Occupation Kernels and Densely Defined Liouville Operators for System Identification

Joel A. Rosenfeld[1] Rushikesh Kamalapurkar[2] Benjamin Russo[3] Taylor T. Johnson[1]

***Abstract*— This manuscript introduces the concept of Liouville operators and occupation kernels over reproducing kernel Hilbert spaces (RKHSs). The combination of these two concepts allow for the embedding of a dynamical system into a RKHS, where function theoretic tools may be leveraged for the examination of such systems. These tools are then turned toward the problem of system identification where an inner product formula is developed to provide constraints on the parameters in a system identification setting. This system identification routine is validated through several numerical experiments, where each experiment examines various contributions to the parameter identification error via numerical integration methods and parameters for the kernel functions themselves.**

I. Introduction

A dynamical system is given as $\dot{x} = f(x)$, 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 [1], [2], [3], biology [4], [5], neuroscience [6], physics [7], and mathematics [8], [9]. However, in many cases even physically motivated dynamical systems can have unknown parameters (i.e. a gray box), such as mass and length of mechanical components, or the dynamics may be completely unknown (i.e. a black box) [10]. In such cases, system identification methods are leveraged to gain estimates on the dynamics of the system based on data generated by the system itself [10].

For linear dynamics, many classical tools are available for systems identification through the Fourier and Laplace transforms of the dynamical systems by the exploitation of the impulse response, and linear system identification still remains numerically challenging. On the other hand, the identification of nonlinear systems proves even more challenging as nonlinearities may manifest in a variety of ways, and linear transform methods for general nonlinear systems are unavailable [11], [10].

To address these challenges a variety of nonlinear system identification methods have been developed, such as NARMAX methods [11], Volterra series [12], Lyapunov methods [13], and Neural Networks [10]. However, given the rich variety of nonlinear systems, there is no modal approach to resolving the system identification problem for nonlinear systems [10]. A recent development in nonlinear systems identification was the introduction of dynamic mode decompositions (DMDs) and their connection with the Koopman operator [14], [15], [16].

One technical challenge that arises in many of the system identification methods described above comes from the estimation of the state derivative [14], [13]. Frequently only the state trajectory is available and numerical estimation methods are employed to obtain samples of the dynamical system. Unfortunately, state derivative estimates are prone to error, and the use of numerical estimates of the state derivatives introduce an artificial noise component that requires additional filtering before it may be used as an estimate of the dynamical system [14].

In an online parameter estimation context, [13] 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 stable under the integral concurrent learning method [13].

The present manuscript develops a method that is close in spirit to the integral concurrent learning method for system identification. Specifically, the method presented in Section V leverages novel kernel techniques presented in Section III, 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 [17]. The present manuscript lifts the theory of occupation measures to that of function theory by examining the integration functionals over reproducing kernel Hilbert spaces (RKHSs) rather than the Banach spaces of continuous functions. What is gained by restricting the examination to that of Hilbert function spaces is that the tools of approximation and function theory can be brought to bear on those of occupation kernels, where these tools were much more limited in their scope for occupation measures. That is, 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,

*This research was supported by the Air Force Office of Scientific Research (AFOSR) under contract numbers FA9550-15-1-0258, FA9550-16-1-0246, and FA9550-18-1-0122. 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.

[1]Institute for Software Integrated Systems, Department of Electrical Engineering and Computer Science, Vanderbilt University, Nashville, TN 37212 USA

[2]Department of Mechanical and Aerospace Engineering, Oklahoma State University, Stillwater, OK 74078 USA

[3]Department of Mathematics, Farmingdale State College, Farmingdale, NY 11735 USA

which allows for different aspects of the trajectory to be emphasized. In constrast, 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 principle reason for this limitation is that these methods aim to exploit the moment problem for occupation measures.

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 the dynamics in a Hilbert space setting.
- Occupation measures are generalized to occupation kernels, where a trajectory is represented inside a Hilbert space as a function.
- Occupation kernels and Liouville operators are leveraged to provide constraints for a system identification method, which is presented in Section V. These constraints use more general test functions than polynomials, which is an advantage that arises in the use of occupation kernels over occupation measures.

The manuscript is organized as follows. Preliminaries necessary for the development of occupation kernels and densely defined Liouville operators are presented in Section II, and the densely defined Liouville operators and occupation kernels themselves are introduced in Section III. These tools are then turned toward the problem of system identification, where the dynamics of a system are parameterized into a collection of basis functions. The tools of Section III are leveraged to provide a collection of linear constraints on the parameters of the dynamics, where state derivatives are replaced via a collection of integral constraints. The system identification approach is then examined through a collection of numerical experiments in Section VI and the experiments are discussed in Section VII.

II. Preliminaries

A. Reproducing Kernel Hilbert Spaces

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 $E_x g := g(x)$ is bounded [18]. As such, 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. The function k_x is called the reproducing kernel function at x, and the function $K(x, y) = \langle k_y, k_x \rangle_H$ is called the kernel function corresponding to H.

This manuscript utilizes two RKHSs, which are defined through their kernel functions [18]. For $\mu > 0$, the kernel function $K_E(x, y) = e^{\mu x^T y}$ is called the exponential dot product kernel function, and for $\mu > 0$, the kernel function given as $K_G(x, y) = \exp\left(-\frac{1}{\mu}\|x - y\|_2^2\right)$ is called a Gaussian radial basis function. Both K_E and K_G are kernel functions for RKHSs over $\mathbb{R}^n$ [19].

B. Densely Defined Operators

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 [20, Chapter 5]. The operator T is closed if, for every sequence $\{g_m\}_{m=0}^{\infty} \subset \mathcal{D}(T)$, such that $g_m \to g \in H$ and $Tg_m \to h \in H$, then $g \in \mathcal{D}(T)$ and $Tg = h$.

The adjoint of a possibly unbounded operator is given first by its domain: $\mathcal{D}(T^*) = \{g \in H : h \mapsto \langle Th, g \rangle_H$ is bounded over $\mathcal{D}(T)\}$ [20]. For each $g \in \mathcal{D}(T^*)$ there exists a member $T^*g \in H$ such that $\langle Th, g \rangle_H = \langle h, T^*g \rangle_H$. Thus, the operator T^* may be defined as taking $g \in \mathcal{D}(T^*)$ to T^*g, which was obtained through the Riesz representation theorem. The closedness of the operator guarantees the nonemptiness of the domain of its adjoint. In fact, the following stronger statement holds.

Lemma 1: (c.f. [20, Chapter 5]) The adjoint of a closed operator is densely defined.

III. 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 [17].

Definition 2: Let $\dot{x} = f(x)$ be a dynamical system with the dynamics, $f : \mathbb{R}^n \to \mathbb{R}^n$, 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

$$\mathcal{D}(A_f) := \{g \in H : \nabla_x g \cdot f \in H\}.$$

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 exponential dot product kernel's native RKHS [19]. 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. Adjusting the RKHS will also adjust the Liouville operators that are admissible. 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 [19], it can be similarly established that A_f is closed under the same circumstances.

Theorem 3: 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 [19, 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 Lemma 1. Associated with Liouville operators in particular are a special class of functions within the domain of the Liouville operators' adjoints, and these functions are also the main object of study of this manuscript.

Definition 4: 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, and may be respresented 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: 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 4 that satisfies $\dot{\gamma} = f(\gamma)$, then $\Gamma_\gamma \in \mathcal{D}(A_f^*)$.

Proof: The establishment of $\Gamma_\gamma \in \mathcal{D}(A_f^*)$, requires the demonstration that the functional $g \mapsto \langle A_f g, \Gamma_\gamma \rangle_H$ is bounded over $\mathcal{D}(A_f)$. Note that

$$\int_0^T \nabla_x g(\gamma(t)) f(\gamma(t)) dt = g(\gamma(T)) - g(\gamma(0)) \tag{1}$$

as the integrand of (1) is the total derivative of $g(\gamma(t))$. The left hand side of (1) may be expressed as $\langle A_f g, \Gamma_\gamma \rangle_H$, while the right hand side satisfies the bound

$$\begin{aligned} |g(\gamma(T)) - g(\gamma(0))| &= |\langle g, K(\cdot, \gamma(T)) - K(\cdot, \gamma(0)) \rangle_H| \\ &\leq \|g\|_H \|K(\cdot, \gamma(T)) - K(\cdot, \gamma(0))\|_H, \end{aligned}$$

which establishes the boundedness of $g \mapsto \langle A_f g, \Gamma_\gamma \rangle_H$. ■

Proposition 5 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

$$\langle A_f g, \Gamma_\gamma \rangle_H = g(\gamma(T)) - g(\gamma(0)) \text{ for all } g \in \mathcal{D}(A_f)$$

uniquely determines Γ_γ. Consequently, this relation will be used subsequently to establish constraints for parameter identification in a system identification setting.

IV. Estimation of Occupation Kernels

Approximating the value of an inner product against an occupation kernel in a RKHS can be performed leveraging quadrature techniques for integration. The occupation kernels themselves can be expressed as an integral against the kernel function in a RKHS as demonstrated in Proposition 6.

Proposition 6: 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 4. The occupation kernel corresponding to γ in H, Γ_γ, may be expressed as

$$\Gamma_\gamma(x) = \int_0^T K(x, \gamma(t)) dt. \tag{2}$$

Proof: Note that $\Gamma_\gamma(x) = \langle \Gamma_\gamma, K(\cdot, x) \rangle_H$, by the reproducing property of K. Consequently,

$$\begin{aligned} \Gamma_\gamma(x) &= \langle \Gamma_\gamma, K(\cdot, x) \rangle_H = \langle K(\cdot, x), \Gamma_\gamma \rangle_H \\ &= \int_0^T K(\gamma(t), x) dt = \int_0^T K(x, \gamma(t)) dt, \end{aligned}$$

which establishes the result. ■

Leveraging Proposition 6, 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 7: Under the hypothesis of Proposition 6, 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

$$\hat{\Gamma}_\gamma(x) := \sum_{i=1}^{F} (t_i - t_{i-1}) K(x, \gamma(t_i)). \tag{3}$$

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

Proof: Consider,

$$\|\Gamma_\gamma - \hat{\Gamma}_\gamma\|_H^2 = \|\Gamma_\gamma\|^2 + \|\hat{\Gamma}_\gamma\|^2 - 2\langle \Gamma_\gamma, \hat{\Gamma}_\gamma \rangle_H.$$

The norm of the approximation can be expanded as

$$\begin{aligned} \|\hat{\Gamma}_\gamma\|_H^2 &= \langle \hat{\Gamma}_\gamma, \hat{\Gamma}_\gamma \rangle_H = \\ \sum_{i=1}^{F} \sum_{j=1}^{F} &(t_i - t_{i-1})(t_j - t_{j-1}) K(\gamma(t_j), \gamma(t_i)) \end{aligned} \tag{4}$$

via the reproducing property of K. Now compare each term in (4) to the corresponding integral,

$$\begin{aligned} &\int_{t_{i-1}}^{t_i} \int_{t_{j-1}}^{t_j} K(\gamma(t), \gamma(\tau)) dt d\tau \\ &-(t_i - t_{i-1})(t_j - t_{j-1}) K(\gamma(t_j), \gamma(t_i)). \end{aligned} \tag{5}$$

By the mean value theorem, there is a point $(\tau^*, t^*) \in [t_{i-1}, t_i] \times [t_{j-1}, t_j]$ such that

$$\begin{aligned} &\int_{t_{i-1}}^{t_i} \int_{t_{j-1}}^{t_j} K(\gamma(t), \gamma(\tau)) dt d\tau \\ &= (t_i - t_{i-1})(t_j - t_{j-1}) K(\gamma(t^*), \gamma(\tau^*)). \end{aligned}$$

Hence, (5) may be written as

$$(t_i - t_{i-1})(t_j - t_{j-1})(K(\gamma(t^*), \gamma(\tau^*)) - K(\gamma(t_j), \gamma(t_i))).$$

Leveraging the mean value inequality [21],

$$|K(\gamma(t^*), \gamma(\tau^*)) - K(\gamma(t_j), \gamma(t_i))| \leq \sup_{x,y \in X} \|\nabla K(x,y)\|_2 \max_{0<t<T} |\dot{\gamma}(t)| \|(\tau^*, t^*) - (t_i, t_j)\|_2.$$

Taking $h = \max_{i=1,\dots,F} |t_i - t_{i-1}|$ and combining the above equations, it can be observed that

$$\|\hat{\Gamma}_\gamma\|_H^2 = \|\Gamma_\gamma\|_H^2 + O(h).$$

Similarly, it may be demonstrated that $\langle \hat{\Gamma}_\gamma, \Gamma_\gamma \rangle_H = \|\Gamma_\gamma\|_H^2 + O(h)$, and the conclusion follows. ∎

It should be clear from the proof of Proposition 7 that higher order quadrature rules for estimating the integral in (2) 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.

V. System Identification via Reproducing Kernel Hilbert Spaces

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 $i = 1, \dots, M$, as

$$f(x) = \sum_{i=1}^{M} \theta_i Y_i(x). \tag{6}$$

The goal of the system identification problem given a collection of trajectories, $\{\gamma_j\}_{j=1}^N$, satisfying the dynamics as in Definition 4, is to determine the values of the parameters, θ_i for $i = 1, \dots, M$, such that (6) may be used to reproduce the trajectories.

For the sake of the succeeding algorithm, the following assumptions are made 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 $A_{Y_1}, \dots, A_{Y_M}$ have a common dense domain.

Assumption 2: Given a RKHS, H, over a set X, and a collection of Liouville operators, $\{A_{Y_i}\}_{i=1}^M$ satisfying Assumption 1, the collection of kernel functions $\{K(\cdot, c) : c \in X\}$ is contained within $\cap_{i=1}^M \mathcal{D}(A_{Y_i})$.

Assumption 1 ensures the validity of decomposing A_f into a linear combination of densely defined Liouville operators, $\{A_{Y_i}\}_{i=1}^M$. Assumption 1 is pivotal for the system identification approach contained in this manuscript. Assumption 1 provides additional restrictions on the dynamics of the system beyond Lipschitz continuity. Liouville operators are closely connected to densely defined multiplication operators (c.f. [22], [23], [24], [25]), 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.

Assumption 2 asks for the domain of the Liouville operator to contain the kernel functions of the RKHS. These may be replaced by other collections of basis functions that have dense span in the RKHS, such as polynomials. However, it is convenient in that each RKHS has a dense collection of kernel functions, which may be used in this context. Thus, Assumption 2 allows a unifying result that applies to all RKHSs, and it also helps the exposition of this manuscript.

A. Parameter Identification via Occupation Kernels

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, H of continuously differentiable functions over X. Suppose that $\{c_s\}_{s=1}^\infty \subset X$ is dense. Constraints on θ_i are then established as

$$\langle A_f K(\cdot, c_s), \Gamma_{\gamma_j} \rangle_H = \tag{7}$$

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

for each $s = 1, \dots, \infty$ and $j = 1, \dots, N$.

After the selection of a finite and representative collection of centers, $\{c_s\}_{s=1}^S$, (7) may be expressed as a matrix equation. Let $\{n_i\}_{i=1}^{S \cdot N}$ be an enumeration of $\{(s, j)\}_{s=1,j=1}^{S,N}$, then the matrix equation in (8) holds.

$$\mathbf{A}\theta = \mathbf{K}(T) - \mathbf{K}(0), \text{ where} \tag{8}$$

$$\mathbf{A} = \left(\langle A_{Y_i} K(\cdot, c_{n_{j,1}}), \Gamma_{\gamma_{n_{j,2}}} \rangle_H \right)_{j=1,i=1}^{j=SN,i=M} \in \mathbb{R}^{SN \times M},$$

$$\theta = \begin{pmatrix} \theta_1 & \cdots & \theta_M \end{pmatrix}^T \in \mathbb{R}^M, \text{ and}$$

$$\mathbf{K}(t) = \begin{pmatrix} K(\gamma_{n_{1,2}}(t), c_{n_{1,1}}) \\ \vdots \\ K(\gamma_{n_{SN,2}}(t), c_{n_{SN,1}}) \end{pmatrix} \in \mathbb{R}^{SN}.$$

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

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

Hence, by Proposition 7

$$\langle A_{Y_i} K(\cdot, c_s), \hat{\Gamma}_{\gamma_j} \rangle_H = \langle A_{Y_i} K(\cdot, c_s), \Gamma_{\gamma_j} \rangle_H + O\left(\sqrt{h}\right), \tag{9}$$

so that quadrature techniques for the estimation of the inner products contained in (8) can be successfully employed. Note that other quadrature techniques can also lead to better convergence estimates. Since the matrix $\mathbf{A}$ must be numerically estimated, the parameter values obtained using this method are approximate, and will be represented as $\hat{\theta}$.

VI. Numerical Experiments

Two systems were examined to evaluate the system identification method of Section V. For each system, the trajectories were generated using the Runge-Kutta 4 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) 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 [26]. The two dimensional dynamics are given as

$$\dot{x} = f(x) = \begin{pmatrix} 2x_1 - x_1x_2 \\ 2x_1^2 - x_2 \end{pmatrix}. \tag{10}$$

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 1.

The centers for the kernel functions 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 V. In this experiment two kernel functions 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 3.

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 4.

System 2: The second system is the three dimensional Lorenz system [9], [14],

$$\dot{x} = f(x) = \begin{pmatrix} \sigma(x_2 - x_1) \\ x_1(\rho - x_3) - x_2 \\ x_1x_2 - \beta x_3 \end{pmatrix}. \tag{11}$$

Following [14] 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 $x_0 = (-8, 7, 27)^T$. The plot of this trajectory is given in Figure 2.

Numerical Method	Convergence Order	Error $\|\theta - \hat{\theta}\|_2$
Right Hand Rule	$O(h)$	$2.1696E + 0$
Trapezoid Rule	$O(h^2)$	$3.8136E - 2$
Simpson's Rule	$O(h^4)$	$7.6920E - 5$

TABLE I

This table presents a comparison between the errors in parameter estimation based on the selection of typical numerical integration schemes for the system identification routine for System 2. Each numerical integration scheme is listed along with the convergence rate of the algorithm. Of the three routines, the Simpson's rule demonstrates the strongest performance. The step-size was kept consistent between each experiment at $h = 0.001$.

Number of Segments	Error $\|\theta - \hat{\theta}\|_2$
1	$7.6920E - 5$
10	$5.2175E - 6$
100	$5.8506E - 6$

TABLE II

This table contrasts the parameter estimation errors committed by the system identification routine applied to System 2 when the single trajectory is segmented into smaller trajectories. It can be observed that an order of mangnitude improvement was realized when the single trajectory was segmented into 10 and 100 trajectories. However, there was no improvement in the error when using 100 segments over 10 segments.

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 investigates the error contribution committed by the use of different numerical integration schemes. In this setting System 2 was identified using the Gaussian RBFs with kernel width $\mu = 10$. The results are displayed in Table I.

Experiment 4: The last experiment is a product of the method used to generation of the trajectory data. Runge-Kutta 4 has a high rate of convergence. However, as with any time-stepping method the global error bound is proportional e^{LT} where L is the Lipschitz constant of the dynamics [27]. As such, the accumulated global error could be large in the long term evaluation of the trajectory of System 2. Experiment investigates the effect on the error when the trajectory of System 2 is segmented into smaller trajectories. Each smaller trajectory is then treated as a new initial value problem with a smaller time horizon and thus a hypothetically smaller global error. Here the Gaussian RBF was leveraged in the system identification algorithm of Section V with kernel width $\mu = 10$.

VII. Discussion

It may be observed through the numerical experiments performed in Section VI that the system identification method of Section V is effective at identifying the parameters for

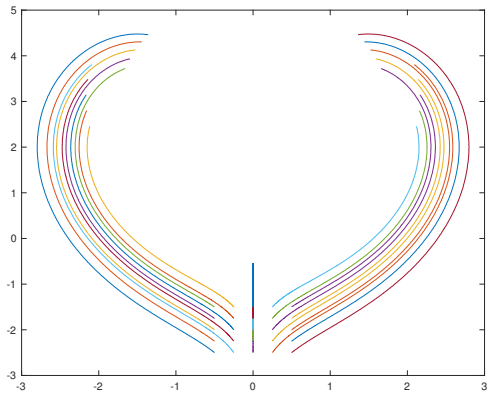

Fig. 1. Twenty five trajectories corresponding to System 1.

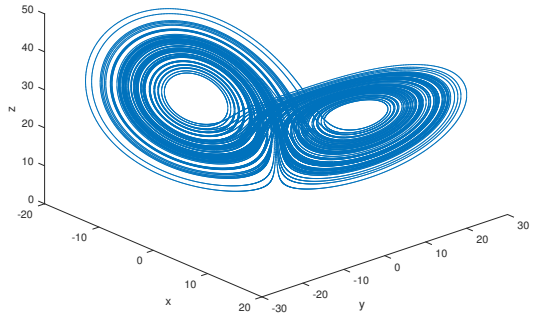

Fig. 2. A single trajectory for the three dimensional Lorenz system given in Example 2.

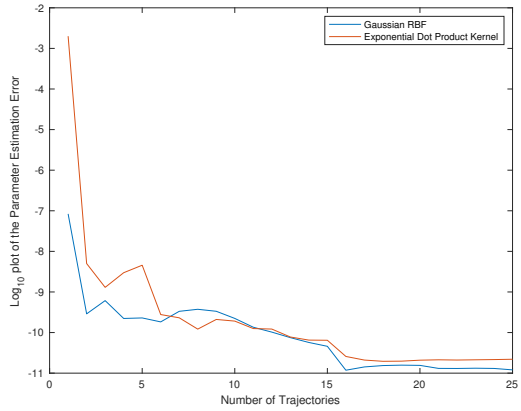

Fig. 3. A log-plot of the parameter estimation error, $\|\theta - \hat{\theta}\|_2$, for System 1 committed by the system identification method in Section V as determined by the number of trajectories utilized by the method. It may be observed that an accurate estimate of θ is established using a single trajectory. However, the inclusion of additional data dramatically improves the parameter estimation error. The two graphs represent the results for two different kernel functions, with a slight advantage exhibited by the Gaussian RBF.

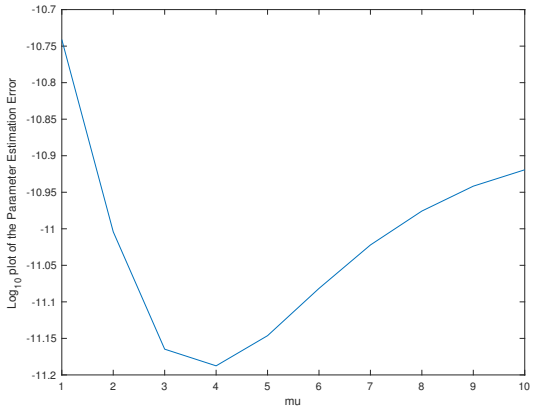

Fig. 4. A log-plot of the parameter estimation error versus adjustments in the kernel width μ for the Gaussian RBF applied to System 1. Note that the error is small for any selection of μ, but reaches its minimum near $\mu = 4$. Additionally note that 4 is approximately the radius of the workspace. One reason for the loss of accuracy at larger μ values could be poorer conditioning of the matrix $\mathbf{A}$ for large μ.

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 VI 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.

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 kernel function. 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). 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 RBF, the size of the kernel width has an impact on the accuracy of the system identification method as shown in Figure 4. 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). However, it is well known in approximation contexts that larger values of μ lead to faster convergence [28]. The minimum error at $\mu = 4$ in

Figure 4 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 in Proposition 7 gives an estimation of the occupation kernel via a right hand rule method of numerical integration, and while Proposition 7 provides a proof of concept demonstrating norm convergence to the occupation kernel in question, it is observed in (9) 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 as demonstrated in Table I. Consequently, the fourth order method of Simpson's rule was utilized for most of the results presented in Section VI.

Two other factors that contribute to the success of the system identification algorithm of Section V are the selection of the centers of the kernel functions as well as 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). For the algorithm in Section V, each kernel function is evaluated for every trajectory, but this isn't technically necessary and kernel functions that will contribute less or redundant information may be ignored for a specific trajectory.

If only a single trajectory is available from a system, as with the Lorenz example in Section VI, then this trajectory may be segmented to provide more constraints in $\mathbf{A}$ of (8). It was observed that segmenting the long trajectory of System 2 improved the parameter estimation error as presented in Table II. This improvement in likely due to the accumulated global error due to numerical time stepping methods in the generation of the trajectory itself. Through segmentation, each segment is treated as a new initial value problem with a smaller time horizon and thus a smaller accumulated global error. Therefore the elements of $\mathbf{A}$ in (8) are closer to the true values of the dynamical system.

VIII. 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.

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 function. Moreover, it was demonstrated in the system identification routine that the selection of kernel function 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] R. Kamalapurkar, P. Walters, J. Rosenfeld, and W. Dixon, *Reinforcement Learning for Optimal Feedback Control: A Lyapunov-Based Approach*. Springer, 2018.

[2] 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.

[3] H. K. Khalil, "Nonlinear systems," 2002.

[4] M. S. Bartlett and R. W. Hiorns, Eds., *The mathematical theory of the dynamics of biological populations*. Academic Press, London-New York, 1973, based on a Conference held in Oxford, September, 1972.

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

[6] E. M. Izhikevich and J. Moehlis, "Dynamical systems in neuroscience: The geometry of excitability and bursting," *SIAM review*, vol. 50, no. 2, p. 397, 2008.

[7] D. E. Stewart, R. L. Dewar *et al.*, "Non-linear dynamics," *Complex systems*, p. 167, 2000.

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

[9] L. Perko, *Differential equations and dynamical systems*. Springer Science & Business Media, 2013, vol. 7.

[10] O. Nelles, *Nonlinear system identification: from classical approaches to neural networks and fuzzy models*. Springer Science & Business Media, 2013.

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

[12] G.-O. Glentis, P. Koukoulas, and N. Kalouptsidis, "Efficient algorithms for volterra system identification," *IEEE Transactions on Signal Processing*, vol. 47, no. 11, pp. 3042–3057, 1999.

[13] 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*, vol. 0, no. 0.

[14] 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*, vol. 113, no. 15, pp. 3932–3937, 2016.

[15] M. S. Hemati, M. O. Williams, and C. W. Rowley, "Dynamic mode decomposition for large and streaming datasets," *Physics of Fluids*, vol. 26, no. 11, p. 111701, 2014.

[16] R. Taylor, J. N. Kutz, K. Morgan, and B. A. Nelson, "Dynamic mode decomposition for plasma diagnostics and validation," *Review of Scientific Instruments*, vol. 89, no. 5, p. 053501, 2018.

[17] 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*, vol. 47, no. 4, pp. 1643–1666, 2008.

[18] V. I. Paulsen and M. Raghupathi, *An introduction to the theory of reproducing kernel Hilbert spaces*. Cambridge University Press, 2016, vol. 152.

[19] I. Steinwart and A. Christmann, *Support vector machines*. Springer Science & Business Media, 2008.

[20] G. K. Pedersen, *Analysis now*. Springer Science & Business Media, 2012, vol. 118.

[21] W. Rudin *et al.*, *Principles of mathematical analysis*. McGraw-hill New York, 1964, vol. 3.
[22] J. A. Rosenfeld, "Densely defined multiplication on several sobolev spaces of a single variable," *Complex Analysis and Operator Theory*, vol. 9, no. 6, pp. 1303–1309, 2015.
[23] ——, "Introducing the polylogarithmic hardy space," *Integral Equations and Operator Theory*, vol. 83, no. 4, pp. 589–600, 2015.
[24] ——, "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*, vol. 440, no. 2, pp. 911–921, 2016.
[25] D. Sarason, "Unbounded toeplitz operators," *Integral Equations and Operator Theory*, vol. 61, no. 2, pp. 281–298, 2008.
[26] A. Sogokon, K. Ghorbal, and T. T. Johnson, "Non-linear Continuous Systems for Safety Verification (Benchmark Proposal)," in *ARCH@CPSWeek 2016 - 3rd International Workshop on Applied Verification for Continuous and Hybrid Systems*, ser. EPiC Series in Computing, G. Frehse and M. Althoff, Eds., vol. 43. Vienna, Austria: EasyChair, Apr. 2016, pp. 42–51. [Online]. Available: https://hal.archives-ouvertes.fr/hal-01660900
[27] K. E. Atkinson, *An introduction to numerical analysis*. John wiley & sons, 2008.
[28] G. E. Fasshauer, *Meshfree approximation methods with MATLAB*. World Scientific, 2007, vol. 6.