H}}$ and the number M of collected data points. The second term on the right-hand side of (92) can be bounded as in Theorem 10, assuming the regularity of $U_f g_N^\delta$. Theorem 2 gives necessary and sufficient conditions that justify this post-projection step. Section 5.4 provides an alternative set of results that justify this step for certain diagonal operator-valued kernels that define Sobolev spaces.

7 Numerical Example

In this section, we apply the proposed results on Petrov-Galerkin methods for noise-free and noisy Koopman operator equations for a discretized version of the Lotka-Volterra system. This system has been selected because its Hamiltonian function is an integral invariant [21], and, consequently, this system exhibits closed periodic orbits. We argue that these orbits can be identified with one-dimensional smooth submanifolds regularly embedded in $\mathbb{X} \triangleq \mathbb{R}^2$. Hence, we can study the convergence of approximations of the Koopman operator based on samples from a single initial condition that defines a single one-dimensional submanifold $\mathcal{M}$ of $\mathbb{X} \triangleq \mathbb{R}^n$, which has zero Lebesgue measure in $\mathbb{R}^2$. Alternatively, we can combine initial conditions to study the convergence of approximations of the Koopman operator over a bounded domain in $\Omega \subset \mathbb{R}^2$.

7.1 The Lotka-Volterra System

The Lotka-Volterra system examined in this paper is given by

$$\dot{x}_1(t) = x_1(t)(x_2(t) - \alpha), \quad x_1(0) = x_{1,0}, \quad t \geq 0, \tag{93a}$$

$$\dot{x}_2(t) = x_2(t)(\beta - x_1(t)), \quad x_2(0) = x_{2,0}, \tag{93b}$$

where $x_1, x_2 : [0, \infty) \to [0, \infty)$ represent the populations of a predator and a prey species, respectively. The coefficients $\alpha, \beta > 0$ are critical thresholds in the predator-prey interaction. If the prey population exceeds α, then the predator

population grows at a rate proportional to its current size. Alternatively, if the predator population is below β, then the prey population grows at a rate proportional to its current size.

The *Hamiltonian function*, from [21] given by

$$\mathfrak{H}(x_1(t), x_2(t)) \triangleq \beta \ln(x_1(t)) - x_1(t) + \alpha \ln(x_2(t)) - x_2(t), \tag{94}$$

is evaluated along the trajectories of (93) for all $t \geq 0$ and is an integral invariant for (93). Hence,

$$\mathfrak{H}(x_{1,0}, x_{2,0}) = \mathfrak{H}(x_1(t), x_2(t)) \quad \text{for all } t \geq 0. \tag{95}$$

The Jacobian of $\mathfrak{H}(x_1, x_2)$ does not vanish over a closed trajectory of (93), and, by the implicit function theorem, this invariant describes a one-dimensional, compact, smooth submanifold that is regularly embedded in $\mathbb{R}^2$. It can always be intrepreted, via a suitable definition of an inner product on the tangent space, as a compact, connected, one-dimensional Riemannian manifold $\mathcal{M}$ that is regularly embedded in $\mathbb{R}^2$. Since we use extrinsic approximations methods such as those in [15], we do not need to use the closed form coordinate charts for the Riemannian manifold in our calculations. Figure 1 illustrates multiple level sets of the Hamiltonian function.

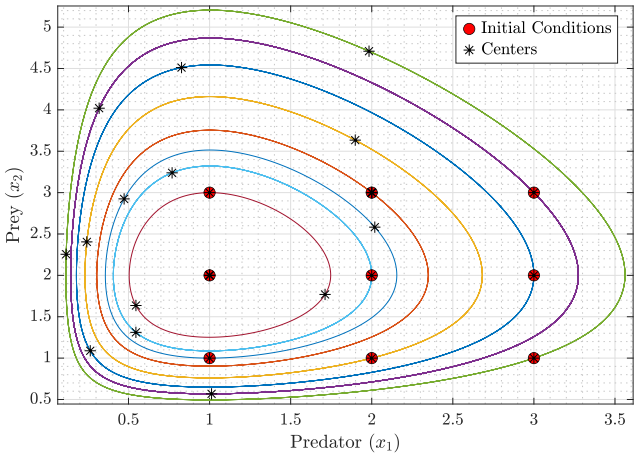

Fig. 1 Phase portrait of the Lotka-Volterra system (93). Each trajectory corresponds to a solution of the system from a different initial condition. Red circles represent selected initial conditions used for data generation, and black stars denote the kernel centers used in the Koopman approximation

To discretize the continuous-time system (93), we apply the symplectic Euler method, which updates x_2 implicitly and x_1 explicitly. This geometric integration method is designed to better preserve geometric features of a flow, such as integral invariants. The use of this discretization approach has already been shown in [21] to preserve the closed-orbit behavior of discrete approximations of the Lotka–Volterra system compared to the exclusively explicit or

implicit Euler methods. The resulting discrete-time dynamics are in the same form as (1) with

$$f(x_i) = \begin{bmatrix} x_{1,i} + hx_{1,i}\left(\dfrac{x_{2,i}}{1-h(\beta - x_{1,i})} - \alpha\right) \\ \dfrac{x_{2,i}}{1-h(\beta - x_{1,i})} \end{bmatrix}, \tag{96}$$

where $h > 0$ denotes the width of time step, and $x_{1,i}$ and $x_{2,i}$ denote the predator and prey populations at the i-th time step, respectively. The trajectories shown in Figure 1 were obtained setting $\alpha = 1$ and $\beta = 1$.

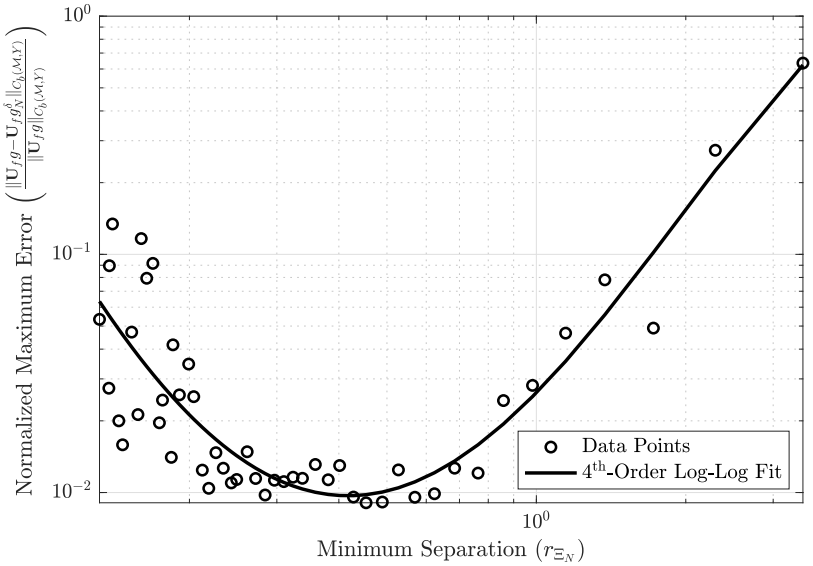

Fig. 2 Log-log plot of the normalized maximum prediction error versus the minimum separation distance r_{Ξ_N}, computed using Method 1 under deterministic sinusoidal measurement uncertainty with a magnitude of $\delta = 0.5\%$. Each marker corresponds to the normalized maximum error, and the solid curve represents a fourth-order polynomial fit in the log-log scale. Results are shown for the Gaussian kernel

Let the output be defined so that, at step i,

$$y_i = g(x_i) = \begin{bmatrix} a_1 & 0 \\ 0 & a_2 \\ a_1 & a_2 \end{bmatrix} x_i, \quad i \in \mathbb{N}_0, \tag{97}$$

where a_1 and a_2 are arbitrary scaling constants representing the contributions of prey and predator, respectively, to the measured quantities (*e.g.*, food production). The proposed numerical simulations are performed by setting $a_1 = 5$ and $a_2 = 10$. To model the measurement uncertainty, we define the observed output as

$$y_{m+1}^\delta = (1 + \delta_{m+1}) \cdot (U_f g)(x_m), \quad m \in \mathbb{N}_0, \tag{98}$$

where δ_{m+1} represents measurement noise. The theoretical bounds derived in this paper are examples of finite error bounds for deterministic systems. However, the final, finite sample error bounds still hold if the centers and output observations are generated from stochastic dynamics. Therefore, we present the results of simulations involving both deterministic and stochastic measurement noises. The finite sample bounds are the same regardless of whether the samples are generated deterministically or randomly, but only hold for the one realization of samples if they are generated stochastically.

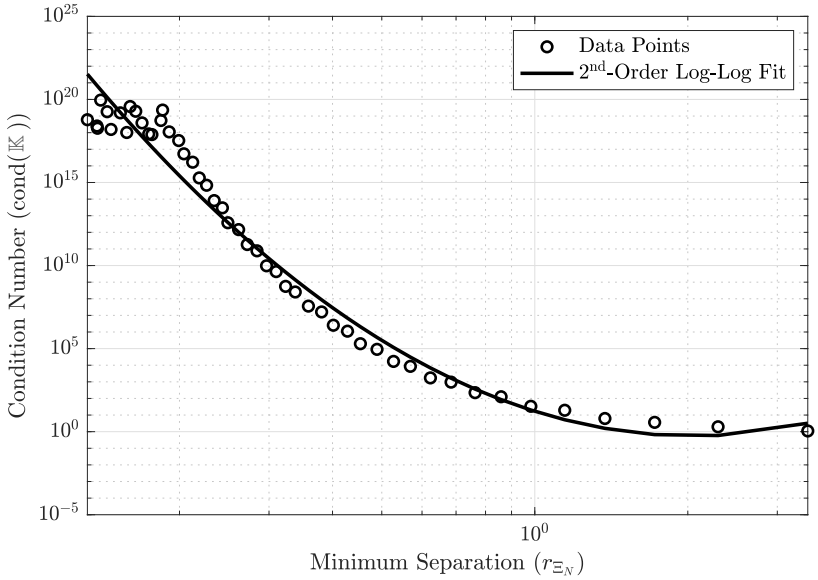

Fig. 3 Log-log plot of the condition number of the Grammian matrix versus the minimum separation distance r_{Ξ_N}, corresponding to Method 1 with the Gaussian kernel. Each marker represents the condition number evaluated at a given r_{Ξ_N}, and the solid curve represents a second-order polynomial fit in the log-log scale. High condition numbers produce potential numerical instabilities in the factorization or inversion of the generalized Grammian matrix

We employ the Galerkin approximations defined in (73), referring to them as "collocation without quadrature," and the approximations defined by (76), which are described as "collocation with quadrature." Furthermore, we consider two different choices of the initial conditions:

1. Single Initial Condition. In this case, we collect samples $\Xi_N \triangleq \{\xi_i\}_{i=1}^N$ along a single orbit $\mathcal{M}$ of (93) for a fixed initial condition. These samples are progressively more dense in $\mathcal{M}$. However, these samples are not dense in any open connected subset Ω of $\mathbb{R}^2$. The proposed results are applied to $g \in \boldsymbol{\mathcal{H}}(\mathcal{M}, \mathbb{Y})$ with the approximating subspace

$$\boldsymbol{\mathcal{H}}_{\mathcal{M},N} \triangleq \text{span}\{\mathcal{K}_{\xi_i} y : \xi_i \in \Xi_N \subset \mathcal{M}, y \in \mathbb{Y}\}. \tag{99}$$

The proposed error bounds will be in the norm on $\boldsymbol{\mathcal{H}}(\mathcal{M}, \mathbb{Y})$ on the errors in $C_b(\mathcal{M}, \mathbb{Y})$.

2. Multiple Initial Conditions. In this case, we define the *sublevel set*

$$\Omega = \{(x_1, x_2) \in \mathbb{R}^2 : \mathfrak{H}(x_1, x_2) \leq C_\Omega\} \tag{100}$$

whose boundary is the level set characterized by $C_\Omega > 0$. Samples are collected along multiple trajectories with initial conditions in Ω. The larger the number of initial conditions and trajectories, the smaller the fill distance of the samples in Ω. The proposed results are applied to $g \in \boldsymbol{\mathcal{H}}(\Omega, \mathbb{Y})$ with the approximating subspace

$$\boldsymbol{\mathcal{H}}_{\Omega,N} \triangleq \operatorname{span}\{\mathcal{K}_{\xi_i} y : \xi_i \in \Xi_N \subset \Omega, y \in \mathbb{Y}\}. \tag{101}$$

The proposed error bounds will be in the norm on $\boldsymbol{\mathcal{H}}(\Omega, \mathbb{Y})$ on the errors in $C_b(\Omega, \mathbb{Y})$.

To assess the quality of the approximations of the Koopman operator, we evaluate the normalized maximum prediction error using the uniform norm. Given the bounded domain Ω, this norm can be approximated numerically by

$$\frac{\left\|\mathbf{U}_f g - \mathbf{U}_f g_N^\delta\right\|_{C_b(\Omega,\mathbb{Y})}}{\|\mathbf{U}_f g\|_{C_b(\Omega,\mathbb{Y})}} \approx \max_{i=1,\ldots,M} \frac{\left\|(\mathbf{U}_f g)(x_i) - (\mathbf{U}_f g_N^\delta)(x_i)\right\|_\infty}{\|\mathbf{U}_f g\|_{C_b(\Omega,\mathbb{Y})}}. \tag{102}$$

Such errors can be computed as functions of either the minimal separation radius (9) or the fill distance (10); in this section, we assume that the sample distribution is quasi-uniform. Equation (102) has been stated for the bounded domain $\Omega \subset \mathbb{X}$. The same result applies by replacing the bounded domain Ω with a smooth manifold $\mathcal{M}$ regularly embedded in $\mathbb{X}$.

Although this numerical example provides evidence that the error bounds derived in this paper can be observed in practical situations, several complicating factors are unavoidable in interpreting or numerically verifying the new bound in Theorem 11. Four of these factors are discussed in the following. First, the approximation error term includes the constant C that is unknown in general in closed form. This constant is typical of Galerkin approximations in general, and its origin can be traced to conditions sufficient to establish that the Galerkin approximations are convergent. We do not treat such sufficiency conditions in this paper. It would require substantial additional work to deduce explicit forms for C for the Koopman problems at hand. We leave this problem for future research. Secondly, the supremum of the power function can *in principle* be estimated empirically or numerically through an optimization problem. However, the estimation of the minimum of the power function numerically is substantially equivalent to computing $\mathbb{K}_N^{-1}$, since this inverse appears in the definition of the power function. Hence, invariably, a numerical estimate of the supremum necessarily suffers from perturbations due to the ill-conditioning of $\mathbb{K}_N$ as N grows. Numerical studies of the error bound in (87) induce numerical perturbations to both the sample and approximation error terms. Thirdly, we presume to choose δ that bounds the perturbation to the problem data in (98). In reality, we choose a parameter, say δ_y, that differs from δ by numerical inaccuracies in forming the matrix $\mathbb{U}_N$. However,

these perturbations, say $\delta_{\mathbb{U}}$ are unknown, and, hence, $\delta = \delta_y + \delta_{\mathbb{U}}$ is unknown. Lastly, in many cases, the power function can be bounded by a known function $F(h_{\Xi_N,\Omega})$ as in Table 11.1 of [55]. For example, both of the classes of Wendland and Matérn kernels have power functions that are bounded by $h^s_{\Xi_N,\Omega}$ for a positive integer s that depends on the specific kernel and measures smoothness. However, if $\Omega \subset \mathbb{R}^n$ is a parallelopiped, then collections of N quasiuniform samples scale like

$$N(\Omega) \sim \left(\frac{1}{h_{\Xi_N,\Omega}}\right)^n.$$

Similarly, for N quasiuniform samples in an ℓ-dimensional, compact, connected, smooth Riemannian submanifold $\mathcal{M}$ that is regularly embedded in $\mathbb{R}^n$, we have

$$N(\mathcal{M}) \sim \left(\frac{1}{h_{\Xi_N,\mathcal{M}}}\right)^\ell.$$

For this reason, we chose to construct logarithmic plots of the error versus the logarithm of the fill distance or minimal separation radius, since

$$\log(N) \sim -\log h_{\Xi_N,\Omega} \sim -\log r_{\Xi_N}.$$

Thus, while all the plots of the empirical error from numerical solutions are plotted in terms of the fill distance or minimum separation of the centers, these plots also can be interpreted, modulo a scaling factor, as the trends of the errors as functions of $\log(N)$, where N denotes the number of samples. In view of the comments, we note that it is extremely difficult, even in principle, to make precise assessments using numerical studies of the exact contributions of the two terms in (87) that correspond to the sampling error and the approximation error. However, Figures 4 and 5 provide qualitative evidence that supports the trends implied in the bound given in (87).

7.2 Method 1: Collocation without Quadrature

In this section, we employ the Galerkin approximation defined by (73), the scalar-valued *Gaussian kernel*

$$\mathfrak{K}(x_1, x_2) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{\|x_1 - x_2\|_2^2}{2\sigma^2}\right), \quad \text{for all } (x_1, x_2) \in \mathbb{R} \times \mathbb{R}, \tag{103}$$

where $\sigma > 0$, and the operator kernel $\mathcal{K}(x_1, x_2) = \mathfrak{K}(x_1, x_2)I_3$, where I_3 denotes the identity matrix in $\mathbb{R}^3$. Noisy samples are collected along a single trajectory, and approximations are constructed from subspaces defined as in (99). The centers Ξ_N used to define the approximating subspaces are subselected from the collection of all samples $\{x_i\}_{i=1}^M$ so that they are approximately uniformly distributed over the manifold $\mathcal{M}$.

Figures 2 and 3 illustrate the results of the numerical simulations with $(x_{1,0}, x_{2,0}) = (1, 1)$, $h = 0.025$, $M = 1865$, and $\delta_i = 0.005\sin(40\pi t)$. Figure 2

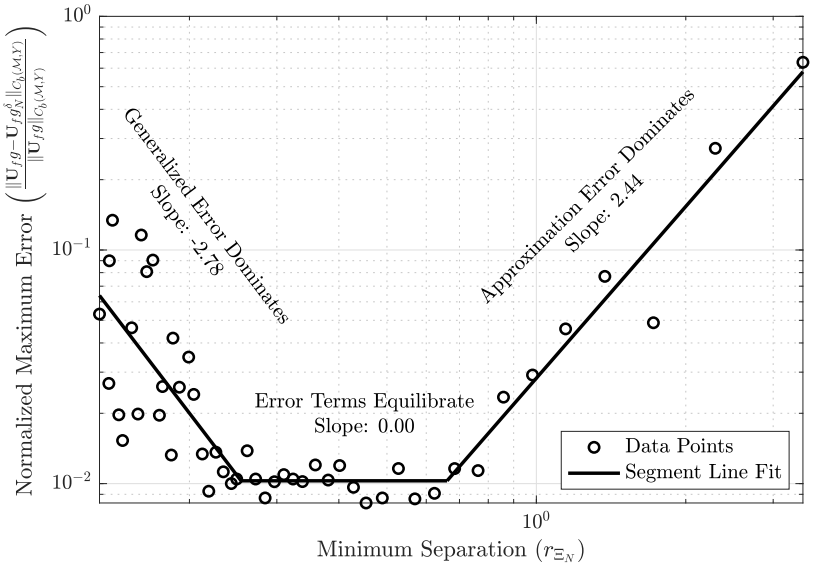

Fig. 4 Normalized maximum error graph for $\delta = 0.5\%$, using the same settings as in Figure 2

shows the normalized prediction error for the Koopman operator approximations as a function of the separation distance r_{Ξ_N}. It appears that the approximations of the Koopman operator exhibit a tradeoff behavior that is well-known in statistical and machine learning theory [35, Ch. 7] and distribution-free regression theory [20]. Figure 3 shows the condition number of the Grammian matrix as a function of r_{Ξ_N}, indicating how numerical stability is affected by the minimum separation distance. Even for this example, which has a low-dimensional state space $\mathbb{X} \triangleq \mathbb{R}^2$, the condition number grows considerably as the minimal separation decreases and the dimension increases. This is a well-documented phenomenon for many kernels, and especially exponential kernels, and numerous authors have studied methods to address the ill-conditioning that accompanies such growth [22, 23]. Fasshauer and McCourt provided an analytic method for addressing the ill-conditioning issue in [13], which is effective for small dimensions when using the Gaussian radial basis function kernel. We only note here that the same qualitative results necessarily arise in assessing the fidelity of predictions based on approximations of the Koopman operator.

Figure 4 illustrates the normalized maximum error and reflects the rate of convergence under the same experimental settings as in Figure 2. Three overlapping data sets are manually selected, and piecewise linear curve fitting is performed in the log-log scale. The fitted lines are trimmed at their intersection points, with the intermediate region constrained to have zero slope. This figure shows the typical three error regimes. The right-most portion of the plot is dominated by the approximation error term in (87). The left-most portion is

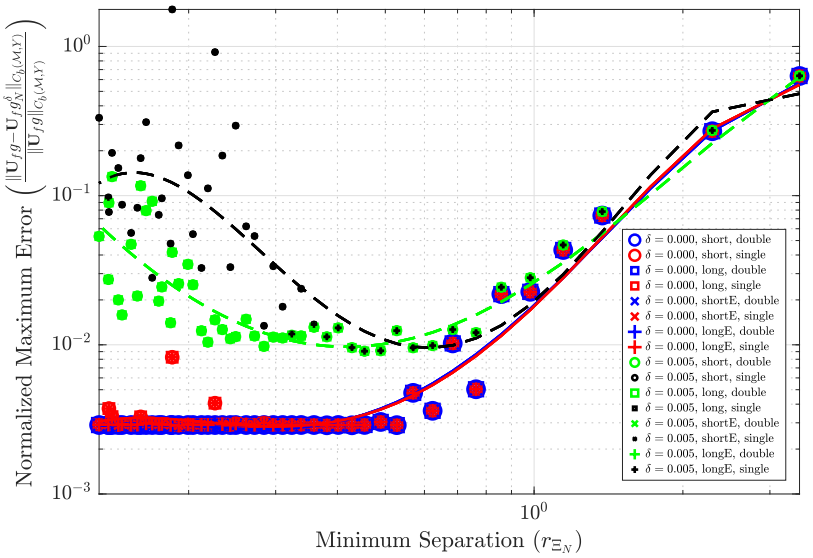

Fig. 5 Normalized maximum error graph for $\delta = 0$ and $\delta = 0.5\%$ with different numerical precision in computations. The same settings as in Figures 2 and 4 are employed

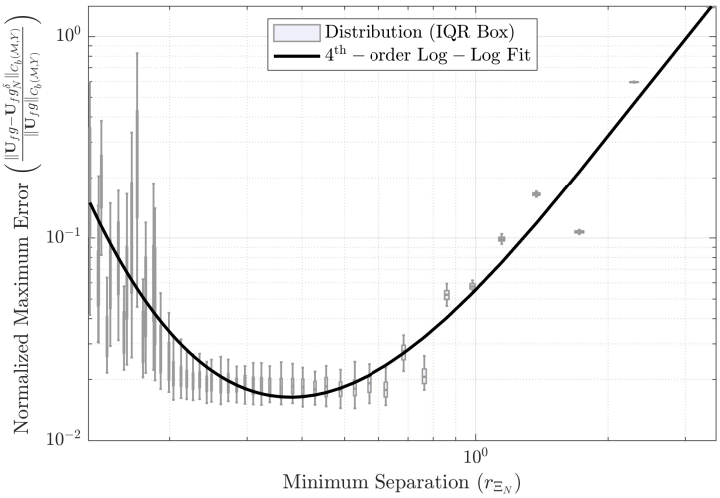

Fig. 6 Log-log plot of the normalized maximum prediction error versus the minimum separation distance r_{Ξ_N}, computed using Method 1 under zero-mean stochastic measurement uncertainty with a magnitude of 0.2%. For each r_{Ξ_N}, the corresponding boxplot shows the distribution of normalized maximum errors over 200 simulations. The solid curve represents a fourth-order polynomial fit of the average error in the log-log scale. Results are shown for the Gaussian kernel

dominated by the sample error. The central region, having the lowest error, occurs when the two error contributions are said to equilibrate.

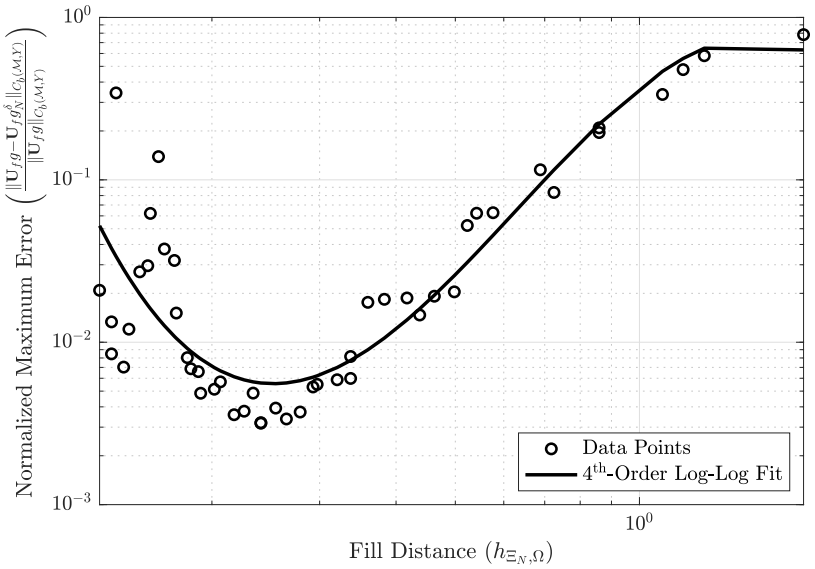

Fig. 7 Log-log plot of the normalized maximum prediction error versus the fill distance $h_{\Xi_N,\Omega}$, computed along 100 manifolds generated from initial conditions uniformly spaced along the x_2 axis from $(1, 1)$ to approximately $(1, 1.99)$, approaching the singular point at $(1, 2)$. The simulation was performed using Method 1 under deterministic sinusoidal measurement uncertainty with a magnitude of 0.01%. Each marker corresponds to the normalized maximum error, and the solid curve represents a fourth-order polynomial fit in the log-log scale. Results are shown for the Gaussian kernel (103)

Thus far, we assessed how changing the maximum perturbation δ to the sample values affects the sensitivity of the error equations. Figure 5 illustrates the effect of numerical precision. In particular, Figure 5 shows simulation results for $\delta = 0$ and $\delta = 0.5\%$, and applying the same settings as for the results in Figures 2 and 4, while varying the solver's precision. Double-precision results were generated using 64-bit calculations to compute $\alpha = \mathbb{K}^{-1}\bar{y}$, and single-precision results were generated using 32-bit calculations. We see that the influence of numerical noise in transitioning from 64-bit to 32-bit calculations can decrease the accuracy substantially during the regime when the sample error dominates. This increase in error varies with the minimum separation value. As an example, consider the typical value of the separation distance $r_{\Xi_N} = .3$ when $\delta = 0.5\%$. Thirty-two-bit calculations show a three-times larger error than 64-bit calculations. Some smaller values of the separation distance increase the error by an even larger factor. The induced numerical noise is clearly significant. We conclude that, practically speaking, we should interpret δ in the sampling contribution in the error bound (87) as capturing both the deterministic noise due to the combination of the perturbation of the sample data and the perturbation due to induced numerical noise.

Figure 6 shows simulation results for stochastic perturbations with the initial condition $(x_{1,0}, x_{2,0}) = (1, 1)$, $h = 0.025$, $M = 1865$, and $\delta_i \sim \mathcal{N}(0, 0.002^2)$.

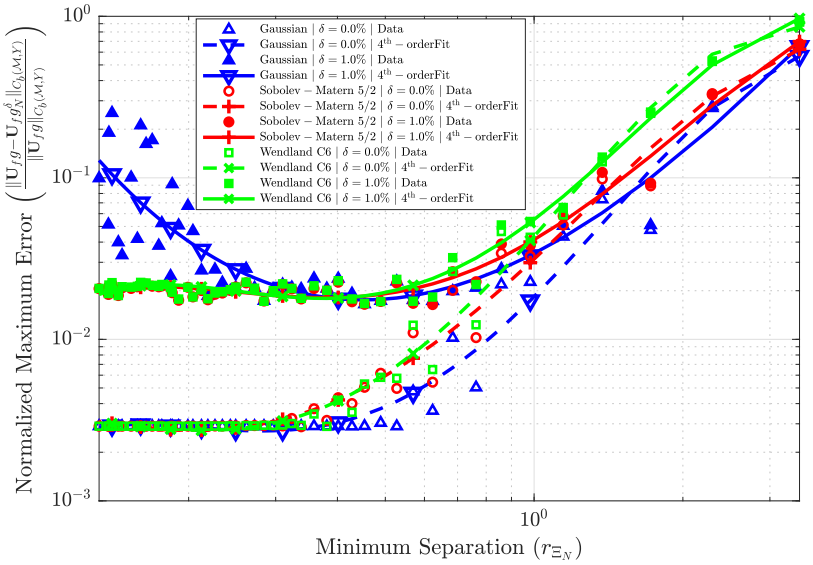

Fig. 8 Log-log plot of the normalized maximum prediction error versus the minimum separation distance r_{Ξ_N}, computed using Method 1 with three different kernel functions: Gaussian, Sobolev–Matérn 5/2, and Wendland C6; see [35, pp. 54-57]. Each marker represents the normalized maximum error computed under deterministic sinusoidal measurement uncertainty. Filled and unfilled markers represent measurement uncertainties with magnitudes of 1% and 0% (noise-free), respectively. Solid and dashed lines with unfilled markers represent fourth-order polynomial fits in the log-log scale for the 1% and noise-free cases, respectively.

It appears that the maximum prediction error ultimately increases as the minimum separation decreases and as the number of centers increases.

Figure 7 presents the results of a simulation conducted with 100 initial conditions uniformly spaced along the x_2 axis, ranging from $(x_{1,0}, x_{2,0}) = (1, 1)$ to approximately $(1, 1.99)$, approaching the singular point at $(1, 2)$. For each initial condition, a trajectory was generated using a fixed time step size $h = 0.025$, resulting in a sequence of M discrete time samples, where M decreased from 1865 to 1778 as x_2 increased. The measurement uncertainty was applied as a deterministic sinusoidal perturbation of the form $\delta_i = 10^{-4}\sin(40\pi t)$. This plot demonstrates that the trend between normalized error and center spacing in the one-dimensional case with minimum separation persists in the two-dimensional setting using the fill distance under measurement uncertainty.

To generate the kernel center set Ξ_N used in the approximation, a set of reference centers was first selected along the outermost trajectory using points equally spaced in arc length, as in the single-manifold case, to compute the two-dimensional minimum separation distance r_{Ξ_N}. Then, while traversing data points from all 100 manifolds in a random order, a new center was added to Ξ_N only if its Euclidean distance to all previously selected centers exceeded r_{Ξ_N}. To quantify the spatial coverage of the selected center set Ξ_N over the

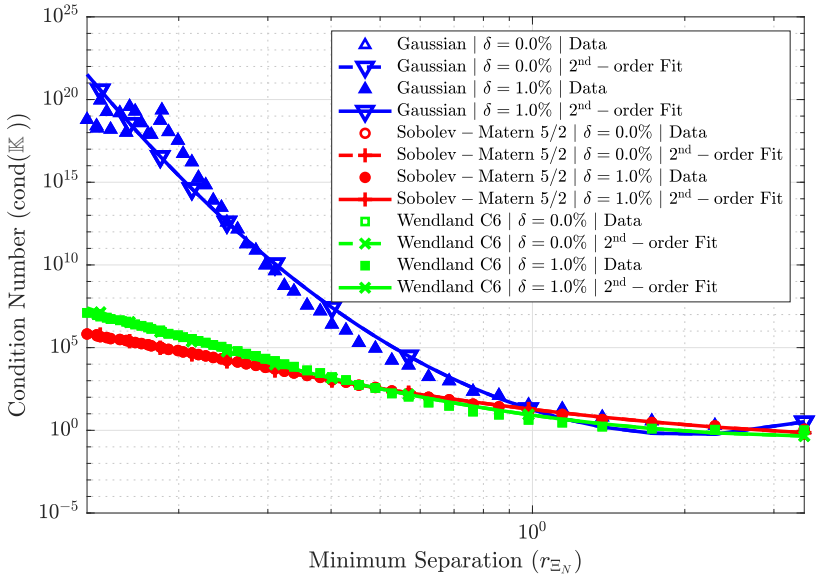

Fig. 9 Log-log plot of the condition number of the Grammian matrix $\mathbb{K}_N$ versus the minimum separation distance r_{Ξ_N}, corresponding to Method 1 with three different kernel functions: Gaussian, Sobolev–Matérn 5/2, and Wendland C6; see [35, pp. 54-57]. Each marker represents the condition number evaluated under deterministic sinusoidal measurement uncertainty. Filled and unfilled markers represent measurement uncertainties with magnitudes of 1% and 0% (noise-free), respectively. Solid and dashed lines with unfilled markers represent second-order polynomial fits in the log-log scale for the 1% and noise-free cases, respectively

manifold embedded in $\mathbb{R}^2$, the fill distance $h_{\Xi_N,\Omega}$ was recomputed as defined in (10) and used as the horizontal axis in Figure 7.

Figures 8 and 9 compare simulation results employing the Gaussian kernel (103), the Sobolev-Matérn 5/2 kernel

$$\mathfrak{K}_{5/2}(x_1,x_2) \triangleq \frac{p!}{(2p)!}\exp\left(-\sqrt{2p+1}\,\|x_1-x_2\|\right)\sum_{i=0}^{p}\binom{p}{i}\left(2\sqrt{2p+1}\,\|x_1-x_2\|\right)^{p-i},$$
$$\text{for all } (x_1,x_2)\in\mathbb{R}\times\mathbb{R}, \tag{104}$$

with $p=2$, and the Wendland C^6 kernel

$$\mathfrak{K}_{C^6}(x_1,x_2) \triangleq \begin{cases}(1-r)^8(32r^3+25r^2+8r+1), & 0\le r\le 1,\\ 0, & r>1,\end{cases} \tag{105}$$

where $r \triangleq \frac{\|x_1-x_2\|}{3}$ for all $x_1,x_2\in\mathbb{X}$. The amplification of the sampling error is smaller for the selected Sobolev-Matérn and Wendland kernels. This behavior is partly justified by the better numerical conditioning associated with these kernels, which is evident in Figure 9. For the finest spacing of centers, the condition number of the Grammian matrix $\mathbb{K}_N$ is at most $\sim O(10^7)$ for the Wendland C^6 kernel and $\sim O(10^4)$ for the Matérn-Sobolev 5/2 kernel, respectively. For the Sobolev-Matérn kernel, the sampling error just starts to

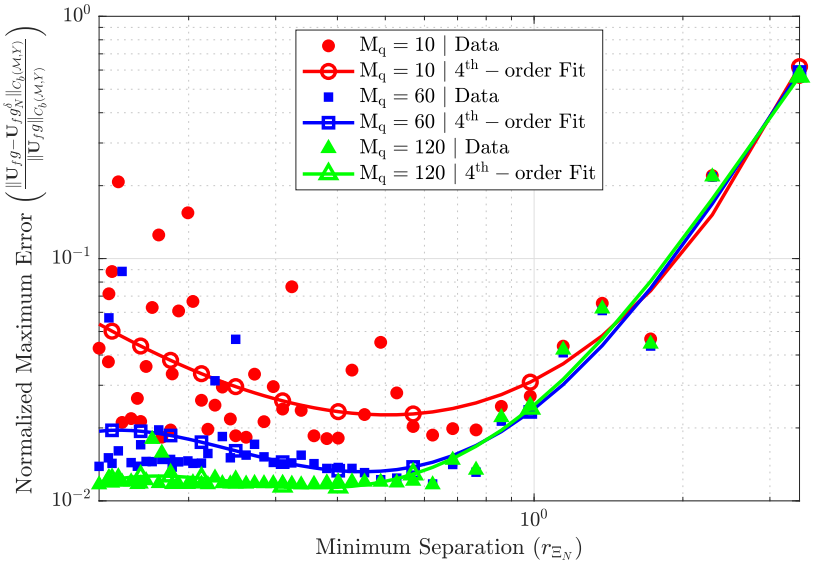

Fig. 10 Log-log plot of the normalized maximum prediction error versus the minimum separation distance r_{Ξ_N}, computed using Method 2 for three different numbers of quadrature points $M_q = 10, 60, 120$, under deterministic sinusoidal measurement uncertainty of magnitude 0.1%. Each filled marker represents the normalized maximum error computed at a given r_{Ξ_N}. Solid curves with unfilled markers represent fourth-order polynomial fits in the log-log scale for each corresponding quadrature point setting

curve upwards when the centers are most closely spaced for a noise level of $\delta = 1\%$. This behavior is considerably better than using Gaussian kernels with the same separation radius, where the condition number is $\sim O(10^{20})$. Overall, these results confirm the general trend that is so well-known when computing with reproducing kernels: smoother kernels generate approximations with higher convergence rates, but at the expense of increased condition numbers of the associated Grammian matrices. From these observations, we note that when using vRKHS to approximate Koopman operators in constructing forecasts, we should use the least regular kernel that meets the desired accuracy requirements. However, if the smoothness index of the propagation function f is unknown, applying this design criterion may be challenging.

In Figure 8, we note that, employing the Sobolev-Matérn and Wendland kernels, the pointwise error is well approximated by straight lines for larger values of the separation radius r_{Ξ_N} or, equivalently, for larger values of the fill distance for our quasiuniform samples. Using such observations to further assess the "empirical or effective regularity" of the unknown functions f or Koopman action $U_f g$ would be a fruitful topic for future study.

7.3 Method 2: Collocation with Quadrature

In this section, we employ the Galerkin approximation defined by (76). Figure 10 shows the maximum relative prediction error as a function of the minimum separation radius r_{Ξ_N}, parameterized by the number of quadrature points M_q used in (76), along the orbit with initial condition $(1, 1)$. The quadrature error increases with the number of quadrature points M_q. The quadrature error can therefore be interpreted as a source of computational noise, in addition to the noise δ_i associated with the imprecise measurements along an orbit of the system. As expected, increasing the size of the perturbation due to the computational noise term amplifies the sampling error term. Figure 11 compares results obtained by applying the method captured by (73) and the method captured by (76). In both cases, we employed a Gaussian kernel, $M_q = 60$ quadrature points, and noise levels $\delta \in \{0.0\%, 0.1\%, 1.0\%\}$. Figure 11 shows a trend similar to Figure 10, that is, for smaller values of r_{Ξ_N}, the sampling error predominates, and for larger values of r_{Ξ_N}, the approximation error dominates in the prediction error. Remarkably, both methods exhibit similar convergence rates during the approximation-error-dominant regime. However, employing (73), the approximation error is more sensitive to noise for smaller values of r_{Ξ_N}.

8 Conclusions and Future Work

Motivated by the problem of identifying the dynamics of a discrete-time nonlinear dynamical model through some measured output, this paper presented several novel theoretical results on Koopman invariance and the fidelity of deterministic approximations of the discrete Koopman operator. The results on the approximation error of the discrete Koopman operator can be used to study the error of output predictions for the next time step based on the current state. Among others, key results are Theorem 2, which provides necessary and sufficient conditions for the Koopman invariance of the pullback vRKHS $\boldsymbol{\mathcal{P}} \triangleq U_f(\boldsymbol{\mathcal{H}})$ in the vRKHS $\boldsymbol{\mathcal{H}}$ defined in terms of an operator-valued kernel, and Theorem 10, which extends the "doubling trick," a popular tool for the improvement of rates of convergence in scalar-valued RKHS, to a vector-valued RKHS setting. This paper also introduced a general constructive procedure in Section 5.3, which is based on the necessary and sufficient conditions in Theorem 2, to build vRKHS that are invariant under the Koopman operator.

The proposed results for approximating Koopman operators in a vRKHS are qualitatively similar to some well-known principles in statistical and machine learning theory. Indeed, the total prediction error can be decomposed into the sum of an approximation error term and a sampling error term. While the former typically decreases as $d(N) \to \infty$ and the dimension of the approximants increases, the latter can increase as $d(N) \to \infty$. This fact has not been noted nor emphasized in existing studies on the error of deterministic approx-

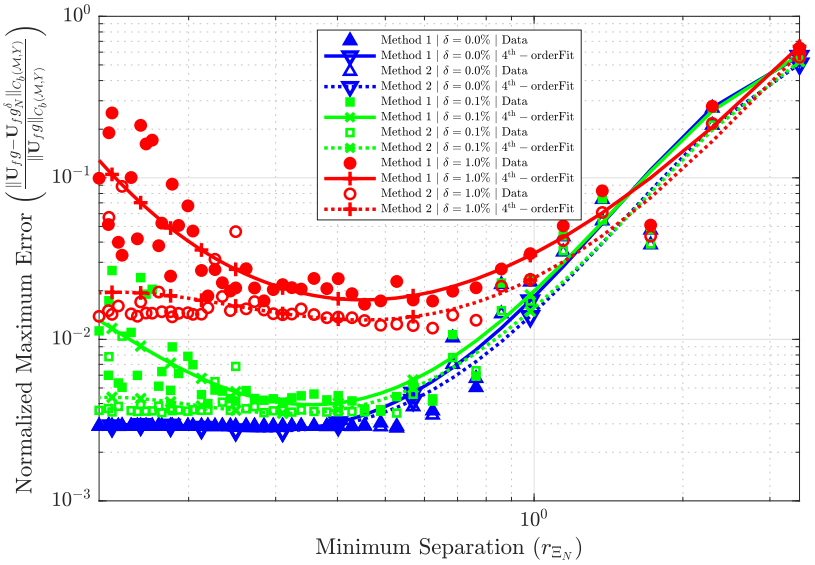

Fig. 11 Log-log plot of the normalized maximum prediction error versus the minimum separation distance r_{Ξ_N}, comparing Method 1 and Method 2 under three levels of deterministic sinusoidal uncertainty: 0%, 0.1%, and 1%. The number of quadrature points is fixed at $M_q = 60$. Each marker represents the normalized maximum error computed at a given r_{Ξ_N}. Filled markers represent Method 1, while unfilled markers represent Method 2. Solid and dotted curves with markers correspond to fourth-order polynomial fits in the log-log scale for each method and uncertainty level.

imations of Koopman operators, and it arises when observation values along a trajectory are subject to noise or are imprecisely measured.

While the bound in (87) is new in the context of approximations of Koopman operators, error bounds of this form have been studied for years in approximations of inverse problems posed in a wide collection of applications. The novelty of this paper lies in the application of the power function and the doubling trick in a vRKHS setting, which enables the general bound for error in Galerkin approximations to be applied to Koopman operators. The structure of this error bound suggests several extensions of the analysis in this paper. For instance, the overall structure of the classic bound in Theorem 8 may serve as the starting point for refinements based on the regularization.

The theory of regularization of inverse problems accounts for the sample error, which, for compact operators, grows with the dimension of the approximation space $\boldsymbol{\mathcal{H}}_N$ as $d(N) \to \infty$. This coupling between sampling error and dimensions of the approximation space can be accomplished in various ways. One way is to couple the dimension $d(N)$ and the number of input-output samples M to equilibrate the sample and approximation errors, and thereby ensure that the final approximation "lies at the bottom" of the error curve in Figure 4. Equivalently, methods like Tikhonov regularization use a regularization parameter, which depends on the dimension $d(N)$ and the number of

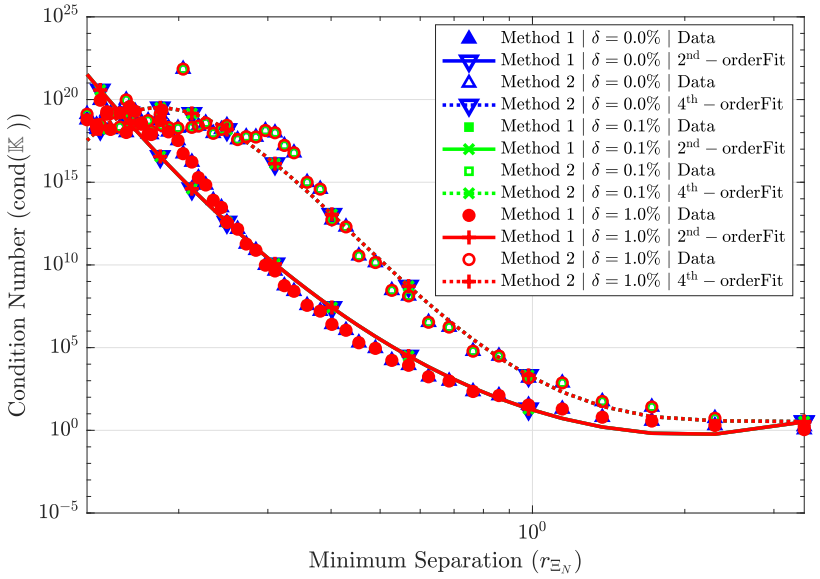

Fig. 12 Log-log plot of the condition number of the Grammian matrix versus the minimum separation distance r_{Ξ_N}, comparing Method 1 and Method 2 under three levels of deterministic sinusoidal measurement uncertainty: 0%, 0.1%, and 1%. The number of quadrature points is fixed at $M_q = 60$. Each marker represents the condition number evaluated at a given r_{Ξ_N}. Filled markers represent the results of Method 1, and unfilled markers represent the results of Method 2. Solid and dotted curves with markers correspond to second-order and fourth-order polynomial fits in the log-log scale for Method 1 and Method 2, respectively

samples M, to generate approximations that "lie at the bottom" of the error curve.

In this paper, we primarily focused on the baseline error bounds in (87). This result can be the starting point for understanding and designing subsequent approximations that use regularization. The analysis of the choice of the regularization parameter and the subsequent associated error analysis exceeds the scope of this initial paper, focused on deterministic methods. It is noteworthy, however, that the choice of the regularization parameter is treated in [37, 39, 44], which study KME for stochastic dynamics. These error bounds are typically stated in their strongest form for function norms associated with spaces that depend on rates of decay of eigenvalues of the integral operator that defines the KME method. Correspondingly precise regularization methods for approximations of deterministic Koopman operators remain an open question.

For future research, the baseline error bound in (87) could be further improved by studying how this worst-case error bound scales with the dimension n of the state space of the dynamics. General considerations regarding how some errors of approximation in vRKHS scale with the dimension of the state space are discussed in [35]. In the following, we make some observations that could motivate or guide finer studies of how this bound scales with n. For

interpolation in vRKHS, the scaling of errors depends to a large degree on whether the samples are quasi-uniform over an open connected set $\Omega \subset \mathbb{R}^n$, or whether they are supported on some ℓ-dimensional submanifold $\mathcal{M} \subset \mathbb{R}^n$. We outline this fact just considering scalar-valued RKHSs $\mathcal{K}$ defined in terms of kernels $\mathfrak{K}$ and interpolation from functions in $\mathcal{K}_N \triangleq \text{span}\{\mathfrak{K}_{\xi_i} \mid \xi_i \in \Xi_N\}$. Choosing to construct approximations from the class of Wendland or Matérn kernels, it is then possible to derive bounds on the L^∞ norm of the interpolation error that scale like $h^s_{\Xi_N,\Omega}$, where the smoothness index $s > 0$ depends on the specific kernel that defines $\mathcal{K} = \mathcal{K}(\Omega, \mathbb{R})$; for details, see [35, 55]. However, for quasi-uniform samples in a parallelopiped $\Omega \subset \mathbb{R}^n$, the number of centers N scales like $N(\Omega) \sim 1/h^n_{\Xi_N,\Omega}$, which means that the L^∞ norm of the error in approximating f by its interpolant $f_N \in \mathcal{K}_N$ scales like

$$\|f - f_N\|_{L^\infty(\Omega,\mathbb{R})} \sim \left(\frac{1}{N(\Omega)}\right)^{s/n} \sim h^s_{\Xi_N,\Omega},$$

where $f, f_N \in \mathcal{K}(\Omega, \mathbb{R})$. This bound shows that the number $N(\Omega)$ of basis functions needed for some prescribed error of approximation grows rapidly with n. However, if the samples are restricted to an ℓ-dimensional submanifold $\mathcal{M} \subset \mathbb{R}^n$, and the reproducing kernel initially defined on all of Ω is restricted to $\mathcal{M} \subset \mathbb{R}^n$, the approximation error can be captured by

$$\|f - f_N\|_{L^\infty(\mathcal{M},\mathbb{R})} \sim \left(\frac{1}{N(\mathcal{M})}\right)^{\bar{s}/\ell} \sim h^{\bar{s}}_{\Xi_N,\mathcal{M}}.$$

for $f, f_I \in \mathcal{K}(\mathcal{M}, \mathbb{R})$, with $\bar{s} < s$ capturing a reduced measure of smoothness [15]. Remarkably, if the system's dynamics are supported on a submanifold, then the dimension of the submanifold limits the error bound. These observations suggest that further studies of how the worst-case error scales with the dimension of the state space n can be improved for dynamics that actually are supported on a submanifold of dimension $\ell < n$. To the authors' knowledge, this fact has never been studied in approximations of Koopman operators.

In the scalar-valued RKHS setting, references [47, 49] show that the error bound presented in Theorem 11 can be further refined with a careful analysis of the error due to the use of quadrature points as in the approach captured by (76). This result is based on the regularity of the integrand. In the future, this analysis can be extended to the vRKHS setting to enable balancing the contributions of the experimental and numerical noise to the total error in the forecast predictions.

Finally, the proposed numerical simulations show how the approach captured by (76) is substantially less sensitive to sampling noise than the approach captured by (73) for smaller values of the separation radius among kernel centers. No explanation was given in this paper for this behavior, and a careful analytical explanation for this phenomenological observation would be a valuable extension of this work.

Authors' Contribution Statement

Soonyong Yang and Haoran Wang were responsible for designing and implementing the proposed numerical simulations. Dr. Kurdila was responsible for the idea underlying this paper and the outline of the key results. Dr. L'Afflitto was primarily responsible for detailing the proofs. Dr. Paruchuri was primarily responsible for the theoretical aspects concerning native space theory. Dr. Kamalapurkar and Dr. Rosenfeld were primarily responsible for the theoretical aspects concerning Koopman operator theory.

Funding Declaration

This work was partly supported by the US National Science Foundation through the grant no. 2137159, and the US Department of the Navy through the grant no. N004212520003

References

1. Adams, R.A., Fournier, J.: Sobolev spaces, vol. 140. Elsevier, Amsterdam (2003)
2. Alla, A., Kutz, J.N.: Nonlinear model order reduction via dynamic model decomposition. SIAM Journal on Scientific Computing **39**(5), B778–B796 (2017). DOI {10.1137/16M1059308}
3. Bollt, E.M., Li, Q., Dietrich, F., Kevrekidis, I.: On matching, and even rectifying, dynamical systems through Koopman operator eigenfunctions. SIAM Journal on Applied Dynamical Systems **17**(2), 1925–1960 (2018). DOI {10.1137/17M116207X}
4. Brunton, S.L., Brunton, B.W., Proctor, J.L., Kutz, J.N.: Koopman invariant subspaces and finite linear representations of nonlinear dynamical systems for control. PLOS ONE **11**(2), 1–19 (2016). DOI 10.1371/journal.pone.0150171
5. Carmeli, C., De Vito, E., Toigo, A., Umanitá, V.: Vector valued reproducing kernel Hilbert spaces and universality. Analysis and Applications **8**(01), 19–61 (2010)
6. Carswell, B.J., MacCluer, B.D., Schuster, A.: Composition operators on the fock space. Acta Sci. Math.(Szeged) **69**(3-4), 871–887 (2003)
7. Colbrook, M.J., Townsend, A.: Rigorous data-driven computation of spectral properties of Koopman operators for dynamical systems. Communications on Pure and Applied Mathematics **77**(1), 221–283 (2024)
8. De Vito, E., Rosasco, L., Toigo, A.: Learning sets with separating kernels. Applied and Computational Harmonic Analysis **37**(2), 185–217 (2014)
9. Degennaro, A.M., Urban, N.M.: Scalable extended dynamic mode decomposition using random kernel approximation. SIAM Journal on Scientific Computing **41**(3), A1482–A1499 (2019). DOI {10.1137/17M115414X}
10. Drmac, Z., Mezic, I., Mohr, R.: Data-driven modal decompositions: Analysis and enhancements. SIAM Journal on Scientific Computing **40**(4), A2253–A2285 (2018). DOI {10.1137/17M1144155}
11. Dunford, N., Schwartz, J.T.: Linear operators, part 1: General theory. John Wiley & Sons, New York (1988)
12. Engl, H.W., Hanke, M., Neubauer, A.: Regularization of inverse problems, vol. 375. Springer Science & Business Media, Berlin (1996)
13. Fasshauer, G.E., McCourt, M.J.: Stable evaluation of Gaussian radial basis function interpolants. SIAM Journal on Scientific Computing **34**(2), A737–A762 (2012)
14. Fuselier, E., Wright, G.B.: Scattered data interpolation on embedded submanifolds with restricted positive definite kernels: Sobolev error estimates. SIAM Journal on Numerical Analysis **50**(3), 1753–1776 (2012)

15. Fuselier, E., Wright, G.B.: Scattered data interpolation on embedded submanifolds with restricted positive definite kernels: Sobolev error estimates. SIAM Journal on Numerical Analysis **50**(3), 1753–1776 (2012)
16. Gelss, P., Klus, S., Eisert, J., Schuette, C.: Multidimensional approximation of nonlinear dynamical systems. Journal of Computational and Nonlinear Dynamics **14**(6) (2019). DOI {10.1115/1.4043148}
17. Giannakis, D.: Data-driven spectral decomposition and forecasting of ergodic dynamical systems. Applied and Computational Harmonic Analysis **47**(2), 338–396 (2019). DOI {10.1016/j.acha.2017.09.001}
18. Giannakis, D., Ourmazd, A., Slawinska, J., Zhao, Z.: Spatiotemporal pattern extraction by spectral analysis of vector-valued observables. Journal of Nonlinear Science **29**(5), 2385–2445 (2019)
19. Griebel, M., Oswald, P.: Hilbert Space Splittings and Iterative Methods. Springer, Berlin (2024)
20. Gyorfi, L., Kohler, M., Krzyzak, A., Walk, H.: A Distribution-Free Theory of Nonparametric Regression. Springer, Berlin (2002)
21. Hairer, E., Lubich, C., Wanner, G.: Geometric Numerical Integration: Structure-Preserving Algorithms for Ordinary Differential Equations. Springer, Berlin (2006)
22. Hangelbroek, T., Narcowich, F., Ward, J.: Kernel approximation on manifolds I: Bounding the Lebesgue constant. SIAM J. Math. Anal. **42**, 1732–1760 (2010)
23. Hangelbroek, T., Narcowich, F., Ward, J.: Polyharmonic and related kernels on manifolds: Interpolation and approximation. Foundations of Computational Mathematics **12**, 625–670 (2012)
24. Hangelbroek, T., Rieger, C.: Extending error bounds for radial basis function interpolation to measuring the error in higher order Sobolev norms. Mathematics of Computation **94**(351), 381–407 (2025)
25. Hua, J.C., Noorian, F., Moss, D., Leong, P.H.W., Gunaratne, G.H.: High-dimensional time series prediction using kernel-based Koopman mode regression. Nonlinear Dynamics **90**(3), 1785–1806 (2017)
26. Kirsch, A.: An introduction to the mathematical theory of inverse problems, vol. 120. Springer, Berlin (2021)
27. Klus, S., Koltai, P., Schütte, C.: On the numerical approximation of the Perron-Frobenius and Koopman operator. Journal of Computational Dynamics **3**(1), 51–79 (2016). DOI 10.3934/jcd.2016003
28. Klus, S., Nuske, F., Koltai, P., Wu, H., Kevrekidis, I., Schuette, C., Noe, F.: Data-Driven Model Reduction and Transfer Operator Approximation. Journal of Nonlinear Science **28**(3), 985–1010 (2018). DOI {10.1007/s00332-017-9437-7}
29. Köhne, F., Philipp, F.M., Schaller, M., Schiela, A., Worthmann, K.: L^∞-error bounds for approximations of the Koopman operator by kernel extended dynamic mode decomposition. SIAM Journal on Applied Dynamical Systems **24**(1), 501–529 (2025)
30. Korda, M., Mezic, I.: On convergence of extended dynamic mode decomposition to the Koopman operator. Journal of Nonlinear Science **28**, 687–710 (2018)
31. Korda, M., Putinar, M., Mezić, I.: Data-driven spectral analysis of the Koopman operator. Applied and Computational Harmonic Analysis **48**(2), 599–629 (2020). DOI https://doi.org/10.1016/j.acha.2018.08.002
32. Kostic, V.R., Novelli, P., Maurer, A., Ciliberto, C., Rosasco, L., Pontil, M.: Learning dynamical systems via Koopman operator regression in reproducing kernel Hilbert spaces. In: A.H. Oh, A. Agarwal, D. Belgrave, K. Cho (eds.) Advances in Neural Information Processing Systems (2022)
33. Kress, R., Maz'ya, V., Kozlov, V.: Linear integral equations, vol. 82. Springer, Berlin (1989)
34. Kurdila, A., Bobade, P.: Koopman theory and linear approximation spaces. Arxiv (2018). DOI arXiv:1811.10809
35. Kurdila, A.J., L'Afflitto, A., Burns, J.A.: Data-Driven, Nonparametric, Adaptive Control Theory. Springer, London, UK (2025)
36. Macesic, S., Crnjari-Zic, N., Mezic, I.: Koopman Operator Family Spectrum for Nonautonomous Systems. SIAM Journal on Applied Dynamical Systems **17**(4), 2478–2515 (2018). DOI {10.1137/17M1133610}

37. Meunier, D., Shen, Z., Mollenhauer, M., Gretton, A., Li, Z.: Optimal rates for vector-valued spectral regularization learning algorithms. Advances in Neural Information Processing Systems **37**, 82514–82559 (2024)
38. Mezić, I.: On numerical approximations of the koopman operator. Mathematics **10**(7) (2022). DOI 10.3390/math10071180
39. Mollenhauer, M., Koltai, P.: Nonparametric approximation of conditional expectation operators. arXiv preprint arXiv:2012.12917 (2020)
40. Nashed, M.Z., Wahba, G.: Convergence rates of approximate least squares solutions of linear integral and operator equations of the first kind. Mathematics of Computation **28**(125), 69–80 (1974)
41. Nashed, M.Z., Wahba, G.: Generalized inverses in reproducing kernel spaces: An approach to regularization of linear operator equations. SIAM Journal on Mathematical Analysis **5**(6), 974–987 (1974)
42. Nashed, M.Z., Walter, G.G.: General sampling theorems for functions in reproducing kernel Hilbert spaces. Mathematics of Control, Signals and Systems **4**(4), 363–390 (1991)
43. Nüske, F., Peitz, S., Philipp, F., Schaller, M., Worthmann, K.: Finite-data error bounds for Koopman-based prediction and control. Journal of Nonlinear Science **33**(1), 14 (2023)
44. Park, J., Muandet, K.: A measure-theoretic approach to kernel conditional mean embeddings. Advances in neural information processing systems **33**, 21247–21259 (2020)
45. Paulsen, V.I., Raghupathi, M.: An introduction to the theory of reproducing kernel Hilbert spaces, vol. 152. Cambridge University Press, Cambridge (2016)
46. Philipp, F.M., Schaller, M., Worthmann, K., Peitz, S., Nüske, F.: Error bounds for kernel-based approximations of the Koopman operator. Applied and Computational Harmonic Analysis **71**, 101657 (2024)
47. Powell, N., Bouland, A., Burns, J.A., Kurdila, A.J.: Convergence rates for approximations of deterministic Koopman operators via inverse problems. In: Conference on Decision and Control. IEEE, Singapore (2023). DOI 10.1109/CDC49753.2023.10383441
48. Powell, N., Liu, B., Kurdila, A.J.: Koopman methods for estimation of motion over unknown, regularly embedded submanifolds. In: American Control Conference, pp. 2584–2591. IEEE (2022)
49. Powell, N., Paruchuri, S.T., Bouland, A., Niu, S., Kurdila, A.: Invariance and approximation of Koopman operators in native spaces. In: American Control Conference, pp. 2871–2878. IEEE (2024)
50. Rosenfeld, J.A., Kamalapurkar, R.: Singular dynamic mode decomposition. SIAM Journal on Applied Dynamical Systems **22**(3), 2357–2381 (2023)
51. Saitoh, S., Sawano, Y.: Theory of reproducing kernels and applications. Springer, Berlin (2016)
52. Stepaniants, G.: Learning partial differential equations in reproducing kernel Hilbert spaces. Journal of Machine Learning Research **24**(86), 1–72 (2023)
53. Vito, E.D., Mucke, N., Rosasco, L.: Reproducing kernel hilbert spaces on manifolds: Sobolev and diffusion spaces. arXiv.org pp. 1–31 (2019). DOI arXiv:1905.10913v1
54. Weidmann, J.: Linear Operators in Hilbert Spaces. Springer, Berlin (1980)
55. Wendland, H.: Scattered data approximation, vol. 17. Cambridge University Press, Cambridge (2004)
56. Williams, M.O., Kevrekidis, I.G., Rowley, C.W.: A data-driven approximation of the Koopman operator: Extending dynamic mode decomposition. Journal of Nonlinear Science **25**(6), 1307–1346 (2015). DOI {10.1007/s00332-015-9258-5}
57. Wittwar, D.: Approximation with matrix-valued kernels and highly effective error estimators for reduced basis approximations. Ph.D. thesis, Dissertation, Stuttgart, Universität Stuttgart, 2021 (2022)
58. Wittwar, D., Santin, G., Haasdonk, B.: Interpolation with uncoupled separable matrix-valued kernels. Dolomites Research Notes on Approximation **11**, 23–39 (2018)