THE OCCUPATION KERNEL METHOD FOR NONLINEAR SYSTEM IDENTIFICATION*

JOEL A. ROSENFELD†, BENJAMIN RUSSO‡, RUSHIKESH KAMALAPURKAR§, AND TAYLOR T. JOHNSON¶

Abstract. This manuscript presents a novel approach to nonlinear system identification leveraging densely defined Liouville operators and a new "kernel" function that represents an integration functional over a reproducing kernel Hilbert space (RKHS) dubbed an occupation kernel. The manuscript thoroughly explores the concept of occupation kernels in the context of RKHSs of continuous functions, and studies Liouville operators over RKHSs. The combination of these two concepts facilitates embedding of a dynamical system into a RKHS, where function theoretic tools may be leveraged for the examination of such systems. This framework allows for trajectories of a nonlinear dynamical system to be treated as a fundamental unit of data, and results in a nonlinear system identification (sysID) routine that generalizes weak least-squares sysID approaches. Numerical experiments indicate that the developed approach identifies parameters of nonlinear dynamical systems accurately, while also exhibiting robustness to noise.

1. Introduction. Consider a dynamical system $\dot{x} = f(x)$, where $x : [0, T] \to \mathbb{R}^n$ is the system state and $f : \mathbb{R}^n \to \mathbb{R}^n$ are Lipschitz continuous dynamics. Dynamical systems are prevalent in the sciences, such as engineering [18, 9, 19], biology [1, 12], neuroscience [16], physics [42], and mathematics [8, 29]. However, in many cases even physically motivated dynamical systems can have unknown parameters (i.e. a gray box model), such as mass and length of mechanical components, or the dynamics may be completely unknown (i.e. a black box model) [25]. In such cases, system identification methods are leveraged to estimate the dynamics of the system based on data generated by the system [25].

While many classical tools are available for system identification for linear dynamics using the impulse response and Fourier and Laplace transforms of the dynamical system, the identification of nonlinear systems proves more challenging as nonlinear-

*A subset of the results in this paper appeared in the proceedings of the 2019 IEEE Conference on Decision and Control [34]. This research was supported by the Air Force Office of Scientific Research (AFOSR) under contract numbers FA9550-20-1-0127, FA9550-16-1-0246, FA9550-18-1-0122, and FA9550-21-1-0134 the Office of Naval Research (ONR) under contract number N00014-18-1-2184, the Air Force Research Laboratories (AFRL) under contract number FA8651-19-2-0009 and the Defense Advanced Research Projects Agency (DARPA) under contract number FA8750-23-C-0518, and the National Science Foundation under award numbers 2027999, 2027976, and 2028001. Any opinions, findings and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the sponsoring agencies. The authors would like to express our appreciation for Hongliang Wang for his careful reading of the manuscript and bringing to our attention numerous errors. The following YouTube playlist discusses the contents of this manuscript: https://youtube.com/playlist?list=PLldiDnQu2phtKtASlUrmiPBH705xmRdUC. This manuscript has been authored by UT-Battelle, LLC, under contract DE-AC05-00OR22725 with the US Department of Energy (DOE). The US government retains and the publisher, by accepting the article for publication, acknowledges that the US government retains a nonexclusive, paid-up, irrevocable, worldwide license to publish or reproduce the published form of this manuscript, or allow others to do so, for US government purposes. DOE will provide public access to these results of federally sponsored research in accordance with the DOE Public Access Plan (http://energy.gov/downloads/doe-public-access-plan)

†Department of Mathematics and Statistics, University of South Florida, Tampa, FL 33620 USA

‡Computer Science and Mathematics Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831, United States of America

§Department of Mechanical and Aerospace Engineering, Oklahoma State University, Stillwater, OK 74078 USA

¶Institute for Software Integrated Systems, Department of Computer Science, Vanderbilt University, Nashville, TN 37212 USA

ities may manifest in a variety of ways, and linear transform methods for general nonlinear systems are unavailable [2, 25].

To address these challenges a variety of nonlinear system identification methods have been developed, such as NARMAX methods [2], Volterra series [14], Lyapunov methods [26], and neural networks [25]. Recent developments in nonlinear system identification include reasoning over infinite-dimensional function spaces via kernel methods [30, 7, 3, 39] and dynamic mode decompositions (DMDs) and their connection with the Koopman operator [5, 15, 43, 22]. However, given the rich variety of nonlinear systems, there is no modal approach to resolving the system identification problem for nonlinear systems [25].

One technical challenge that arises in many of the system identification methods described above is the estimation of the state derivative [5, 26]. Frequently only the output trajectory is available and numerical estimation methods are employed to obtain samples of the state and the state derivative. Unfortunately, numerical state derivative estimates are prone to error, and introduce an artificial noise component that requires additional filtering [5].

In an online parameter estimation context, [26] leveraged the technique of integral concurrent learning, where state derivative estimates were replaced with integrals of the state. Therein it was demonstrated that the parameters were more precisely estimated via the integral concurrent learning method than by methods using state derivative estimates. Moreover, in the online setting the parameter estimation error was more robust to noise under the integral concurrent learning method [26].

The present manuscript develops a method to compute projection of the dynamics of a system onto the span of a collection of basis functions using samples of the system trajectory. The projection is derived for a class of dynamical systems that arise as symbols for densely defined Liouville operators over a RKHS (RKHS). The developed method also provides a collection of constraints that can be leveraged for sparse identification routines, such as the SINDy algorithm [5].

Specifically, the method presented in Section 8 leverages novel kernel techniques presented in Section 5, where the concept of occupation kernels is introduced along side that of densely defined Liouville operators. Occupation kernels are a generalization of occupation measures, which have been used in dynamical systems theory and optimal control based largely on the seminal work of [20]. The present manuscript links the theory of occupation measures to function theory by examining integration functionals over RKHSs rather than the Banach spaces of continuous functions.

While an occupation measure is a member of the dual of a Banach space, an occupation kernel is a function that resides in the RKHS. Moreover, the representation of a trajectory as an occupation kernel over a RKHS changes with the selection of RKHS, which allows for different aspects of the trajectory to be emphasized. In particular, an inner product on the space of dynamical systems may be determined given observed system trajectories through a combination of densely defined operators over a RKHS and the occupation kernel corresponding to the trajectory. In contrast, due to the limited availability of computational tools for measures, the study of occupation measures has been limited to polynomials in both the dynamics of the dynamical systems as well as the test functions leveraged to provide constraints on the occupation measures themselves.

The contributions of this manuscript are presented below.

- The concept of Liouville operators is integrated with the theory of RKHSs to yield a representation of nonlinear dynamical systems in a Hilbert space setting.

- Occupation measures are generalized to occupation kernels, where a trajectory is represented inside a Hilbert space as a function.
- An inner product on dynamical systems that give rise to densely defined Liouville operators over RKHSs is developed in Section 7 via adjoints applied to occupation kernels.
- Through a factorization of the Gram matrix for projection within the inner product space, a collection of constraints is developed for a system identification method, which is presented in Section 8. These constraints use more general test functions than polynomials, which is an advantage that arises in the use of occupation kernels over occupation measures. In particular, the test functions under consideration correspond to feature maps for the underlying reproducing kernel.
- The pre-inner product and the inner product developed in Section 7 yield unique insights for the system identification problem, such as the possibility to improve numerical stability of the estimation algorithm using a Gram-Schmidt process, see Section 8.2.

The manuscript is organized as follows. Preliminaries necessary for the development of occupation kernels and densely defined Liouville operators are presented in Section 4, and the densely defined Liouville operators and occupation kernels themselves are introduced in Section 5, with commentry on numerical estimation of occupation kernels in Section 6. In Section 7, these tools are then turned towards the development of an inner product on nonlinear dynamical systems, where the parameters resolving a system identification routine arise as the coefficients for the projection of a dynamical system onto its basis functions. Specifically, through a factorization of the Gram matrix, a collection of linear constraints on the parameters of the dynamics in Section 8 arise naturally. These constraints may be compared with [5], where state derivatives are replaced via a collection of integral constraints. Section 6 examines the convergence properties of occupation kernels associated with various numerical methods, while Section 9 demonstrates a robustness to noise of the samples used in Section 8. Finally, the system identification approach is then examined through a collection of numerical experiments in Section 10 and the experiments are discussed in Section 11.

2. Problem formulation. In a gray box system identification setting, the system dynamics, $f : \mathbb{R}^n \to \mathbb{R}^n$, is parameterized in terms of a collection of basis functions, $Y_i : \mathbb{R}^n \to \mathbb{R}^n$ for $i = 1, \ldots, M$, as

$$\dot{x} = f(x) = \sum_{i=1}^{M} \theta_i Y_i(x). \tag{2.1}$$

Let $X \subset \mathbb{R}^n$ be compact, H be a RKHS of continuous functions over X, and $0 < T \in \mathbb{R}$. The goal of system identification is to determine values of the parameters, θ_i for $i = 1, \ldots, M$, such that (2.1) may be used to reproduce a given a collection of continuous trajectories, $\{\gamma_j : [0, T] \to X\}_{j=1}^{N}$.

The identification method needs to be data-efficient and robust to process and sensor noise. In this paper, an operator theoretic generalization of a derivative-free method similar to integral least squares [26] and weak-SINDy [24] is developed to solve the identification problem. The generalization is realized by developing the theory of occupation kernels and Liouville operators and using it to formulating a novel inner product between dynamics that are symbols of densely defined Liouville operators. In addition to the system identification algorithm, the inner product provides a new

measure of distance between two dynamical systems and the Gram-Schmidt process developed in Section 8.2.

3. System Identification from the Operator Theoretic Viewpoint. The operator theoretic viewpoint of this article offers not only practical benefits such as the Gram Schmidt process in Section 8.2 and the improved performance demonstrated in Experiment 3, but also avenues towards new insights into and new approaches for data-driven learning. For example, extended dynamic mode decomposition [45] encodes the action of the dynamics as a Koopman (composition) operator and then builds finite rank representations. Kernelized support vector machines [41] use a feature map to send data points to observables in a RKHS. The success of these techniques is due, in part, to transformation of nonlinear problems into linear problems over Hilbert spaces, and indicates that the operator theoretic viewpoint is valuable.

In this section, we offer a generalized framework for parameter identification schemes along with a *translation* between the operator theoretic viewpoint presented in this paper and a more conventional weak least-squares viewpoint. It should be noted that most of the material presented below first appeared in [36].

3.1. Parameter Identification in State Space Models. An overarching assumption in parameter identification techniques is that f is a linear combination of a finite set of functions $\{Y_i\}_{i=1}^M$, i.e., $f(\cdot) = \sum_{i=1}^M \theta_i Y_i(\cdot), \quad \theta_i \in \mathbb{R}, \quad i = 1, \dots, M$. The goal is to estimate the weights $\{\theta_i\}_{i=1}^M$ from data that contains (part of) the solution x to the underlying system.

3.1.1. The Least-squares/SINDy Method. Given $\{\gamma(t_k)\}_{k=1}^K$ at time steps $\{t_k\}_{k=1}^K$, the least squares method seeks weights $\theta := \begin{pmatrix}\theta_1 & \cdots & \theta_M\end{pmatrix}^\top$ that satisfy $\dot\gamma(t_k) = \sum_{i=1}^M \theta_i Y_i(\gamma(t_k))$, $k = 1, \dots, K$, which is then formulated as a linear system

$$\boldsymbol{b} = \boldsymbol{G}\theta \quad \text{with entries} \quad \boldsymbol{b}_k = \dot\gamma(t_k), \quad \text{and} \quad \boldsymbol{G}_{k,i} = Y_i(\gamma(t_k)). \tag{3.1}$$

The SINDy method in [5] makes a second fundamental assumption that the underlying dynamics are sparsely represented in the correct choice of basis, and as such, the above linear system is solved using a sparsity enforcing solver such as LASSO [44].

3.1.2. The Weak Least-squares/Weak-SINDy Method. The weak-SINDy method proposed in [24, 23] introduces a class of *test functions* $\{\psi_k : [0,T] \to \mathbb{R}\}_{k=1}^K$ in addition the basis functions $\{Y_i\}_{i=1}^M$ used in the SINDy method. We refer to $\{Y_i\}_{i=1}^M$ as a *projection basis* and $\{\psi_k\}_{k=1}^K$ as a *test function basis*. Given $\gamma(t)$, $\forall t \in [0,T]$, the weak-SINDy method finds the weights $\{\theta_i\}_{i=1}^M$ by solving

$$\langle \dot\gamma,\ \psi_k\rangle_{L^2} = \left\langle \textstyle\sum_{i=1}^M \theta_i Y_i(\gamma(\cdot)),\ \psi_k \right\rangle_{L^2}, \quad k = 1, \dots, K, \tag{3.2}$$

where $\langle \cdot,\ \cdot \rangle_{L^2}$ denotes the L^2 inner product on $[0,T]$. In matrix form, Equation 3.2 is written as

$$\boldsymbol{b} = \boldsymbol{G}\theta \quad \text{with entries} \quad \boldsymbol{b}_k = \langle \dot\gamma,\ \psi_k\rangle_{L^2} \quad \text{and} \quad \boldsymbol{G}_{k,i} = \langle Y_i(\gamma(\cdot)),\ \psi_k\rangle_{L^2}, \tag{3.3}$$

where the entries of $\boldsymbol{b}$ can be computed using integration by parts, i.e.,

$$\langle \dot\gamma,\ \psi_k\rangle_{L^2} = -\left\langle \gamma,\ \dot\psi_k \right\rangle_{L^2}, \tag{3.4}$$

under the assumption that $\{\psi_k\}_{k=0}^K$ have compact support in $[0,T]$.

The derivatives of the test functions, $\dot{\psi_k}$, are often known a priori, and hence, the weak-SINDy method avoids direct computations of $\dot{\gamma}$. For example, with $\psi_k(t) = t$ for all $t \in [0,T]$ and $k = 1, \cdots, K$, and a collection of N trajectories $\{\gamma_j : [0,T] \to X\}_{j=1}^N$, one recovers the integral least squares (ILS) problem[1]

$$\theta^* = \min_\theta \sum_{j=1}^{N} \left\| \gamma_j(T) - \gamma_j(0) - \sum_{i=1}^{M} \theta_i \int_0^T Y_i(\gamma_j(t))dt \right\|_2^2 . \tag{3.5}$$

In this paper, ILS/weak-SINDy methods are generalized by posing the system identification problem as a projection problem in a RKHS using the so-called occupation kernels. The ILS formulation is recovered when the underlying RKHS is the one corresponding to the linear kernel $K(x,y) = x^\top y$.

3.1.3. The Occupation Kernel Method. The basic form of the operator-theoretic technique developed in this article is similar to weak-SINDy. Given $\gamma(t)$, $\forall t \in [0,T]$ and a set of test functions $\{\psi_k : X \to \mathbb{R}\}_{k=1}^K$, the occupation kernel method solves

$$\langle \nabla\psi_k(\gamma(\cdot)),\ \dot{\gamma} \rangle_{L^2} = \left\langle \nabla\psi_k(\gamma(\cdot)),\ \textstyle\sum_{i=1}^M \theta_i Y_i(\gamma(\cdot)) \right\rangle_{L^2} , \quad k = 1, \ldots, K, \tag{3.6}$$

where the left-hand side can be reformulated as

$$\langle \nabla\psi_k(\gamma(\cdot)),\ \dot{\gamma} \rangle_{L^2} = \int_0^T \nabla\psi_k(\gamma(t))\dot{\gamma}\, dt = \int_0^T \tfrac{d}{dt} \left(\psi_k(\gamma(t))\, dt = \psi_k(\gamma(b)) - \psi_k(\gamma(a)) \right). \tag{3.7}$$

Thus, the matrix formulation of Equation 3.6 becomes

$$\boldsymbol{b} = \boldsymbol{G}\theta \text{ with entries } \boldsymbol{b}_k = \psi_k(\gamma(b)) - \psi_k(\gamma(a)) \text{ and } \boldsymbol{G}_{k,i} = \langle Y_i(\gamma(\cdot)),\ \nabla\psi_k(\gamma(\cdot)) \rangle_{L^2} , \tag{3.8}$$

which also avoid direct measurements or evaluations of $\dot{\gamma}$. The generalization results in several practical and theoretical advantages. As evidenced by the results of Experiment 3 (see Figure 10.7 and Table 11.1), the method developed in this paper can yield better results than various implementations of ILS, including the case where sufficiently many trajectories available to yield a well-conditioned ILS regression matrix.

3.2. A Cohesive Vision of System Identification Problems. In this section we adapt the use of both the test functions $\{\psi_k : [0,T] \to \mathbb{R}\}_{k=1}^K$ and the basis functions $\{Y_i\}_{i=1}^M$ as done in weak-SINDy. Foremost, the least-squares/SINDy technique can be cast into this framework under the assumptions that $\gamma(\cdot)$, $\dot{\gamma}(\cdot)$, and $Y_i(\gamma(\cdot))$ lie in the same RKHS (defined in Section 4), denoted $H_{[0,T]}$, and that the test functions are chosen to be, K_{t_k}, the reproducing kernels centered at the time-steps t_k. The reproducing kernels centered at t_k are functions in $H_{[0,T]}$ with the property $h(t_k) = \langle h, K_{t_k} \rangle_{H[0,T]}$. The right hand side of Equation 3.2 can thus be evaluated as

$$\langle Y_i(\gamma(\cdot)),\ \psi_k \rangle_{L^2} = \langle Y_i \circ \gamma(\cdot),\ K_{t_k} \rangle_{H_{[0,T]}} = Y_i(\gamma(t_k)). \tag{3.9}$$

While the test functions could also be selected to be Dirac delta distributions to yield a similar result, this paper takes the view point that the test functions will form a finite dimensional subspace of the underlying Hilbert function space such that Hilbert space

[1] Since the system is time-invariant, the collection of N trajectories $\{\gamma_j : [0,T] \to X\}_{j=1}^N$ can also be obtained by segmenting a single trajectory $\gamma : [0, NT] \to X$ as $\gamma_j(t) = \gamma(jt)$.

theory applies to the analysis. The theoretical sacrifice made to enable the Hilbert space formulation is the assumption that the solution and the basis functions lie in a RKHS. This assumption relates to decay rate constraints on Fourier coefficients for functions inside the RKHS (via Mercers theorem) [41, Theorem 4.49].

3.2.1. Operator form of Occupation Kernel System Identification. Consider the Liouville operator $A_f g = \nabla_x g \cdot f$ and the occupation kernel Γ_γ, with the property $\langle f,\ \Gamma_\gamma \rangle_H = \int_0^T f(\gamma(t))dt$, as defined in Section 5.

We note that $\int_0^T \nabla_x g(\gamma(t)) f(\gamma(t))dt = \langle A_f g,\ \Gamma_\gamma \rangle_H$. Therefore, given the test functions $\{\psi_k : X \to \mathbb{R}\}_{k=1}^K$, the entries in the linear system Equation 3.8 can be written as

$$\boldsymbol{b}_k = \left\langle \psi_k,\ A_f^*(\Gamma_\gamma) \right\rangle_H, \quad \text{and} \quad \boldsymbol{G}_{k,i} = \left\langle \psi_k,\ A_{Y_i}^*(\Gamma_\gamma) \right\rangle_H, \tag{3.10}$$

where $A_{Y_i}^*$ denotes the adjoint of A_{Y_i}, as subsequently defined in Section 4.2. The success of this method is then is tied to how similar the functions $A_f^*(\Gamma_\gamma)$ and $A_{\sum_i \theta_i Y_i}^*(\Gamma_\gamma)$ are on the finite dimensional space generated by the test functions ψ_k. In the implementation given in this paper, the test functions ψ_k are chosen to be reproducing kernels.

3.2.2. Operator form of Least-squares/ SINDy/ Weak-SINDy. By invoking a composition operator $C_\gamma(f) := f \circ \gamma$, Equations (3.1) and (3.2) can be written as

$$\boldsymbol{b}_k = \langle C_\gamma(f),\ \psi_k \rangle, \quad \text{and} \quad \boldsymbol{G}_{i,k} = \langle C_\gamma(Y_i),\ \psi_k \rangle. \tag{3.11}$$

Again, the least squares/SINDy technique makes the assumption that the test functions are reproducing kernels at the time steps yielding Equation 3.9, while Weak-SINDy makes the assumption that the test functions have compact support in the relevant domain allowing the use of Equation 3.4.

Overall, this manuscript takes the viewpoint that under the assumption of increased regularity of the involved functions, parameter estimation problems can be placed in a setting other than L^2. By analyzing in this new setting, we gain several theoretical advantages. Often, the assumptions of increased regularity allow for stronger notions of convergence. For example, in the case of truncation error estimates, we establish convergence in an RKHS, which implies pointwise everywhere convergence. We also gain new avenues for analysis by considering soft analysis approaches via the properties of the operator, and the additional structure of an RKHS can yield new approaches such as the Gram-Schmidt procedure in Section 8.2.

4. Preliminaries.

In this section, fundamental properties of RKHSs are summarized before introducing the idea of occupation kernels.

4.1. Reproducing Kernel Hilbert Spaces.

DEFINITION 4.1. *A RKHS, H, over a set X is a Hilbert space of real valued functions over the set X such that for all $x \in X$ the evaluation functional, $E_x : H \to \mathbb{R}$, given as $E_x g := g(x)$ is bounded.*

The Riesz representation theorem guarantees, for all $x \in X$, the existence of a function $k_x \in H$ such that $\langle g, k_x \rangle_H = g(x)$, where $\langle \cdot, \cdot \rangle_H$ is the inner product for H [27, Chapter 1]. The function $K(x, y) = \langle k_y, k_x \rangle_H$ is called the reproducing kernel of H. Each reproducing kernel has an associated feature map, $\Psi : X \to \ell^2(\mathbb{N})$, such that $K(x, y) = \langle \Psi(x), \Psi(y) \rangle_{\ell^2(\mathbb{N})}$. The feature map can be obtained by using an

orthonormal basis for H, and if $K : X \times X \to \mathbb{R}$ can be represented through a feature map, then there is a unique RKHS for which K is its reproducing kernel [4].

This manuscript utilizes two RKHSs, which are defined through their reproducing kernels. For $\mu > 0$, the reproducing kernel $K_E(x, y) = e^{\mu x^\top y}$ is called the exponential dot product kernel, and for $\mu > 0$, the reproducing kernel given as $K_G(x, y) = \exp\left(-\frac{1}{\mu}\|x - y\|_2^2\right)$ is called a Gaussian radial basis function (RBF) kernel. Both K_E and K_G are reproducing kernels for RKHSs over $\mathbb{R}^n$ [41, Chapter 4].

4.2. Densely Defined Operators. For many RKHSs of continuously differentiable functions, the differential operator is unbounded, which means that there are frequently functions in such a RKHS, H, such that their derivative is not a member of H. The focus of this manuscript is in the study of Liouville operators, which implement the gradient operation on members of a RKHS. As such, care will be required in defining these operators and their domain.

Given a Hilbert space, H, and a subspace $\mathcal{D}(T) \subset H$, a linear operator $T : \mathcal{D}(T) \to H$ is called densely defined if $\mathcal{D}(T)$ is a dense subspace of H [28, Chapter 5]. The operator T is closed if $g \in \mathcal{D}(T)$ and $Tg = h$ for every sequence $\{g_m\}_{m=0}^{\infty} \subset \mathcal{D}(T)$ such that $g_m \to g \in H$ and $Tg_m \to h \in H$.

The adjoint of a possibly unbounded operator is defined through its domain $\mathcal{D}(T^*) = \{g \in H : h \mapsto \langle Th, g\rangle_H \text{ is bounded over } \mathcal{D}(T)\}$. Since $\mathcal{D}(T)$ is dense in H, for each $g \in \mathcal{D}(T^*)$, the functional $h \mapsto \langle Th, g\rangle_H$ extends to a bounded functional defined on H [28, Theorem 2.1.11]. As a result, by the Riesz representation theorem, there exists a unique member $T^*g \in H$ such that $\langle Th, g\rangle_H = \langle h, T^*g\rangle_H$. The adjoint T^* is then defined as the operator taking $g \in \mathcal{D}(T^*)$ to T^*g [28, Section 5.1.2]. The closedness of the operator guarantees the nonemptiness of the domain of its adjoint. In fact, the following stronger statement holds.

LEMMA 4.2. *(c.f. [28, Chapter 5]) The adjoint of a closed operator is densely defined.*

5. Liouville Operators and Occupation Kernels. To establish a connection between RKHSs and nonlinear dynamical systems, the following operator is introduced, which is inspired by the study of occupation measures [20].

DEFINITION 5.1. *Let $\dot{x} = f(x)$ be a dynamical system with the dynamics, $f : \mathbb{R}^n \to \mathbb{R}^n$, locally Lipschitz continuous, and suppose that H is a RKHS over a set X, where $X \subset \mathbb{R}^n$ is compact. The* Liouville operator with symbol *f, $A_f : \mathcal{D}(A_f) \to H$, is given as $A_f g := \nabla_x g \cdot f$, where $\mathcal{D}(A_f) := \{g \in H : \nabla_x g \cdot f \in H\}$.*

A small subset of Liouville operators can be obtained as infinitesimal generators for semigroups of Koopman operators. For a Liouville operator to be such an infinitesimal generator, the associated symbol (or continuous time dynamics) must be forward invariant (cf. [13, Assumption 1]). Specifically, dynamics that lead to finite escape times, such as $\dot{x} = 1 + x^2$, cannot be discretized to yield a Koopman operator, and as a result, the Liouville operator corresponding to the symbol $f(x) := 1 + x^2$ cannot be obtained as an infinitesimal generator. More generally, Liouville operators corresponding to globally Lipschitz symbols can be obtained as infinitesimal generators. However, techniques that rely on global Lipschitz continuity would exclude most polynomial systems and other nonlinear systems with finite escape times. Hence, methods that allow generators to inherit desirable properties, such as being densely defined and closed, from their associated semigroups, cannot be leveraged to study a majority of

Liouville operators. These properties must be established separately to ensure full generality.

Liouville operators embed the nonlinear dynamics inside of an unbounded operator. The first question to address is that of existence. In particular, are there reasonable classes of dynamics for which the Liouville operator is densely defined over a RKHS?

EXAMPLE 1. *The most commonly investigated dynamical systems are those with polynomial dynamics. In the case that f is a polynomial over $\mathbb{R}^n$, a Liouville operator with those dynamics maps polynomials to polynomials, when polynomials are contained in the RKHS in question. One example, where polynomials are not only contained in the RKHS but are also dense, is the native RKHS of the exponential dot product kernel [41, Chapter 4]. Moreover, for this space, the collection of monomials forms an orthogonal basis.*

The above example guarantees the existence of densely defined Liouville operators for a large class of dynamics. In the case when a Liouville operator is not known to be densely defined, some of the methods of this manuscript may still be applied as a heuristic algorithm. As a differential operator, A_f is not expected to be a bounded over any RKHS. However, as differentiation is a closed operator over RKHSs consisting of continuously differentiable functions, which follows from [41, Corollary 4.36], it can be similarly established that A_f is closed under the same circumstances.

THEOREM 5.2. *Let H be a RKHS of continuously differentiable functions over a set X and $f : \mathbb{R}^n \to \mathbb{R}^n$ be a function such that A_f has nontrivial domain, then A_f is a closed operator.*

Proof. By [41, Corollary 4.36], it can be observed that if $\{g_m\}_{m=1}^{\infty} \subset H$ such that $\|g_m - g\|_H \to 0$ in H then $\left\{\frac{\partial}{\partial x_i} g_m\right\}_{m=0}^{\infty}$ converges to $\frac{\partial}{\partial x_i} g$ uniformly in X. Hence, if $\{g_m\}_{m=0}^{\infty} \subset \mathcal{D}(A_f) \subset H$ converges to g and $\{A_f g_m\}_{m=0}^{\infty}$ converges to $h \in H$ then $\nabla_x g_m(x) f(x)$ converges to $\nabla_x g(x) f(x)$ pointwise. As $A_f g_m(x) = \nabla_x g_m(x) f(x)$, it follows that $h(x) = \lim_{m\to\infty} A_f g_m(x) = \nabla_x g(x) f(x)$. By the definition of $\mathcal{D}(A_f)$, $g \in \mathcal{D}(A_f)$ and $A_f g = h$. ∎

Thus, A_f is a closed operator for RKHSs consisting of continuously differentiable functions. Consequently, the adjoints of densely defined Liouville operators are themselves densely defined by Lemma 4.2. The main object of study of this manuscript is a class of functions,dubbed occupation kernels, within the domain of the adjoint of the Liouville operator.

DEFINITION 5.3. *Let $X \subset \mathbb{R}^n$ be compact, H be a RKHS of continuous functions over X, and $\gamma : [0, T] \to X$ be a continuous trajectory. The functional $g \mapsto \int_0^T g(\gamma(\tau))d\tau$ is bounded over H, and may be represented as $\int_0^T g(\gamma(\tau))d\tau = \langle g, \Gamma_\gamma \rangle_H$, for some $\Gamma_\gamma \in H$ by the Riesz representation theorem. The function Γ_γ is called the occupation kernel corresponding to γ in H.*

PROPOSITION 5.4. *Let H be a RKHS of continuously differentiable functions over a compact set X, and suppose that $f : \mathbb{R}^n \to \mathbb{R}^n$ is Lipschitz continuous. If $\gamma : [0, T] \to X$ is a trajectory as in Definition 5.3 that satisfies $\dot{\gamma} = f(\gamma)$, then $\Gamma_\gamma \in \mathcal{D}(A_f^*)$, and $A_f^* \Gamma_\gamma = K(\cdot, \gamma(T)) - K(\cdot, \gamma(0))$.*

Proof. For Γ_γ to be in $\mathcal{D}(A_f^*)$, the functional $g \mapsto \langle A_f g, \Gamma_\gamma \rangle_H$ needs to be bounded

over $\mathcal{D}(A_f)$. Note that

$$\int_0^T \nabla_x g(\gamma(t)) f(\gamma(t)) dt = g(\gamma(T)) - g(\gamma(0)) = \langle g, K(\cdot, \gamma(T)) - K(\cdot, \gamma(0)) \rangle_H \tag{5.1}$$

as the integrand of (5.1) is the total derivative of $g(\gamma(t))$. The left hand side of (5.1) may be expressed as $\langle A_f g, \Gamma_\gamma \rangle_H$, while the right hand side satisfies the bound $|g(\gamma(T)) - g(\gamma(0))| = |\langle g, K(\cdot, \gamma(T)) - K(\cdot, \gamma(0)) \rangle_H| \leq \|g\|_H \|K(\cdot, \gamma(T)) - K(\cdot, \gamma(0))\|_H$, which establishes the boundedness of $g \mapsto \langle A_f g, \Gamma_\gamma \rangle_H$. Moreover, since $\langle A_f g, \Gamma_\gamma \rangle_H = \langle g, K(\cdot, \gamma(T)) - K(\cdot, \gamma(0)) \rangle_H$, it follows that $A_f^* \Gamma_\gamma = K(\cdot, \gamma(T)) - K(\cdot, \gamma(0))$ ∎

Proposition 5.4 completes the integration of nonlinear dynamical systems with RKHSs. In particular, valid trajectories for the dynamical system appear as occupation kernels within the domain of the adjoint of the Liouville operator corresponding to the dynamics. This intertwining allows for the expression of finite dimensional nonlinear dynamics as linear systems in infinite dimensions.

Moreover, the relation $\langle A_f g, \Gamma_\gamma \rangle_H = g(\gamma(T)) - g(\gamma(0))$ for all $g \in \mathcal{D}(A_f)$ uniquely determines Γ_γ. This relation will be used subsequently to establish constraints for parameter identification in a system identification setting.

While occupation kernels representing trajectories of a dynamical systems have a clear image under the adjoint of the corresponding Liouville operator, it can be demonstrated that all occupation kernels corresponding to continuous functions are also in the domains of the adjoints of Liouville operators.

THEOREM 5.5. *Let H be a RKHS of continuously differentiable functions over a set $X \subset \mathbb{R}^n$, let f be a symbol for a densely defined Liouville operator, $A_f : \mathcal{D}(A_f) \to H$, and let $\gamma : [0, 1] \to X$ be a continuous trajectory. Then $\Gamma_\gamma \in \mathcal{D}(A_f^*)$ and for each $x_0 \in X$, $K(\cdot, x_0) \in \mathcal{D}(A_f^*)$. Moreover, the images of each function may be expressed as $A_f^* K(\cdot, x_0) = \nabla_2 K(\cdot, x_0) f(x_0)$, and $A_f^* \Gamma_\gamma = \int_0^T \nabla_2 K(\cdot, \gamma(t)) f(\gamma(t)) dt$, where ∇_2 indicates that the gradient is performed on the second variable.*

Proof. For each $i = 1, \ldots, n$, the boundedness of the linear functional $g \mapsto \frac{\partial}{\partial x_i} g(x_0)$ follows from [41, Corollary 4.36]. Let $\{e_m\}_{m=1}^\infty \subset H$ be an orthonormal basis for H, then $\frac{\partial}{\partial x_i} K(\cdot, \cdot) = \sum_{m=1}^\infty e_m(\cdot) \frac{\partial}{\partial x_i} e_m(\cdot)$. More generally, if $g = \sum_{m=1}^\infty a_m e_m$ then $\frac{\partial}{\partial x_i} g(x_0) = \sum_{m=1}^\infty a_m \frac{\partial}{\partial x_i} e_m(x_0)$. Thus,

$$\frac{\partial}{\partial x_i} g(x_0) f_i(x_0) = \left\langle g, \frac{\partial}{\partial x_i} K(\cdot, x_0) \right\rangle_H f_i(x_0) = \left\langle g, \frac{\partial}{\partial x_i} K(\cdot, x_0) f_i(x_0) \right\rangle_H.$$

It follows that $A_f^* K(\cdot, x_0) = \nabla_2 K(\cdot, x_0) f(x_0)$ via a linear combination. The formula for $A_f^* \Gamma_\gamma$ follows from a limiting argument leveraging Proposition 6.2 below and the closedness of A_f^*. ∎

The last theorem of this section characterizes symbols that give rise to densely defined Liouville operators over RKHSs consisting of real analytic functions. Many frequently used RKHSs are of this form, including the native RKHSs of exponential dot product kernels and Gaussian RBF kernels [41, 21]. The characterization below is essential for the construction of one of the inner products in Section 7.

THEOREM 5.6. *If f is the symbol for a densely defined Liouville operator over a RKHS consisting of real analytic functions of several variables over a set $X \subset \mathbb{R}^n$ such that the collection of gradient of the functions in the space are universal in $\mathbb{R}^n$, then f must be a vector valued real analytic function of several variables.*

REMARK 1. *The universality of the gradients of the kernels are readily established for the exponential dot product kernel, where the gradients of the monomials will span the collection of vectors of monomials in $\mathbb{R}^n$. Gradients of the Gaussian RBF kernel are shown to be universal in [6].*

Proof. Write $f = (f_1, \ldots, f_n)^\top$. Given any $v \in \mathbb{R}^n \backslash \{0\}$ and $x \in X$, the proof of Theorem 5.5 established that the functional $g \mapsto \nabla g(x)v$ is bounded as real analytic functions are continuously differentiable. Hence, there is a function $h_{x,v}$ that represents that functional through the inner product of the RKHS. The universality of the gradients of functions in the Hilbert space guarantees that there is at least one function for which $\nabla g(x)v$ is nonzero. Hence, $h_{x,v}$ is not the zero function for any $v \in \mathbb{R}^n \setminus \{0\}$ and $x \in \mathbb{R}^n$.

Select $x \in X$, then $W_x := \operatorname{span}\{\nabla g(x)\}_{g \in \mathcal{D}(A_f)} = \mathbb{R}^n$. If not, then there is a $v \in \mathbb{R}^n \setminus \{0\}$ such that $\nabla g(x)v = 0$ for all $g \in \mathcal{D}(A_f)$, and hence, $\langle g, h_{x,v} \rangle_H = 0$ for all $g \in \mathcal{D}(A_f)$. As a result, $\mathcal{D}(A_f)$ has codimension at least 1 in H and is not dense (which is a contradiction).

Thus, for a fixed $x_0 \in \mathbb{R}^n$, a complete basis for $\mathbb{R}^n$ may be selected from W_{x_0} as $\nabla g_1(x_0), \ldots, \nabla g_n(x_0)$, and there are linear combinations of these vectors that yield the standard basis in $\mathbb{R}^n$ at the point x_0. In particular, this means that $x \mapsto \det(B(x))$, with $B(x) := \begin{pmatrix} \nabla g_1(x)^\top & \cdots & \nabla g_n(x)^\top \end{pmatrix}^\top$, is an analytic function that is nonvanishing at x_0 (by linear independence). The analyticity follows since products and sums of real analytic functions are real analytic, and each component of $\nabla g_i(x)$ is a real analytic function.

Now consider $G_i(x) := A_f g_i(x) = \nabla g_i(x) f(x)$. Let $B_i(x)$ be matrix obtained by replacing the i-th column of $B(x)$ by the column vector consisting of the functions $G_i(x)$. Then, $\det(B_i(x))$ is also a real analytic function.

Finally, by Cramer's rule, $f_i(x) = \det(B_i(x))/\det(B(x))$, and since $\det(B(x_0)) \neq 0$, f_i is real analytic at x_0. As x_0 was arbitrary, f_i is real analytic everywhere. ∎

6. Estimation of Occupation Kernels. The occupation kernels can be expressed as an integral against the reproducing kernel in a RKHS as demonstrated in Proposition 6.1. As a result. inner products against occupation kernels in a RKHS can be approximated using quadrature techniques for integration. Proofs of all results in this section follow from numerical integration theory and can be found in [35] (an arXiv pre-print of this article).

PROPOSITION 6.1. *Let H be a RKHS over a compact set X consisting of continuous functions and let $\gamma : [0, T] \to X$ be a continuous trajectory as in Definition 5.3. The occupation kernel corresponding to γ in H, Γ_γ, may be expressed as*

$$\Gamma_\gamma(x) = \int_0^T K(x, \gamma(t))dt. \tag{6.1}$$

Leveraging Proposition 6.1, quadrature techniques can be demonstrated to give not only pointwise convergence but also norm convergence in the RKHS, which is a strictly stronger result.

PROPOSITION 6.2. *Under the hypothesis of Proposition 6.1, let $t_0 = 0 < t_1 < t_2 < \ldots < t_F = T$, suppose that γ is a continuously differentiable trajectory and H is composed of continuously differentiable functions. Consider*

$$\hat{\Gamma}_\gamma(x) := \sum_{i=1}^{F} (t_i - t_{i-1}) K(x, \gamma(t_i)). \tag{6.2}$$

The norm distance is bounded as $\|\Gamma_\gamma - \hat{\Gamma}_\gamma\|_H^2 = O(h)$, *where* $h = \max_{i=1,\dots,F} |t_i - t_{i-1}|$.

It should be clear from the proof of Proposition 6.2 that higher order quadrature rules for estimating the integral in (6.1) will also lead to higher order convergence rates of the difference in Hilbert space norms of the occupation kernel and the quadrature estimate with the added caveat of higher order continuous differentiability of the kernels and trajectories. Additional propositions that address homotopic parameterizations of curves and their respective occupation kernels can be found in [35].

7. Inner Products on Symbols of Densely Defined Liouville Operators. This section presents a method for parameter identification that builds on the Hilbert space and operator theoretic framework presented in Section 5. In particular, the development of this section uses the adjoint relation between Liouville operators and occupation kernels to establish an inner product on the collection of symbols for densely defined Liouville operators, $\mathcal{F}$, over a RKHS. This section begins with a pre-inner product arising from a single trajectory for the dynamical system, and then develops two different inner products based on this pre-inner product.

Let $\gamma : [0, T] \to \mathbb{R}^n$ satisfying $\dot{\gamma} = f(\gamma)$. As the identity $A_f^* \Gamma_\gamma = \sum_{m=1}^M \theta_m A_{Y_m}^* \Gamma_\gamma$ holds the following quadratic form arises as

$$0 = \left\| A_f^* \Gamma_\gamma - \sum_{m=1}^M \theta_m A_{Y_m}^* \Gamma_\gamma \right\|_H^2 \tag{7.1}$$
$$= \left\| A_f^* \Gamma_\gamma \right\|_H^2 - 2 \sum_{m=1}^M \theta_m \langle A_f^* \Gamma_\gamma, A_{Y_m}^* \Gamma_\gamma \rangle_H + \sum_{m,m'=1}^M \theta_m \theta_{m'} \langle A_{Y_m}^* \Gamma_\gamma, A_{Y_{m'}}^* \Gamma_\gamma \rangle_H.$$

The challenge in leveraging (7.1) to generate constraints on θ for system identification lies in the ability to compute the various elements of (7.1). For example the first two terms can be computed without appealing to a particular kernel space or determining the adjoint operator $A_{Y_m}^*$ as

$$\begin{aligned} \|A_f^* \Gamma_\gamma\|_H^2 &= \langle K(\cdot, \gamma(T)) - K(\cdot, \gamma(0)), K(\cdot, \gamma(T)) - K(\cdot, \gamma(0)) \rangle_H \\ &= K(\gamma(T), \gamma(T)) - 2K(\gamma(T), \gamma(0)) + K(\gamma(0), \gamma(0)), \text{ and} \\ \langle A_f^* \Gamma_\gamma, A_{Y_m}^* \Gamma_\gamma \rangle_H &= \int_0^T A_{Y_m} (K(\cdot, \gamma(T)) - K(\cdot, \gamma(0))) \mid_{\gamma(t)} dt, \end{aligned}$$

where the notation $(\cdot) \mid_{\gamma(t)}$ stands for evaluation at $\gamma(t)$.

Closed form expressions of $A_{Y_m}^*$ are not expected to be easily determined for most choices of Y_m and RKHSs (e.g. [37]). Hence, computation of the third term in (7.1) relies on the action of $A_{Y_m}^*$ on an occupation kernel (which is easier to determine, and was given in Theorem 5.5) as

$$\langle A_{Y_m}^* \Gamma_\gamma, A_{Y_{m'}}^* \Gamma_\gamma \rangle_H = \int_0^T \int_0^T \nabla_1 \left(\nabla_2 K(\gamma(t), \gamma(\tau)) Y_m(\gamma(\tau)) \right) Y_{m'}(\gamma(t)) d\tau dt. \tag{7.2}$$

Hence, parameter identification can be performed using only the occupation kernel and the Liouville operators by setting the gradient of (7.1) equal to zero. As the norm squared in (7.1) is zero at the true parameters of the system, which is the smallest value the norm can take, this must be the minimum of the quadratic equation in (7.1), and hence the true parameters must also set the gradient equal to zero. The

parameters $\theta_1, \dots, \theta_M$ must then satisfy

$$\begin{pmatrix} \langle Y_1, Y_1 \rangle_{\mathcal{F},\gamma} & \cdots & \langle Y_1, Y_M \rangle_{\mathcal{F},\gamma} \\ \vdots & \ddots & \vdots \\ \langle Y_M, Y_1 \rangle_{\mathcal{F},\gamma} & \cdots & \langle Y_M, Y_M \rangle_{\mathcal{F},\gamma} \end{pmatrix} \begin{pmatrix} \theta_1 \\ \vdots \\ \theta_M \end{pmatrix} = \begin{pmatrix} \langle f, Y_1 \rangle_{\mathcal{F},\gamma} \\ \vdots \\ \langle f, Y_M \rangle_{\mathcal{F},\gamma} \end{pmatrix}, \tag{7.3}$$

where for a collection, $\mathcal{F}$ of symbols of densely defined Liouville operators, the bilinear form $\langle f, g \rangle_{\mathcal{F},\gamma}$ is given as $\langle f, g \rangle_{\mathcal{F},\gamma} := \langle A_g^* \Gamma_\gamma, A_f^* \Gamma_\gamma \rangle_H$. The bilinear form induces a pre-inner product on the space of dynamical systems giving rise to densely defined Liouville operators over H. The resolution of (7.3) gives the projection of an arbitrary $f \in \mathcal{F}$ onto the span of $\{Y_1, \cdots, Y_M\}$ with respect to the pre-inner product $\langle \cdot, \cdot, \rangle_{\mathcal{F},\gamma}$.

Note that, in contrast with the SINDy method found in [5], (7.3) yields a derivative free approach for the system identification problem. Here, the only derivatives to be performed are those that can be computed symbolically, which are easily computed for many reproducing kernels. Moreover, the above formulation only leverages a single trajectory of the system, and the size of the positive semidefinite Gram matrix $G := (\langle Y_m, Y_{m'} \rangle_{\mathcal{F},\gamma})_{m,m'=1}^{M}$ corresponding to (7.1), is governed only by the number of basis functions.

7.1. Computational Challenges and Feature Space Representations. The conditioning of the Gram matrix, G is frequently poor. One approach to improve the condition number is to recognize that the weights are unchanged when the inner product is adjusted between two different trajectories satisfying the same dynamics. The linear systems in (7.3) corresponding to each trajectory may then be concatenated, and a left psuedo-inverse may be employed to determine the parameters for the system.

However, as is typical of numerical methods and matrix computations (e.g. [10, 46]), a more reliable result may be extracted via a simplification obtained through a factorization of the Gram matrices. If the reproducing kernel K is associated with the feature map $\Psi(x) = (\Psi_1(x), \Psi_2(x), \dots)^\top \in \ell^2(\mathbb{N})$ as $K(x, y) = \sum_{s=1}^{\infty} \Psi_s(x) \Psi_s(y)$, then $A_{Y_i}^* \Gamma_\gamma$ and $A_f^* \Gamma_\gamma$ may be written as

$$A_{Y_i}^* \Gamma_\gamma = \sum_{s=1}^{\infty} \Psi_s(x) \int_0^T \nabla \Psi_s(\gamma(t)) Y_i(\gamma(t)) dt, \text{ and} \tag{7.4}$$

$$A_f^* \Gamma_\gamma = \sum_{s=1}^{\infty} \Psi_s(x) \left(\Psi_s(\gamma(T)) - \Psi_s(\gamma(0)) \right). \tag{7.5}$$

Hence, the Gram matrix on the left hand side of (7.3) may be expressed as

$$V_\gamma^\top V_\gamma, \quad \text{where} \quad V_\gamma := \left(\int_0^T \nabla \Psi_i(\gamma(t)) Y_j(\gamma(t)) dt \right)_{i,j=1,1}^{\infty,M}. \tag{7.6}$$

The right hand side of (7.3) is expressible as

$$\begin{pmatrix} \langle f, Y_1 \rangle_{\mathcal{F},\gamma} \\ \langle f, Y_2 \rangle_{\mathcal{F},\gamma} \\ \vdots \\ \langle f, Y_M \rangle_{\mathcal{F},\gamma} \end{pmatrix} = V_\gamma^\top (\Psi(\gamma(T)) - \Psi(\gamma(0))) = V_\gamma^\top \begin{pmatrix} \Psi_1(\gamma(T)) - \Psi_1(\gamma(0)) \\ \Psi_2(\gamma(T)) - \Psi_2(\gamma(0)) \\ \Psi_3(\gamma(T)) - \Psi_3(\gamma(0)) \\ \vdots \end{pmatrix}. \tag{7.7}$$

Since both (7.7) and (7.6) have the infinite matrix $V_\gamma^\top$ on the left hand side, the resolution of the system

$$V_\gamma \theta = \Psi(\gamma(T)) - \Psi(\gamma(0)) \tag{7.8}$$

also satisfies (7.3).

As Ψ is infinite dimensional for most reproducing kernels, V_γ is an infinite dimensional matrix. For a given reproducing kernel, such as the Gaussian RBF kernel, one option to obtain a finite-dimensional representation is to leverage decaying factors for the feature space representation and set a cutoff after the size of the features fall under a pre-specified precision, as was done in [10] for scattered data interpolation. Conversely, given a finite collection of real-valued functions $\{g_1, \ldots, g_S\}$ over a set X, a reproducing kernel, $K(x, y) = \sum_{s=1}^{S} g_s(x) g_s(y)$, may be constructed yielding a matrix V of finite dimensions. In the sequel, a collection of test functions will be employed for the resolution of the system identification problem given in (2.1). These test functions can be any collection of continuously differentiable functions, provided that the collection is either finite or constitutes the members of a feature map to $\ell^2(\mathbb{N})$. Each collection of test functions gives rise to a reproducing kernel and in turn, give a different inner product on the collection of densely defined Liouville operators over the native space of that reproducing kernel.

It should also be noted that if f is known to be explicitly representable as a linear combination of a finite number of Y_i's, then the matrix V_γ only needs to be evaluated up until its rank matches the number of basis functions, M. At that point, θ is completely determined.

7.2. Inner Products from Pre-Inner Products. Pre-inner products give rise to pseudonorms on vector spaces. When $\nabla_1 \nabla_2 (K(x, y) + K(y, x))$ is positive definite and bounded below for all $x, y \in \gamma([0, T])$, then the pseudonorm induced by the pre-inner product $\langle f, g \rangle_{\mathcal{F},\gamma}$ can be seen to dominate the L^2 norm on $\gamma([0, T])$[2]. As a result, agreement in the pseudonorm implies agreement in the L^2 norm over a trajectory. However, since functions, say $\eta(\cdot)$, that vanish identically on the trajectory cannot be observed through this pseudonorm, there remains some ambiguity, where $f(\cdot)$ and $f(\cdot) + \eta(\cdot)$ are different functions over $\mathbb{R}^n$ but are indistinguishable according to the induced pseudonorm.

It should be noted that in this setting, if one is only interested in modeling the system along the trajectory, then the pre-inner product would be a proper inner product. However, since the purpose of system identification to extend the system model outside of the trajectory to all of $\mathbb{R}^n$, adjustments are necessary to achieve an inner product of this form.

7.2.1. Quotient Approach to Inner Products. In the case where only one trajectory is available for the dynamical system, there is a limited number of options available. First, if f is known to be explicitly a linear combination of the basis functions, then the weights may be determined through the factorization of the Gram matrix above. The pre-inner product can also be made into a proper inner product by restricting it to the quotient space, $\mathcal{F}/\mathcal{N}$, where $\mathcal{N} := \{\eta : \mathbb{R}^n \to \mathbb{R}^n | \eta \in \mathcal{F} \text{ and } \eta(\gamma([0, T])) = \{0\}\}$. The space $\mathcal{F}/\mathcal{N}$ is then an inner product space consisting of equivalence classes of functions from $\mathcal{F}$, where two members of $\mathcal{F}$ are equivalent if

[2]Commonly used kernels such as polynomial, exponential dot product, and Gaussian RBF kernels satisfy these conditions, see Appendix A of [35].

their difference is in $\mathcal{N}$. This is a typical construction of an inner product space from a pre-inner product space. Details can be found in standard references, such as [28].

Computations of the inner product in this form are precisely the same as those for the pre-inner product, where a member of each equivalence class is selected to represent each of f and Y_i for $i = 1, \ldots, M$. This will motivate the algorithm for parameter identification in Section 8.

7.2.2. Integration Approach to Inner Products. When a large collection of trajectories is available from a dynamical system, this collection of trajectories can be used to generate an inner product from the collection of corresponding pre-inner products. In particular, if an $(n-1)$ dimensional sub-manifold of initial points for a collection of observed trajectories is given, where there is at least one point on the submanifold where the dynamics is nonzero, then this collection of trajectories can be leveraged to give an inner product on the collection of dynamical systems giving rise to densely defined Liouville operators.

THEOREM 7.1. *Let $f : \mathbb{R}^n \to \mathbb{R}^n$ be a dynamical system that gives rise to a densely defined Liouville operator over a universal RKHS of real analytic functions on $\mathbb{R}^n$, such that the operator valued kernel, $\nabla_1 \nabla_2 K(x, y)$ is universal. Let $X \subset \mathbb{R}^n$ be a compact smooth Riemann submanifold of $\mathbb{R}^n$. Suppose that $\Omega := \cup_{\xi \in X} \{\gamma_\xi(t) : t \in [0, T]\} \subset \mathbb{R}^n$ is a collection of trajectories corresponding to f, such that $\gamma_\xi(0) = \xi \in X$, and Ω has a nonempty interior. Define the bilinear form on $\mathcal{F}$, $\langle \cdot, \cdot \rangle_{\mathcal{F},\Omega}$, as $\langle p, q \rangle_{\mathcal{F},\Omega} = \int_X \langle p, q \rangle_{\mathcal{F},\gamma_\xi} d\xi$. This bilinear form is an inner product.*

Proof. Linearity and the semidefinite property of the bilinear form follows directly from the same properties of $\langle p, q \rangle_{\mathcal{F},\gamma_\xi}$ for each $\xi \in X$.

Let $\Psi = (\Psi_1, \Psi_2, \ldots)$ be a feature map for the reproducing kernel K, where each $\{\Psi_m\}_{m=1}^\infty$ forms an orthonormal basis for the RKHS. Note that the matrix $\mathbb{K}(x, y) := \nabla_1 \nabla_2 K(x, y) = \sum_{m=1}^\infty \nabla \Psi_m^\top(x) \nabla \Psi_m(x)$ is an operator valued kernel for a RKHS [6]. The universality of $\mathbb{K}$ yields the universality of its features [6].

Let $Y, \tilde{Y} \in \mathcal{F}$, and suppose that $\int_X \langle A^*_{Y-\tilde{Y}} \Gamma_{\gamma_\xi}, A^*_{Y-\tilde{Y}} \Gamma_{\gamma_\xi} \rangle d\xi = 0$. Hence,

$$A^*_{Y-\tilde{Y}} \Gamma_{\gamma_x i} = \sum_{m=1}^\infty \Psi_m(x) \int_0^T \nabla \Psi_m(\gamma_\xi(t))(Y(\gamma_\xi(t)) - \tilde{Y}(\gamma_\xi(t)))dt = 0$$

for almost all $\xi \in X$. By the linear independence of the features,
$\int_0^T \nabla \Psi_m(\gamma_\xi(t))(Y(\gamma_\xi(t)) - \tilde{Y}(\gamma_\xi(t))dt = 0$ for all m. By the universality of Ψ_m, for any $\epsilon > 0$ there is a linear combination of the Ψ_m that uniformly approximates $Y - \tilde{Y}$ over $\gamma_\xi([0, T])$ within ϵ. Hence, $\int_0^T \|Y(\gamma_\xi(t)) - \tilde{Y}(\gamma_\xi(t))\|_2^2 dt = 0$. As Y and $\tilde{Y}$ are both continuous functions, $Y(x) = \tilde{Y}(x)$ for all $x \in \gamma_\xi([0, T])$. As this holds for almost all $\xi \in X$, $Y(x) = \tilde{Y}(x)$ for all $x \in \Omega$ by continuity.

Since, Y and $\tilde{Y}$ are vector valued real analytic functions over $\mathbb{R}^n$ by Theorem 5.6 and Ω has a nonempty interior, it follows that $Y = \tilde{Y}$ as functions over $\mathbb{R}^n$. ∎

8. Parameter Identification via Occupation Kernels. To utilize the integral formulation of the occupation kernel determined in Section 6, the collection of test functions leveraged in the sequel will be a finite collection of reproducing kernels. The method below may be seen as a weak formulation in the sense of Hilbert space inner products, where instead of determining the parameters directly from $A_f = \sum_{m=1}^M \theta_m A_{Y_m}$, the problem is resolved on test functions through an integral. Selection of reproducing kernels as test function allows for the analysis of various error sources using the developed framework. It should be noted that the errors may

also be analyzed using different methods specific to other possible selections of test functions.

For a compact set $X \subset \mathbb{R}^n$, let $\{\gamma_j : [0,T] \to X\}_{j=1}^N$ be a collection of trajectories satisfying the dynamics $\dot{x} = f(x) = \sum_{i=1}^M \theta_i Y_i(x)$, and let Γ_{γ_j} be the corresponding occupation kernels inside a RKHS, $\tilde{H}$ of continuously differentiable functions over X. Suppose that $\text{span}\{c_s\}_{s=1}^\infty \subset X$ is dense. The feature map generated by the set of test functions $\{K(\cdot, c_s)\}_{s=1}^S\}$ is $\Psi(x) = (\tilde{K}(x, c_1), \ldots, \tilde{K}(x, c_s))^\top$. The feature map yields a finite-dimensional representation of the constraints in (7.8) as

$$\langle A_f \tilde{K}(\cdot, c_s), \Gamma_{\gamma_j} \rangle_H = \sum_{i=1}^M \theta_i \langle A_{Y_i} K(\cdot, c_s), \Gamma_{\gamma_j} \rangle_H = \tilde{K}(\gamma_j(T), c_s) - \tilde{K}(\gamma_j(0), c_s), \tag{8.1}$$

for each $s = 1, \ldots, \infty$ and $j = 1, \ldots, N$, which can be expressed in a matrix notation as

$$\mathbf{A}\theta = \mathbf{K}(T) - \mathbf{K}(0), \quad \text{where} \tag{8.2}$$

$$\mathbf{A} = \begin{pmatrix} V_{\gamma_1} \\ \vdots \\ V_{\gamma_N} \end{pmatrix} \in \mathbb{R}^{SN \times M}, \theta = \begin{pmatrix} \theta_1 & \cdots & \theta_M \end{pmatrix}^\top \in \mathbb{R}^M, \text{ and } \mathbf{K}(t) = \begin{pmatrix} \Psi(\gamma_1(t)) \\ \vdots \\ \Psi(\gamma_M(t)) \end{pmatrix} \in \mathbb{R}^{SN}.$$

Under the additional assumption of continuous differentiability of both the reproducing kernels and the trajectories $\{\gamma_j\}_{j=1}^M$, it can be observed through the Cauchy-Schwarz inequality that

$$|\langle A_{Y_i} \tilde{K}(\cdot, c_s), \hat{\Gamma}_{\gamma_j} \rangle_H - \langle A_{Y_i} \tilde{K}(\cdot, c_s), \Gamma_{\gamma_j} \rangle_H| \leq \|A_{Y_i} \tilde{K}(\cdot, c_s)\|_H \|\hat{\Gamma}_{\gamma_j} - \Gamma_{\gamma_j}\|_H.$$

Hence, by Proposition 6.2

$$\langle A_{Y_i} \tilde{K}(\cdot, c_s), \hat{\Gamma}_{\gamma_j} \rangle_H = \langle A_{Y_i} \tilde{K}(\cdot, c_s), \Gamma_{\gamma_j} \rangle_H + O\left(\sqrt{h}\right), \tag{8.3}$$

so that quadrature techniques can be successfully employed for estimation of the inner products contained in (8.2).

Since the matrix $\mathbf{A}$ must be numerically estimated, written as $\hat{\mathbf{A}}$, the parameter values obtained using this method are approximate, and will be represented as $\hat{\theta}$, obtained via

$$\hat{\theta} := (\hat{\mathbf{A}}^\top \hat{\mathbf{A}})^{-1} \hat{\mathbf{A}}^\top (\mathbf{K}(T) - \mathbf{K}(0)).$$

The complete algorithm for the system identification method is given in Algorithm 8.1. Note that in the case of $N = 1$, $A^\top A = V_{\gamma_1}^\top V_{\gamma_1}$ is the Gram matrix given in Section 7, with respect to the reproducing kernel $K(x, y) = \sum_{s=1}^S \tilde{K}(x, c_s)\tilde{K}(y, c_s)$.

8.1. A note on the independence of the algorithm. As was noted above, the constraints for the parameter identification routine can be established independent of the framework of Section 7. However, the utilization of the Hilbert space framework for error estimates requires one additional assumption on the basis functions Y_i.

ASSUMPTION 1. *Given a RKHS, H, over a set X, each of the operators, $A_{Y_i} : \mathcal{D}(A_{Y_i}) \to H$ are densely defined. Moreover, $\cap_{i=1}^M \mathcal{D}(A_{Y_i})$ is dense in H. That is, the operators $A_{Y_1}, \ldots, A_{Y_M}$ have a common dense domain.*

Algorithm 8.1 Pseudocode for the system identification routine of Section 8. In the description some quantities are left in their analytic form, such as the integral of line 7. The choice of quadrature routine can have a significant impact on the overall results, and it is advised that a high accuracy method is leveraged in practice.

1: Input: Trajectories $\{\gamma_j : [0,T] \to \mathbb{R}^n\}_{j=1}^M$, Centers $\{c_s\}_{s=1}^S$, and basis $\{Y_i\}_{i=1}^N$
2: Initialize the empty matrix $\mathbf{A}$ and empty vector $\mathbf{b}$
3: **for** j'=1:M **do**
4: Initialize $S \times N$ matrix $\mathbf{A}_{j'}$ and the length S vector $\mathbf{b}_{j'}$
5: **for** s'=1 to S **do**
6: **for** i'=1 to N **do**
7: Assign the value of the integral $\int_0^T \nabla \tilde{K}(\gamma_{j'}(t), c_{s'}) Y_i(\gamma_{j'}(t))dt$ to the (s', i') entry of $\mathbf{A}_{j'}$.
8: Assign the value $\tilde{K}(\gamma_{j'}(T), c_{s'}) - \tilde{K}(\gamma_{j'}(0), c_{s'})$ to the s' entry of $\mathbf{b}_{j'}$.
9: **end for**
10: **end for**
11: Append $\mathbf{A}_{j'}$ and $\mathbf{b}_{j'}$ to $\mathbf{A}$ and $\mathbf{b}$ respectively.
12: **end for**
13: **return** θ as $(\mathbf{A}^\top \mathbf{A})^{-1} \mathbf{A}^\top \mathbf{b}$.

Assumption 1 ensures the validity of decomposing A_f into a linear combination of densely defined Liouville operators, $\{A_{Y_i}\}_{i=1}^M$. Liouville operators are closely connected to densely defined multiplication operators (c.f. [31, 32, 33, 38]), and the unavailability of complete classifications of densely defined multiplication operators over many RKHSs indicates that characterizing the necessary and sufficient conditions that a dynamical system must meet to allow a Liouville operator to be densely defined may be an intractable problem in many cases. However, sufficient conditions can certainly be established. In particular, Assumption 1 is borne out through examination of the exponential dot product kernel, where a polynomial function f may be decomposed into linear combinations of polynomials, each of which has a corresponding Liouville operator containing polynomials inside of its domain. More sophisticated examples of decompositions can be expressed and treated individually.

For other collections of test functions, where methods outside of the Hilbert space framework can be employed for error analysis, Assumption 1 is not necessary.

8.2. A Gram-Schmidt procedure to improve conditioning. Recall that the main matrix equation involved in the system identification method is $\mathbf{A}\theta = \mathbf{K}(T) - \mathbf{K}(0)$, where $\mathbf{A}$, θ, and $\mathbf{K}$ are introduced in (8.2). Restricting to a single trajectory for simplicity, and picking kernels centered at points $\{c_s\}$ as test functions $\{k_{c_s}\}$, the matrix $\mathbf{A}$ reduces to a single V_γ given by:

$$V_\gamma := \left(\int_0^T \nabla k_{c_i}(\gamma(t)) Y_j(\gamma(t)) dt \right)_{i,j=1,1}^{S,M} \tag{8.4}$$

Selecting $K(x,y) = \sum_{s=1}^S k_{c_s}(x) k_{c_s}(y)$ under reformulation of the system identification problem into a new inner product space via Section 7, we conclude that:

$$(\langle Y_i, Y_j \rangle_{\mathcal{F},\gamma})_{i,j=1,1}^{M,M} = V_\gamma^\top V_\gamma \tag{8.5}$$

and

$$\langle Y_m, Y_{m'}\rangle_{\mathcal{F},\gamma} = \int_0^T \int_0^T Y_{m'}(\gamma(t))^\top \cdot \nabla_1 \left(\nabla_2 K(\gamma(t),\gamma(\tau)) \cdot Y_m(\gamma(\tau))\right) d\tau dt. \tag{8.6}$$

Thus, numerical stability of the developed method can be studied by inspecting either V_γ or the matrix $V_\gamma^\top V_\gamma$, as the singular numbers of one are the roots of the eigenvalues of the other. Moreover, if presented with poor conditioning, the form of Equation 8.5 suggests and enables a Gram-Schmidt procedure to select basis elements Y_i tailored to the data $\gamma(t)$. Specifically, given a selection of basis elements $\{Y_m\}$ we construct a new basis $\{U_m\}$ by $U_1 = Y_1$ and $U_\ell = Y_m - \sum_{\ell=1}^{m-1} \frac{\langle Y_m,\, U_\ell\rangle_{\mathcal{F},\gamma}}{\langle U_\ell,\, U_\ell\rangle_{\mathcal{F},\gamma}} U_\ell$. Here, each new basis element is given as a linear combination of the previous basis elements. Thus to define our new basis $\{U_m\}$ it is only necessary to keep track of the scalar values $\left\{\frac{\langle Y_m,\, U_\ell\rangle_{\mathcal{F},\gamma}}{\langle U_\ell,\, U_\ell\rangle_{\mathcal{F},\gamma}}\right\}_{\ell,m}$.

9. Impact of Signal Noise on Samples. An immediate advantage evident in the usage of occupation kernel methods for system identification over that of methods employing numerical derivative estimates is a robustness to noise. Proofs of all results in this section follow from numerical integration theory and can be found in [35].

9.1. Measurement Noise. Signal noise added to a signal requires sophisticated filtering techniques to allow for reasonable numerical derivative estimates [5]. On the other hand, normally distributed white noise has a smaller effect on integration based methods, since peaks in the noise are infinitesimally small and carry less weight through the integration process.

In the context of occupation kernel based methods, a sample for the system identification method takes the form $\langle A_{Y_l}\tilde{K}(\cdot,c_i), \tilde{\Gamma}_{\gamma_j}\rangle_H = \int_0^T \nabla\tilde{K}(\gamma_j(t),c_i)Y_l(\gamma_j(t))dt$ as in (8.2). Let $\epsilon : [0,T] \to \mathbb{R}^n$ be a disturbance term acting as signal noise, then the noise corrupted sample may be expressed as $\langle A_{Y_l}\tilde{K}(\cdot,c_i), \tilde{\Gamma}_{\gamma_j+\epsilon}\rangle_H = \int_0^T \nabla\tilde{K}(\gamma_j(t)+\epsilon(t),c_i)Y_l(\gamma_j(t)+\epsilon(t))dt$. The following theorem provides a bound on the difference between the corrupted and uncorrupted signals, see Figure 9.1 for a numerical example.

THEOREM 9.1. *Suppose that H is a RKHS consisting of twice continuously differentiable functions and Y_l is continuously differentiable for each l, then the error introduced by a bounded zero mean disturbance*[3]*, $\epsilon \in L^2([0,T],\mathbb{R}^n)$, is $O(T\cdot\sigma(\epsilon))$ where $\sigma(\epsilon)$ is the standard deviation of ϵ with respect to the uniform probability distribution over $[0,T]$.*

Note that the above theorem may be modified to accommodate a possibly unbounded disturbance in $L^2([0,T],\mathbb{R}^n)$ when Y_l and $\nabla\tilde{K}(\cdot,c_i)$ have bounded derivatives and Jacobians respectively.

9.2. Process Noise. The above discussion concerns measurement or signal noise for the system. Another common source of noise is process noise, where a state dependent disturbance, $\eta : \mathbb{R}^n \to \mathbb{R}^n$, impacts the dynamics directly as $\dot{x} = f(x) + \eta(x)$. In this case, there is an impact on the occupation kernel as well as the state variable and Liouville operator.

[3] $L^2([0,T],\mathbb{R}^n)$ denotes the Lebesgue space of functions $g : [0,T] \to \mathbb{R}^n$ such that $\int_0^T \|g(t)\|_2^2 dt < \infty$.

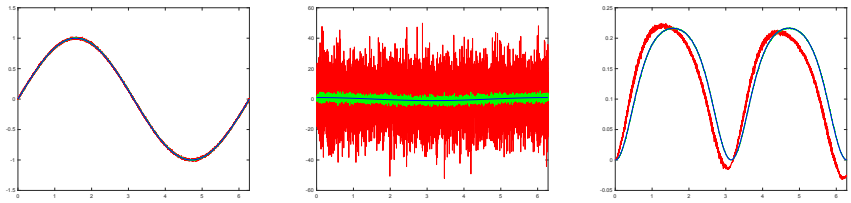

Fig. 9.1: A trajectory, $\sin(t)$ over $[0, 2\pi]$, is shown in the left figure (blue) along with a noise corrupted trajectory (red) and a 10 point moving average filter of the noise corrupted trajectory trajectory (green). The disturbance, ϵ, is normally distributed white noise with mean zero and standard deviation 0.01. The center figure shows numerical derivatives obtained from each trajectory, and the right figure shows samples obtained using occupation kernels. This figure demonstrates the occupation kernel samples' robustness to noise, where even when an unfiltered noisy signal is used, there is a very small error in the sample.

THEOREM 9.2. *If the system is subjected to a process noise, η, that is bounded over a compact set containing the trajectories, Ω, which yields a densely defined Liouville operator, then error induced in the samples may be bounded as*

$$\sqrt{\|\nabla_1\nabla_2\tilde{K}(c,c)\|_F\|\eta(c)\|_2^2}\|\tilde{\Gamma}_{x(\cdot)}\|_H$$
$$+\sqrt{\|\nabla_1\nabla_2\tilde{K}(c,c)\|_F\|f(c)\|_2^2}\sqrt{\tilde{S}^2T\int_0^T e^{2Lt}\left(\int_0^t e^{-L\tau}\|\eta(x(\tau))\|_2 d\tau\right)^2 dt},$$

which is $O(\sup_{x\in\Omega}\|\eta(x)\|_2)$.

The error induced by the process noise is ultimately larger than that induced by measurement noise. However, process noise with small bounds leads to small errors in the samples.

10. Numerical Experiments. Two simulated systems were examined to evaluate the system identification method of Section 8, and one system arising from real world data was also examined. For each simulated system, the trajectories were generated using the Runge-Kutta 4 algorithm with step size $h = 0.001$. On each system several different experiments were performed to evaluate the effects of various parameters, such as the kernel width, the selection of kernel, the numerical integration method, and the number of trajectories utilized. For each system, the centers of the kernel were kept constant throughout the experiments. The dynamics in each example are treated as unknown and are parameterized with respect to the collection monomials of degree up to two. Unless otherwise noted, the matrix $\mathbf{A}$ in (8.2) for each experiment was computed using Simpson's Rule for numerical integration.

SYSTEM 1. *The first dynamical system is sourced from a collection of benchmark examples for the formal verification community presented in [40]. The two dimensional dynamics are given as*

$$\dot{x} = f(x) = \begin{pmatrix} 2x_1 - x_1x_2 \\ 2x_1^2 - x_2 \end{pmatrix}. \tag{10.1}$$

Twenty five trajectories were generated for this system over the time interval $[0,1]$ *and the initial points were selected from the rectangle* $[-0.5,0.5]\times[-2.5,-1.5]$ *through a lattice with width* 0.25*. The collection of trajectories are presented in Figure 10.2.*

The centers for the reproducing kernels for System 1 were selected from a lattice of width 1 *over* $[-3,3]\times[-3,5]$*.*

EXPERIMENT 1. *The first experiment examines the error committed in the parameter estimation by varying the number of trajectories used in the system identification method of Section 8. In this experiment two reproducing kernels were used; the Gaussian RBFs and the exponential dot product kernels. The Gaussian RBFs were used with kernel width* $\mu = 10$*, and the exponential dot product kernels used parameter* $\mu = 1/25$*. The results of Experiment 1 may be observed in Figure 10.4.*

EXPERIMENT 2. *The second experiment explores the effect of the kernel width,* μ*, on the parameter estimation when using the Gaussian RBF in the system identification routine on System 1. The results of Experiment 2 can be observed in Figure 10.5.*

SYSTEM 2. *The second system is the three dimensional Lorenz system [29, 5],*

$$\dot{x} = f(x) = \begin{pmatrix} \sigma(x_2 - x_1) \\ x_1(\rho - x_3) - x_2 \\ x_1 x_2 - \beta x_3 \end{pmatrix}. \tag{10.2}$$

Following [5] a single trajectory was generated over the time interval $[0,100]$ *where* $\sigma = 10$*,* $\beta = 8/3$*,* $\rho = 28$*, and the initial condition was given as* $x_0 = (-8,7,27)^\top$*. The plot of this trajectory is given in Figure 10.3.*

The centers for System 2 were obstained from a lattice with width 10 *within* $[-20,20]\times[-50,50]\times[-20,50]$*.*

EXPERIMENT 3. *This experiment introduces zero mean normally distributed white noise with standard deviation of* 0.01 *to System 1 using the same parameters as in Experiment 1. The system identification method is used on the noise corrupted trajectories as well as the corrupted trajectories treated with a* 20 *point moving average filter. The results of the parameter estimates obtained for this experiment are shown in Table 10.2.*

The experiment also includes a set of a one hundred Monte-Carlo trials that compare the results of the developed occupation kernel system identification method with the ILS approach described in (3.5). Zero mean Gaussian white noise with a standard deviation of 0.01 *is added to trajectories of the Lorenz system. To see the effect of the number,* N*, and the length,* T*, of the trajectories, the experiment compares the performance of the two methods across nine different combinations of* N *and* T*. To ensure all combinations use the same data,* N *and* T *are selected so as to keep the product* NT *constant within rounding errors. The trajectory lengths range from* $T = 0.002s$ *(two samples per trajectory) to* $T = 1.2s$ *(1201 samples per trajectory). Monomials in* 3 *variables, up to order* 3*, are utilized as basis functions to yield a set of* 60 *parameters to be estimated. A set of* 180 *Gaussian RBF kernels with* $\mu = 20^2/3$ *are used to implement the occupation kernel method. The norm of the error,* $\|\theta - \hat{\theta}\|$*, is used in Figure 10.7 and Table 11.1 as a metric for comparison.*

EXPERIMENT 4. *This experiment leveraged the constraints of (8.1) in combination with the LASSO algorithm (*ℓ^1 *regularization) for System 1, where a monomial basis was selected to consist of all monomials of degree* 5 *or less. Of the* 42 *basis functions leveraged in this experiment, the LASSO algorithm selected only* 6 *nonzero weights. These nonzero components corresponded to a collection of basis functions,*

Signal	Benchmark	Occupation Kernels
Position	8.025e-3	8.038e-3
Velocity	4.789e-1	4.729e-1
Acceleration	1.982e+1	1.954e+1
Force	8.941e+0	9.046e+0

Table 10.1: Comparison of percentage errors committed by the developed method and the IDIM-LS method from [17]. A model that uses the identified parameters is simulated forward in time to obtain the simulated position, velocity, force, and acceleration trajectories using the same controller that was used to collect the validation data. The trajectories are then compared against the validation data, where velocity and acceleration are obtained through filtered central difference differentiation.

where the true basis was a proper subset. The parameter values deviated significantly from those of the true parameters, however with the reduced system, the weights can be determined using the pseudo-inverse as in the rest of the manuscript. This method aligns with the SINDy algorithm of [5].

EXPERIMENT 5. *This experiment applies the developed technique to identify the Electro-Mechanical Positioning System (EMPS) from the nonlinear system identification benchmarks archive [17]. The system is a controlled system of the form* $\dot{x}_1 = x_2, \quad \dot{x}_2 = \begin{pmatrix} \tau & -x_2 & -\operatorname{sign}(x_2) & -1 \end{pmatrix} \theta$ *with* $\theta \in \mathbb{R}^4$. *Due to presence of the controller* τ, *this is a time-varying system. To reformulate the problem in terms of an autonomous system, time is augmented to the state vector to get a system of the form*

$$\begin{pmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \dot{x}_3 \end{pmatrix} = \begin{pmatrix} x_2 \\ 0 \\ 1 \end{pmatrix} + \begin{pmatrix} 0 & 0 & 0 & 0 \\ \tau(x_3) & -x_2 & -\operatorname{sign}(x_2) & -1 \\ 0 & 0 & 0 & 0 \end{pmatrix} \theta = h(x) + \sum_{i=1}^{i=4} \theta_i Y_i(x)$$

The system identification is then carried out using a slight modification of the developed technique to accommodate for the known part $h(x) = \begin{pmatrix} x_2 & 0 & 1 \end{pmatrix}^\top$:

$$\hat{\theta} := (\hat{\mathbf{A}}^\top \hat{\mathbf{A}})^{-1} \hat{\mathbf{A}}^\top (\mathbf{K}(T) - \mathbf{K}(0) - \mathbf{B}), \quad \mathbf{B} = \left(\langle A_h K(\cdot, c_{n_{j,1}}), \Gamma_{\gamma_{n_{j,2}}} \rangle_H \right)_{j=1}^{j=SN} \in \mathbb{R}^{SN}.$$

The results of the parameter estimates obtained for this experiment using the validation data provided in [17] are shown in Table 10.1 and Figure 10.1.

11. Discussion. It may be observed through the numerical experiments performed in Section 10 that the system identification method of Section 8 is effective at identifying the parameters for nonlinear systems. In particular, for System 1 parameter estimation errors were as low as 10^{-11} and for System 2 the parameter estimation errors were as low as 10^{-5}. The systems given in Section 10 are of two and three dimensions, and the dynamics are nonlinear. The basis functions utilized to represent the unknown dynamics are monomials of degree up to two with appropriate dimensionality. For example, for a three dimensional system the cardinality of the basis of monomials of degree up to two is 30 when accounting for each dimension (i.e. there is a copy of the 10 monomial basis vectors for each dimension). The actual dynamics in each case use only a handful of the basis functions, which results in a sparse representation of the dynamics in the given basis.

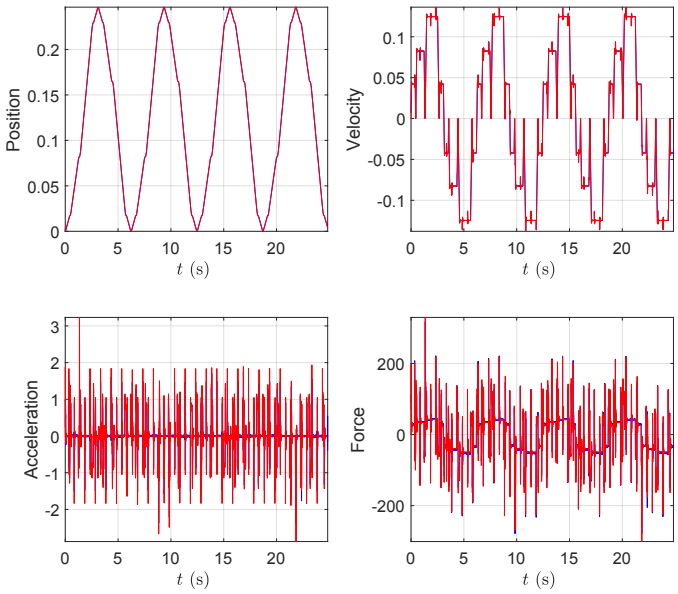

Fig. 10.1: EMPS: Cross-test validation between the simulated and measured data.

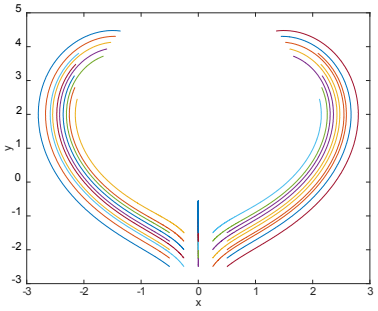

Fig. 10.2: Twenty five trajectories corresponding to System 1.

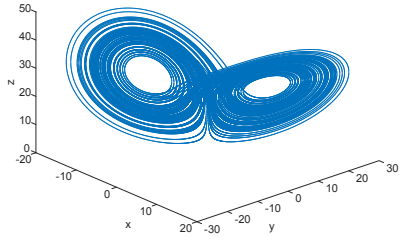

Fig. 10.3: A single trajectory for the three dimensional Lorenz system given in Example 2.

The adjustment of several parameters affect the accuracy of the determine parameters, θ. The most obvious impact on the accuracy of the parameters arises through the selection of the reproducing kernel. While theoretically it is established that Liouville operators with polynomial symbols are densely defined over the exponential dot product kernel's native space, the exponential dot product kernel suffers from poor conditioning. This poor conditioning can lead to inaccuracies that appear from numerical uncertainties in the expression of the (left) inverse matrix for $\mathbf{A}$ in (8.2). The Gaussian RBF exhibits less conditioning issues than the exponential dot product kernel, especially when a small kernel width is selected. In the case of the Gaussian

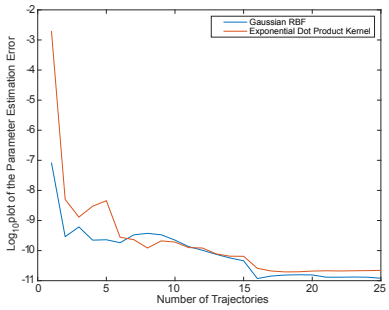

Fig. 10.4: A log-plot of the parameter estimation error, $\|\theta - \hat{\theta}\|_2$, for System 1 using exponential (red) and Gaussian RBF (blue) kernels with varying number of trajectories.

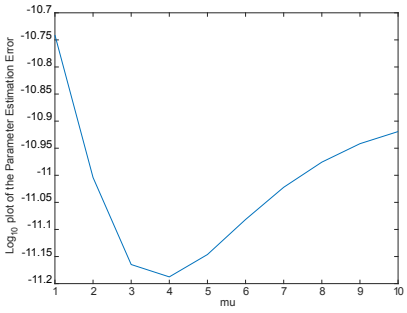

Fig. 10.5: A log-plot of the parameter estimation error versus adjustments in the kernel width μ for the Gaussian RBF applied to System 1.

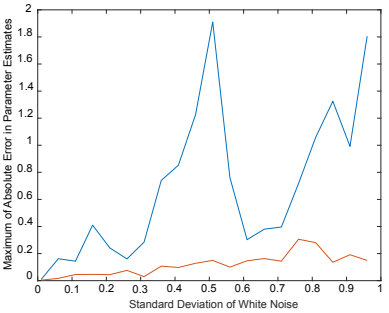

Fig. 10.6: This figure shows the absolute errors of the parameter estimates for System 1 obtained using noisy data (blue), and data filtered using a 20 point moving average filter (red).

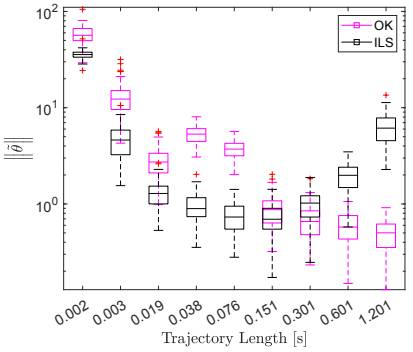

Fig. 10.7: Log-scale box plots of the parameter estimation error committed by the developed occupation kernel (OK) method (magenta) and the ILS method in (3.5) (black) algorithms in 100 repeated trials. See Experiment 3 for details.

RBF, the size of the kernel width has an impact on the accuracy of the system identification method as shown in Figure 10.5. Specifically, occupation kernels corresponding to Gaussian RBFs with smaller kernel widths can distinguish nearby trajectories more effectively than those with larger kernel widths, which leads to better conditioning of $\mathbf{A}$ in (8.2). However, it is well known in approximation contexts that larger values of μ lead to faster convergence [11]. The minimum error at $\mu = 4$ in Figure 10.5 thus strikes a balance between the conditioning of the matrix and the advantages gained from larger μ.

The most significant contribution to errors in the estimation of the parameters is the method of numerical integration performed. The simple example presented

Monomial	Dim	No Noise	Noise	Moving Average	Target
$x_1^0x_2^0$	1	8.882e-16	8.051e-3	-2.093e-3	0
$x_1^1x_2^0$	**1**	**2.000e+0**	**2.000e+0**	**2.000e+0**	**2**
$x_1^2x_2^0$	1	8.882e-16	-3.840e-3	1.346e-3	0
$x_1^0x_2^1$	1	-1.554e-15	3.968e-3	-1.523e-3	0
$x_1^1x_2^1$	**1**	**-1.000e+0**	**-9.988e-1**	**-9.994e-1**	**-1**
$x_1^0x_2^2$	1	-6.939e-17	-2.173e-4	-3.066e-4	0
$x_1^0x_2^0$	2	-8.691e-12	-7.179e-3	-1.363e-3	0
$x_1^1x_2^0$	2	0	1.471e-3	-2.271e-4	0
$x_1^2x_2^0$	**2**	**2.000e+0**	**2.003e+0**	**2.001e+0**	**2**
$x_1^0x_2^1$	**2**	**-1.000e+0**	**-1.007e+0**	**-1.001e+0**	**-1**
$x_1^1x_2^1$	2	-8.327e-17	1.834e-4	9.210e-5	0
$x_1^0x_2^2$	2	2.652e-12	8.280e-4	-2.650e-4	0
Max Error		**8.691e-12**	**8.051e-3**	**2.093e-3**	

Table 10.2: This table presents the results of the nonlinear system identification method applied to the trajectories presented in Figure 10.2. The target parameters are listed in the last column, and the columns "No Noise," "Noise," and "Moving Average" show the obtained parameters from their respective experiments. The "Monomial" column lists the specific basis function that the parameter of that row is tied to, and "Dim" expresses which dimension that particular basis function is contributing to. Note that the bolded rows correspond to the non-zero target values. The presented results demonstrate that even in the case of unfiltered noise, the nonlinear system identification method of Aim 1 obtains parameter estimates while committing an error of a most $8.051e-3$.

in Proposition 6.2 gives an estimation of the occupation kernel via a right hand rule method of numerical integration, and while Proposition 6.2 provides a proof of concept demonstrating norm convergence to the occupation kernel in question, it is observed in (8.3) that this method results in a relatively slow convergence rate. When other methods, such as the trapezoid or Simpson's rule is leveraged for numerical integration, a significant improvement in the performance of the system identification method may be realized.Consequently, the fourth order method of Simpson's rule was utilized for most of the results presented in Section 10.

The results of Experiment 3 are summarized in Figure 10.7, which shows box plots for the norm of the parameter estimation error committed by the two methods for each of the nine selected combinations of N and T. Box plots corresponding to the developed method are in magenta and those corresponding to the ILS method are in black. The trend in Figure 10.7 indicates that for a large number of short trajectories, ILS out-performs the developed occupation kernel method, and for a small number of long trajectories, the developed occupation kernel method out-performs ILS. *However, as indicated by Table 11.1, the average parameter estimation error committed by the developed occupation kernel method with 20 trajectories, 1.2s long, is smaller than the average errors committed by all nine ILS implementations.* The improvement is attributed to the ability of the developed framework to extract more information from the same set of trajectories by integrating them against different test functions.

Two other factors that contribute to the success of the system identification algorithm of Section 8 are the selection of the centers of the reproducing kernels as well as

T	0.002	0.003	0.019	0.038	0.076	0.151	0.301	0.601	1.201
OK	57.873	12.853	2.835	5.259	3.751	0.918	0.688	0.584	**0.489**
ILS	35.718	4.763	1.315	0.945	0.758	**0.727**	0.998	1.957	6.268

Table 11.1: Mean of the norm of the parameter estimation error over 100 repeated trials with different combinations of trajectory lengths and number of trajectories. Across all combinations, the smallest norm of the parameter estimation error for the developed occupation kernel (OK) method method is found to be 0.489 and that for the ILS method in (3.5) is found to be 0.727. See Experiment 3 for details.

the number of trajectories. The contribution of the Gaussian RBFs are largest when the centers are distributed over the working space containing the trajectories. That is, if the centers are too far away from the trajectories, the decay of the Gaussian RBFs will lead to near zero row vectors of $\mathbf{A}$ in (8.2). For the algorithm in Section 8, each reproducing kernel is evaluated for every trajectory, but this isn't technically necessary and reproducing kernels that will contribute less or redundant information may be ignored for a specific trajectory.

Experiment 5 provides a real world test of the system identification method presented in this manuscript. Several features appear in the dynamics of this system that are not present in the other experiments including discontinuities in the acceleration. Table 10.1 demonstrates that performance of the present method closely matches that of the benchmark method from [17]. It should be noted that the poor performance of both methods in estimating the acceleration occurs because of the discontinuities in acceleration, where sharp "ringing" phenomena occur at the jumps in acceleration. This occurs as the basis functions used to approximate the dynamics corresponding to acceleration are continuous.

The key difference between existing kernel based approaches and the present approach is that the basis functions are separated from data integration. Whereas typical kernel based approaches leverage the representer theorem to yield an approximation of the dynamics with respect to a linear combination of reproducing kernels centered at the data points (cf. [30]). The method presented in this manuscript leverages the form of the occupation kernel to incorporate the trajectory inside a RKHS, and the selection of occupation kernel is largely independent of the selection of basis functions. In the present context, the selection of basis functions is such that the functions should give a densely defined Liouville operator in the selected kernel space to guarantee the soundness of the developed methods. The advantage gained in the use of occupation kernels is that the occupation kernel itself is not as strongly influenced by noisy measurements as the kernel counterparts, since it can be represented as the integral of kernels that have centers along the trajectory.

12. Conclusion. In this manuscript a new approach to system identification was developed through the use of Liouville operators and occupation kernels over a RKHS. Liouville operators are densely defined operators whose adjoint contains occupation kernels corresponding to solutions to differential equations within its domain. Hence, a dynamical system may be embedded into a RKHS where methods of numerical analysis, machine learning, and approximation theory affiliated with RKHSs may be brought to bear on problems in dynamical systems theory. Specifically, an inner product on dynamical systems that give rise to densely defined Liouville operators was developed, where the projection of a dynamical system on a linear combination

of basis function is realized through the solution of a parameter identification routine that generalized that found in [5] through a weak formulation using integrals. In this setting, the features of a reproducing kernel become test functions for designing contraints on parameters for system identification.

The domain of Liouville operators depends nontrivially on the selection of RKHS. It was demonstrated that Liouville operators with polynomial symbols are densely defined over the RKHS corresponding to the exponential dot product kernel. Moreover, it was demonstrated in the system identification routine that the selection of reproducing kernel may have an effect on the results of parameter estimation.

The system identification method developed in the manuscript was validated on a two dimensional and a three dimensional system through several different experiments designed to evaluate the effects of various integration and RKHS parameters, such as kernel width for the Gaussian RBF, the selection of numerical integration scheme, the selection of kernel, and so on. Through each experiment, accurate estimations of the parameters were achieved. However, it was demonstrated that the largest error source arises through the choice of numerical integration method, where Simpson's rule provided the most accurate results.

REFERENCES

[1] *The mathematical theory of the dynamics of biological populations*, Academic Press, London-New York, 1973. Based on a Conference held in Oxford, September, 1972.

[2] S. A. Billings, *Nonlinear system identification: NARMAX methods in the time, frequency, and spatio-temporal domains*, John Wiley & Sons, 2013.

[3] G. Birpoutsoukis, A. Marconato, J. Lataire, and J. Schoukens, *Regularized nonparametric Volterra kernel estimation*, Automatica, 82 (2017), pp. 324–327.

[4] C. M. Bishop, *Pattern recognition and machine learning*, Springer, 2006.

[5] S. L. Brunton, J. L. Proctor, and J. N. Kutz, *Discovering governing equations from data by sparse identification of nonlinear dynamical systems*, Proceedings of the National Academy of Sciences, 113 (2016), pp. 3932–3937.

[6] A. Caponnetto, C. A. Micchelli, M. Pontil, and Y. Ying, *Universal multi-task kernels*, The Journal of Machine Learning Research, 9 (2008), pp. 1615–1646.

[7] F. P. Carli, T. Chen, and L. Ljung, *Maximum entropy kernels for system identification*, IEEE Transactions on Automatic Control, 62 (2016), pp. 1471–1477.

[8] E. A. Coddington and N. Levinson, *Theory of ordinary differential equations*, Tata McGraw-Hill Education, 1955.

[9] W. E. Dixon, A. Behal, D. M. Dawson, and S. P. Nagarkatti, *Nonlinear control of engineering systems: a Lyapunov-based approach*, Springer Science & Business Media, 2013.

[10] G. Fasshauer and M. McCourt, *Kernel-based Approximation Methods using MATLAB*, World Scientific Publishing Company, 2015.

[11] G. E. Fasshauer, *Meshfree approximation methods with MATLAB*, vol. 6, World Scientific, 2007.

[12] H. I. Freedman, *Deterministic mathematical models in population ecology*, vol. 57, Marcel Dekker Incorporated, 1980.

[13] D. Giannakis, S. Das, and J. Slawinska, *Reproducing kernel hilbert space compactification of unitary evolution groups*, arXiv preprint arXiv:1808.01515, (2018).

[14] G.-O. Glentis, P. Koukoulas, and N. Kalouptsidis, *Efficient algorithms for volterra system identification*, IEEE Transactions on Signal Processing, 47 (1999), pp. 3042–3057.

[15] M. S. Hemati, M. O. Williams, and C. W. Rowley, *Dynamic mode decomposition for large and streaming datasets*, Physics of Fluids, 26 (2014), p. 111701.

[16] E. M. Izhikevich and J. Moehlis, *Dynamical systems in neuroscience: The geometry of excitability and bursting*, SIAM review, 50 (2008), p. 397.

[17] A. Janot, G. M., and M. Brunot, *Data set and reference models of EMPS*, in 2019 Workshop on Nonlinear System Identification Benchmarks, Eindhoven, The Netherlands, Apr. 2019.

[18] R. Kamalapurkar, P. Walters, J. Rosenfeld, and W. Dixon, *Reinforcement Learning for Optimal Feedback Control: A Lyapunov-Based Approach*, Springer, 2018.

[19] H. K. Khalil, *Nonlinear systems*, Prentice-Hall, 2002.

[20] J. B. Lasserre, D. Henrion, C. Prieur, and E. Trélat, *Nonlinear optimal control via occupation measures and lmi-relaxations*, SIAM journal on control and optimization, 47 (2008), pp. 1643–1666.
[21] F. Lu and H. Sun, *Positive definite dot product kernels in learning theory*, Advances in Computational Mathematics, 22 (2005), pp. 181–198.
[22] A. Mauroy and J. Goncalves, *Koopman-based lifting techniques for nonlinear systems identification*, IEEE Trans. Autom. Control, (2020). to appear.
[23] D. A. Messenger and D. M. Bortz, *Weak SINDy for partial differential equations*, Journal of Computational Physics, 443 (2021), p. 110525.
[24] D. A. Messenger and D. M. Bortz, *Weak SINDy: Galerkin-based data-driven model selection*, Multiscale Modeling & Simulation, 19 (2021), pp. 1474–1497.
[25] O. Nelles, *Nonlinear system identification: from classical approaches to neural networks and fuzzy models*, Springer Science & Business Media, 2013.
[26] A. Parikh, R. Kamalapurkar, and W. E. Dixon, *Integral concurrent learning: adaptive control with parameter convergence using finite excitation*, International Journal of Adaptive Control and Signal Processing, 33 (2019), pp. 1775–1787.
[27] V. I. Paulsen and M. Raghupathi, *An introduction to the theory of reproducing kernel Hilbert spaces*, vol. 152, Cambridge University Press, 2016.
[28] G. K. Pedersen, *Analysis now*, vol. 118, Springer Science & Business Media, 2012.
[29] L. Perko, *Differential equations and dynamical systems*, vol. 7, Springer Science & Business Media, 2013.
[30] G. Pillonetto, F. Dinuzzo, T. Chen, G. De Nicolao, and L. Ljung, *Kernel methods in system identification, machine learning and function estimation: A survey*, Automatica, 50 (2014), pp. 657–682.
[31] J. A. Rosenfeld, *Densely defined multiplication on several sobolev spaces of a single variable*, Complex Analysis and Operator Theory, 9 (2015), pp. 1303–1309.
[32] J. A. Rosenfeld, *Introducing the polylogarithmic hardy space*, Integral Equations and Operator Theory, 83 (2015), pp. 589–600.
[33] J. A. Rosenfeld, *The sarason sub-symbol and the recovery of the symbol of densely defined toeplitz operators over the hardy space*, Journal of Mathematical Analysis and Applications, 440 (2016), pp. 911–921.
[34] J. A. Rosenfeld, R. Kamalapurkar, B. Russo, and T. T. Johnson, *Occupation kernels and densely defined Liouville operators for system identification*, in Proc. IEEE Conf. Decis. Control, Dec. 2019, pp. 6455–6460.
[35] J. A. Rosenfeld, B. Russo, R. Kamalapurkar, and T. T. Johnson, *The occupation kernel method for nonlinear system identification.* arXiv:1909.11792, 2019.
[36] B. Russo and M. P. Laiu, *Convergence of weak-SINDy surrogate models.* arXiv:2209.15573, 2022.
[37] B. P. Russo and J. A. Rosenfeld, *Liouville operators over the hardy space*, Journal of Mathematical Analysis and Applications, 508 (2022), p. 125854.
[38] D. Sarason, *Unbounded toeplitz operators*, Integral Equations and Operator Theory, 61 (2008), pp. 281–298.
[39] J. Schoukens and L. Ljung, *Nonlinear system identification: A user-oriented road map*, IEEE Control Syst. Mag., 39 (2019), pp. 28–99.
[40] A. Sogokon, K. Ghorbal, and T. T. Johnson, *Non-linear Continuous Systems for Safety Verification (Benchmark Proposal)*, in International Workshop on Applied Verification for Continuous and Hybrid Systems, G. Frehse and M. Althoff, eds., vol. 43 of EPiC Series in Computing, Vienna, Austria, Apr. 2016, EasyChair, pp. 42–51.
[41] I. Steinwart and A. Christmann, *Support vector machines*, Springer Science & Business Media, 2008.
[42] D. E. Stewart, R. L. Dewar, et al., *Non-linear dynamics*, Complex systems, (2000), p. 167.
[43] R. Taylor, J. N. Kutz, K. Morgan, and B. A. Nelson, *Dynamic mode decomposition for plasma diagnostics and validation*, Review of Scientific Instruments, 89 (2018), p. 053501.
[44] R. Tibshirani, *Regression shrinkage and selection via the lasso*, Journal of the Royal Statistical Society: Series B (Methodological), 58 (1996), pp. 267–288.
[45] M. O. Williams, I. G. Kevrekidis, and C. W. Rowley, *A data–driven approximation of the Koopman operator: Extending dynamic mode decomposition*, Journal of Nonlinear Science, 25 (2015), pp. 1307–1346.
[46] M. O. Williams, C. W. Rowley, and I. G. Kevrekidis, *A kernel-based method for data-driven Koopman spectral analysis*, J. Comput. Dyn., 2 (2015), pp. 247–265.