

Fractional Order System Identification with Occupation Kernel Regression*

Xiuying Li¹ and Joel A Rosenfeld²

Abstract—While fractional order systems have been employed broadly throughout science and engineering, system identification aimed at nonlinear fractional order dynamical systems remain in their infancy. One reason for this is that local estimates cannot be used to obtain a sample of fractional order dynamics in the same way that is done for integer order systems. This manuscript leverages occupation kernels to poise a trajectory as the fundamental unit of data from a fractional order dynamical system. When combined with a regularized regression problem, an approximation of fractional order dynamics is obtained as a linear combination of occupation kernels. A battery of numerical experiments are executed to validate the developed method, and it is demonstrated over two dynamical systems that accurate estimates of fractional order dynamics can be obtained both along trajectories and also nearby regions.

I. INTRODUCTION

Despite a preponderance of interest in fractional order dynamical systems that has blossomed over the past several decades (cf. [1]–[4]), there remains a gap in the identification of nonlinear fractional order systems. In the case of linear systems, much of classical parameter and system identification routines naturally translate to the fractional order case (cf. [5]–[7]), and this is largely due to the availability of cogent Laplace transformation for linear fractional order systems of Caputo and Riemann-Liouville type. To wit, assuming zero initial conditions, the Laplace transform of each of these operators (of the order q) results in multiplication by s^q in the Laplace domain [5].

However, the Laplace transform is a tool designed for the analysis of linear systems, and as such, the developed tools do not translate to nonlinear fractional order dynamical systems. Some recent work has begun to explore the identification and approximation of nonlinear systems, such as in [3]. In [3], a nonlocal dynamical system was approximated using a deep neural network through fPINNs.

The present manuscript aims to exploit the “fundamental theorem of fractional calculus,” and occupation kernels to resolve an approximation for the dynamics of a fractional order dynamical system from observed trajectories. Reproducing kernel Hilbert spaces (RKHSs) have demonstrated a long history of successful implementation in nonlinear estimation

problems (cf. [8]–[10]). In a RKHS over a set X , the reproducing kernel centered at a point $x \in X$ reproduces *local* point evaluation through the inner product of the Hilbert space. Occupation kernels reside in the same RKHSs, but instead of reproducing function evaluation, they reproduce the functional of evaluation against a given trajectory in the Hilbert space. Hence, occupation kernels may be viewed as nonlocal functionals within a Hilbert space.

This manuscript, leverages a combination of ridge regression and occupation kernels to provide an estimation of the dynamics for a nonlinear and nonlocal fractional order system. The treatment in the sequel focuses on the Caputo fractional derivative, but is applicable to any nonlocal dynamical system that is solvable through a Volterra integral.

In contrast to traditional kernel functions (cf. [11]–[14]), the regression problems contained in this manuscript center the trajectory as the fundamental unit of data (following [15]–[18]) rather than states at individual time points. This perspective has been demonstrated to be less sensitive to noise in the integer order setting, and in the present context, it allows for the fractional order dynamics to be sampled through integrals along trajectories within a domain of interest.

The manuscript is organized as follows. Section III reviews details concerning RKHSs that are relevant for the rest of the paper, Section IV gives a generalization of occupation kernels to the fractional order setting. In particular, the occupation kernels given in the present manuscript correspond to the Riemann-Liouville fractional integral rather than simply integration in [15], [16]. After a description of the problem statement in Section V, Section VI outlines the approach to the estimation problem using ridge regression and a representer theorem. Section VII then demonstrates the effectiveness of this approach, where the dynamics are compared to the approximation along both the sampled trajectories and the domain of interest.

II. FRACTIONAL ORDER DYNAMICAL SYSTEMS

In the context of this manuscript, a fractional order dynamical systems utilize fractional order differential operators to define initial value problems. For $0 < q$ and $m \in \mathbb{N}$ such that $m - 1 < q < m$, two principle operators are the Caputo fractional derivative and the Riemann-Liouville fractional derivative, given as $D_*^q = J^{m-q} \frac{d^m}{dt^m}$ and $D^q = \frac{d^m}{dt^m} J^{m-q}$ respectively, where $J^q x(T) = \frac{1}{\Gamma(q)} \int_0^T (T-t)^{q-1} x(t) dt$ is the Riemann-Liouville fractional integral [2]. Initial value problems using the Caputo fractional derivative may be defined using solely integer order derivatives, representing position, velocity, acceleration, etc. However, initial value

*This research was supported by the Air Force Office of Scientific Research (AFOSR) under contract numbers FA9550-20-1-0127, and the National Science Foundation (NSF) under award 2027976. 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.

¹Department of Mathematics, Changshu Institute of Technology, Suzhou, Jiangsu 215500, P R China xyli1112@sina.com.

²Department of Mathematics and Statistics, University of South Florida, Tampa, FL 33620. rosenfeldj@usf.edu

problems using the Riemann-Liouville derivative require fractional order initial conditions, which makes the Caputo derivative the modal choice for many engineering applications.

Let $f : \mathbb{R}^n \rightarrow \mathbb{R}^n$ be a Lipschitz continuous function. In this setting, the initial value problem $D_*^q x(t) = f(x(t))$ with $\frac{d^i}{dt^i} x(0) = x_{0,i} \in \mathbb{R}^n$ has a unique solution [2]. This initial value problem may also be expressed in Volterra form as

$$x(t) = \sum_{k=0}^{m-1} \frac{t^k}{k!} \frac{d^k}{dt^k} x(0) + \frac{1}{\Gamma(q)} \int_0^t (t-\tau)^{q-1} f(x(\tau)) d\tau. \quad (1)$$

The approach taken in this manuscript is to extract samples from observed trajectories, $\gamma(t)$, satisfying the initial value problem. In the simple case of $0 < q < 1$ this is expressed as $\gamma(t) - \gamma(0) = \frac{1}{\Gamma(q)} \int_0^t (t-\tau)^{q-1} f(\gamma(\tau)) d\tau$. By taking f to be in a RKHS and considering the fractional integral after composition with $\gamma(t)$ as a functional on that RKHS, a linear combination of occupation kernels can be leveraged to resolve a regularized regression problem, (4), and an approximation of f may be achieved.

III. REPRODUCING KERNEL HILBERT SPACES

A reproducing kernel Hilbert space (RKHS) over a set X is a Hilbert space consisting of functions from X to $\mathbb{R}$ such that for each $x \in X$ the evaluation functional $E_x f := f(x)$ is bounded. Hence, by the Riesz representation theorem, for each $x \in X$, there exists a function $k_x \in H$ such that $\langle f, k_x \rangle_H = f(x)$ for all $f \in H$. The function k_x is called the kernel function “centered” at x , and the function of two variables, $K : X \times X \rightarrow \mathbb{R}$, given as $K(x, y) = \langle k_y, k_x \rangle_H$ is called the kernel function corresponding to H .

Each kernel function is expressible as an inner product of feature maps, $\Psi : X \rightarrow \ell^2(\mathbb{N})$, as $K(x, y) = \langle \Psi(y), \Psi(x) \rangle_{\ell^2} = \sum_{m=1}^{\infty} \Psi_m(x) \Psi_m(y)$. Moreover, any function K that may be represented through features in this way is a kernel function for a RKHS. It should be noted that multiple feature maps can result in the same kernel function.

Two frequently employed kernel functions over $\mathbb{R}^n$ are tied to the exponential function. The exponential dot product kernel, given as $K(x, y) = \exp\left(\frac{1}{\mu} x^T y\right)$, is the real valued counterpart to the Bargmann-Fock space kernel, and the Gaussian radial basis function (RBF) is given as $K(x, y) = \exp\left(-\frac{1}{\mu} \|x - y\|_2^2\right)$. The Gaussian RBF is used extensively in scattered data interpolation and support vector machine classification [8], [9], [19]. The exponential dot product kernel on the other hand, is important for its close ties to quantum mechanics and operator theory [20]–[22].

Both kernels given above are *universal*, which means that given a compact set $Y \subset X$, $\epsilon > 0$, and a continuous function on Y , $\hat{f}$, there is a function $f \in H$ such that $\sup_{x \in Y} |f(x) - \hat{f}(x)| < \epsilon$ (cf. [19]). That is, the restriction of the functions in H are dense in the collection of continuous functions over Y with respect to the uniform norm.

IV. FRACTIONAL ORDER OCCUPATION KERNELS

In [15], occupation kernels were introduced to facilitate the integration of trajectory information into a single function within a RKHS. The resultant collection of functions were then turned to provide parameter estimates in a system identification routine. One of the benefits of this perspective is that it not only provided a signal noise robust representation of the data, but also cast trajectory information into an infinite dimensional feature space, which allowed for more information to be extracted from a single trajectory than otherwise available. The present method utilized (and generalizes) the occupation kernel method of [15] in a different way, where the occupation kernels are being utilized as basis functions for function approximation directly rather than just for data interpretation as was done in [15]. Though some of the definitions and evaluations of the fractional order occupation kernels (introduced below) overlap with occupation kernels in [15], their implementation here is distinct and largely self-contained.

Given a continuous signal $\theta : [0, T] \rightarrow \mathbb{R}^n$, $q > 0$, and a RKHS, H , over $\mathbb{R}^n$ consisting of continuously differentiable functions, the functional¹

$$f \mapsto C_q \int_0^T (T-t)^{q-1} f(\theta(t)) dt \quad (2)$$

is bounded. By the Riesz representation theorem, there exists a $\Gamma_{\theta, q} \in H$ that implements (2) via the Hilbert space inner product as $\langle f, \Gamma_{\theta, q} \rangle_H = C_q \int_0^T (T-t)^{q-1} f(\theta(t)) dt$.

Let $(\cdot)_j$ denote the j th coordinate of a vector in $\mathbb{R}^n$. If $\gamma : [0, T] \rightarrow \mathbb{R}^n$ is a trajectory satisfying the dynamical system $D_*^q(\gamma(t))_j = f_j(\gamma(t))$ for $f_j \in H$ for each $j = 1, \dots, n$, then

$$\begin{aligned} \langle f_j, \Gamma_{\gamma, q} \rangle_H &= C_q \int_0^T (T-t)^{q-1} f_j(\gamma(t)) dt \\ &= (\gamma(T))_j - (\gamma(0))_j. \end{aligned} \quad (3)$$

This relation follows from (1), and is similar to that found in [15], but in this case the dynamics are assumed to lie within a RKHS and not a symbol for a densely defined operator. This relation will be exploited in the development of a loss function in the sequel.

Occupation kernels have a particular advantage of being robust to sensor noise, as was demonstrated for the first order occupation kernels introduced in [15]. This property carries over to fractional order occupation kernels due to the integrability of the integration kernel $(T-t)^{q-1}$. Theorem 1 demonstrates that the impact of signal noise on the data embedding within the RKHS is relatively small. This stands in contrast to integer order methods for system and parameter identification that rely on numerical derivatives, which tend to be highly sensitive to noise and require considerable filtering (cf. [23]).

¹Here the notation C_q is used to denote the quantity $1/\Gamma(q)$, where Γ is the classical Gamma function. This avoids confusion with the occupation kernels in this manuscript.

Theorem 1. Let $\gamma : [0, T] \rightarrow \mathbb{R}^n$ be a continuous signal and let $\epsilon : [0, T] \rightarrow \mathbb{R}^n$ be measurable white noise with standard deviation $\sigma(\mu)$ according to the uniform metric on $[0, T]$. Let H be a RKHS of continuous functions over $\mathbb{R}^n$ with a continuously differentiable kernel function with bounded derivative. The pointwise impact of noise on the occupation kernel, which manifests as $|\Gamma_\gamma(x) - \Gamma_{\gamma+\epsilon}(x)|$ is $O(\sigma(\epsilon))$ when $q > 1/2$. More generally, this quantity is $O(J^q|\epsilon|)$ for $q > 0$.

Proof. Let $\mathcal{M} = \sup_{y \in \mathbb{R}^n} |\nabla_2 K(x, y)|$ be the bound on the derivative of K with respect to the second variable. Consider $|\Gamma_{\gamma,q}(x) - \Gamma_{\gamma+\epsilon,q}(x)| \leq C_q \int_0^T (T-t)^{q-1} |K(x, \gamma(t)) - K(x, \gamma(t) + \epsilon(t))| dt \leq C_q \cdot \mathcal{M} \int_0^T (T-t)^{q-1} |\epsilon(t)| dt \leq C_q \cdot \mathcal{M} \cdot \frac{T^{2q-\frac{1}{2}}}{2q-1} \cdot \sqrt{\frac{1}{T} \int_0^T |\epsilon(t)|^2 dt}$ when $q > 1/2$. For $q > 0$, the inequality is obtained in a similar manner. Hence, the theorem follows. $\square$

V. PROBLEM STATEMENT

Given a collection of observed continuously differentiable trajectories $\{\gamma_i : [0, T_i] \rightarrow \mathbb{R}^n\}_{i=1}^M$ that satisfy $D_*^q \gamma_i(t) = f(\gamma_i(t))$, the objective of this manuscript is to provide a means of determining an approximation of f .

To facilitate the analysis, each component of the function $f = (f_1, \dots, f_n)$ is assumed to reside in a RKHS, which allows for the framing of regression problems related to the dynamics as

$$\min_{f_j \in H} \sum_{i=1}^M \left(C_q \int_0^{T_i} (T_i - \tau)^{q-1} f_j(\gamma_i(\tau)) d\tau - (\gamma_i(T_i) - \gamma_i(0))_j \right)^2 + \lambda \|f_j\|_H^2, \quad (4)$$

for each $j = 1, \dots, n$ with $\lambda > 0$.

VI. OCCUPATION KERNELS AND REGRESSION PROBLEMS

For regularized regression problems of the form (4), Representer Theorems are commonly developed to give a closed form expression for a minimizer of (4). Typically, these representer theorems are used to directly show that a minimizer, f_j , is a sum of kernel functions. In the case where the functional involved in the regression problem is not function evaluation, f is represented as a sum of related functions from the Hilbert space. The Representer Theorem given below follows directly from standard proofs for Representer Theorems involving kernels, and it is presented here for completeness of exposition.

Proposition 1 (A Representer Theorem). *The solution to (4) is of the form $f_j(x) = \sum_{i=1}^M w_{i,j} \Gamma_{\gamma_i,q}(x)$.*

Proof. Let $f_j \in H$ and consider $W = \text{span}\{\Gamma_{\gamma_1,q}, \dots, \Gamma_{\gamma_M,q}\}$. Since W is a finite dimensional subspace of H , it is closed, and f is expressible as $f_j = f_W + f_P$ where $f_W \in W$ and $f_P \in W^\perp$. Note that, for each i , $C_q \int_0^{T_i} (T_i - t)^{q-1} f(\gamma_i(t)) dt = \langle f_j, \Gamma_{\gamma_i,q} \rangle_H =$

$\langle f_W, \Gamma_{\gamma_i,q} \rangle_H$. Hence, the replacement of f_j with f_W has no impact on the loss function. Moreover, by the Pythagorean theorem, $\|f_W\|_H^2 \leq \|f_j\|_H^2$ with equality only when $f_P = 0$.

Hence, the replacement of f_j with f_W decreases (4). Therefore, a minimizer of (4) must reside in W . $\square$

By virtue of Proposition 1, (4) may be adjusted to

$$\min_{w_j \in \mathbb{R}^M} \sum_{i=1}^M \left(\sum_{s=1}^M w_{s,j} \langle \Gamma_{\gamma_s,q}, \Gamma_{\gamma_i,q} \rangle_H - (\gamma_i(T) - \gamma_i(0))_j \right)^2 + \lambda w_j^T G w_j,$$

where $G = (\langle \Gamma_{\gamma_s,q}, \Gamma_{\gamma_i,q} \rangle_H)_{i,s=1}^M$ is the Gram matrix corresponding to the collection of occupation kernels. This Gram matrix may be computed by expressing the inner product $\langle \Gamma_{\gamma_s,q}, \Gamma_{\gamma_i,q} \rangle_H$ as

$$(C_q)^2 \int_0^{T_s} \int_0^{T_i} ((T_s - \tau)(T_i - t))^{q-1} K(\gamma_s(\tau), \gamma_i(t)) d\tau dt.$$

Writing $V_i := (\langle \Gamma_{\gamma_1,q}, \Gamma_{\gamma_i,q} \rangle_H, \dots, \langle \Gamma_{\gamma_M,q}, \Gamma_{\gamma_i,q} \rangle_H)^T \in \mathbb{R}^M$, and $c_j := ((\gamma_i(T) - \gamma_i(0))_j)_{i=1}^M$, w may be computed as

$$w_j = \left(\sum_{i=1}^M V_i V_i^T + \lambda G \right)^{-1} G c_j,$$

which may be obtained by taking a gradient to the loss function with respect to w_j .

VII. NUMERICAL EXPERIMENTS

Two systems were examined to validate the method of Section VI. For each system, the trajectories were generated through a kernel driven numerical method over the space $H^1 = W^{1,2}$ which is the Sobolev space of functions that have derivatives in L^2 . This numerical method transforms the system into an equivalent integral system as found in [24]. Subsequently, a piece-wise iterative reproducing kernel method was employed to solve the transformed integral system [25].

The occupation kernels and their associated Gram matrix, G , were computed using Gaussian numerical integration. For each system several experiments were performed to evaluate the effect of various parameters, such as the number of trajectories, the value of λ and μ , and the selection of kernel function for approximation.

The systems examined below were selected based on characteristics of the dynamics themselves. To align with the hypothesis of Theorem 1, System 1 was selected to have a bounded first derivative. System 2 is composed of transcendental functions that commonly appear in function theory.

System 1: Consider the following two dimensional fractional order dynamics,

$$D^{\frac{4}{5}} x = f(x) = \begin{pmatrix} f_1(x) \\ f_2(x) \end{pmatrix} = \begin{pmatrix} \frac{1}{1+x_2^2} \\ \frac{1}{1+x_1^2} \end{pmatrix},$$

where $x = (x_1, x_2)^T$.

The following experiments leverage 10 trajectories from this system, $\{\gamma_i : [0, 1] \rightarrow \mathbb{R}^2\}_{i=1}^{10}$. The initial points were utilized in the generation of the trajectories, $\{\gamma_i(0)\}_{i=1}^{10} = \{(0, 0)^\top, (0, 0.3)^\top, (0, 0.6)^\top, (0, 0.9)^\top, (0.3, 0.3)^\top, (0.3, 0.6)^\top, (0.3, 0.9)^\top, (0.6, 0.6)^\top, (0.6, 0.9)^\top, (0.9, 0.9)^\top\}$.

Experiment 1 This experiment examines the error between f and its approximation $\bar{f}$ along some of the given trajectories and in the region of System 1. By using the Gaussian RBF, $K(x, y) = e^{-\frac{1}{\mu} \|x-y\|_2^2}$, with kernel width $\mu = 2$ and parameter λ selected as 10^{-6} , the absolute errors between $f_i(x)$ and obtained approximations $\bar{f}_i(x)$ along several trajectories are shown in Figure 1 and Figure 2. The absolute errors between $f_i(x)$ and the obtained approximations $\bar{f}_i(x)$ on region $[0, 1] \times [0, 1]$ are shown in Figure 3.

Experiment 2 The second experiment explores the effect of the parameters λ and μ on the approximations. The effect of the kernel width μ for a fixed selection of $\lambda = 10^{-6}$ on the numerical results is shown in Table I. The impact of λ with fixed $\mu = 2.5$ is shown in Table II.

Experiment 3 The third experiment explores the impact of the number of occupation kernels utilized on the absolute error committed by the approximation. In this case, each trajectory $\gamma_i, i = 1, 2, \dots, 10$, is divided into two trajectories, one for $t \in [0, 1/2]$ and the other for $t \in [0, 1]$. Consequently, twenty occupation kernels are obtained:

$$\Gamma_{\gamma_i, [0, 1/2], q}(x) = C_q \int_0^{1/2} (1/2 - s)^{q-1} K(x, \gamma_i(s)) ds, \text{ and}$$

$$\Gamma_{\gamma_i, [0, 1], q}(x) = C_q \int_0^1 (1 - s)^{q-1} K(x, \gamma_i(s)) ds.$$

Using the Gaussian RBF $K(x, y) = e^{-\frac{1}{\mu} \|x-y\|_2^2}$ with kernel width $\mu = 2$ and selecting $\lambda = 10^{-6}$, the obtained absolute errors between $f_i(x)$ and the approximations $\bar{f}_i(x)$ for $i = 1, 2$ over region $[0, 1] \times [0, 1]$ are presented in Figure 4. For the purpose of comparison, Figure 3 presents the results of using ten occupation kernels.

Experiment 4 This experiment shortens the time horizon to $T = 0.2$, and examines the approximation power of this method in the smaller region of $[0, 1/2] \times [0, 1/2]$ using the same parameters with experiment 3.

The trajectories leveraged in this experiment correspond to the initial points $\{(0, 0), (0, 0.2), (0.2, 0), (0.2, 0.2)\}$ and the time horizons $[0, 0.1]$ and $[0, 0.2]$, resulting in eight trajectories and occupation kernels. The absolute errors between $f_i(x)$ and obtained approximations $\bar{f}_i(x)$ on region $[0, 0.5] \times [0, 0.5]$ is shown in Figure 5.

Experiment 5 This experiment introduces zero mean normally distributed white noise with standard deviation of σ to System 1 using the same parameters as in Experiment 3. The system identification method was implemented on the noise corrupted trajectories. The absolute errors between $f_i(x)$ and the obtained approximations $\bar{f}_i(x)$ on region $[0, 0.5] \times [0, 0.5]$ are shown in Table III.

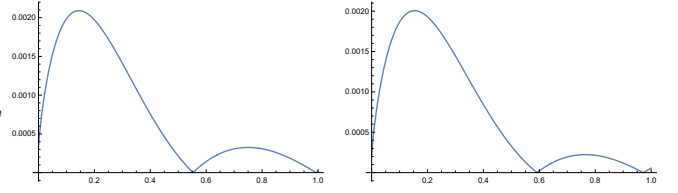


Fig. 1. Left: Absolute errors for the first dimension of $f(x)$ on trajectory γ_5 for System 1. Right: Absolute errors for the second dimension of $f(x)$ on trajectory γ_5 for System 1.

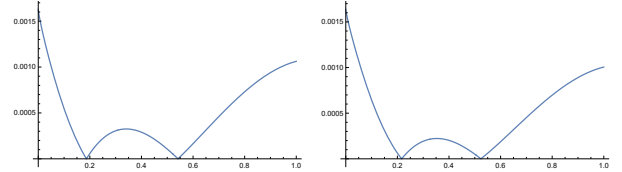


Fig. 2. Left: Absolute errors for the first dimension of $f(x)$ on trajectory γ_9 for System 1. Right: Absolute errors for the second dimension of $f(x)$ on trajectory γ_9 for System 1.

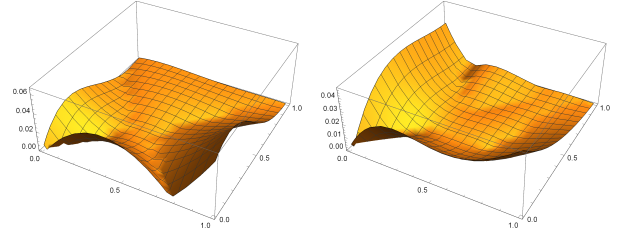


Fig. 3. Absolute errors between $f_i(x)$ and obtained approximations on $[0, 1] \times [0, 1]$ using 10 occupation kernels for System 1 (left: $|f_1 - \bar{f}_1|$, right: $|f_2 - \bar{f}_2|$).

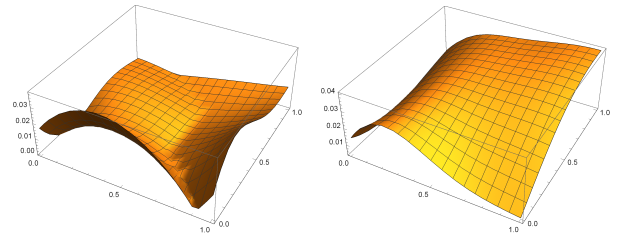


Fig. 4. Absolute errors between $f_i(x)$ and obtained approximations on $[0, 1] \times [0, 1]$ using 20 occupation kernels for System 1 (left: $|f_1 - \bar{f}_1|$, right: $|f_2 - \bar{f}_2|$).

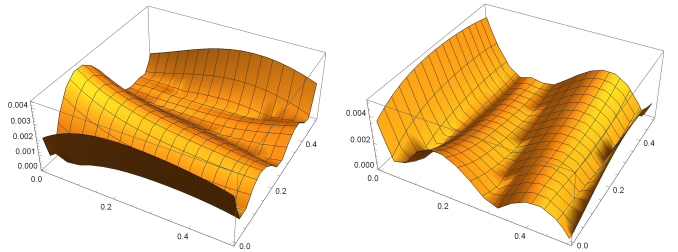


Fig. 5. Absolute errors between $f_i(x)$ and obtained approximations on $[0, 0.5] \times [0, 0.5]$ using 8 occupation kernels for System 1 (left: $|f_1 - \bar{f}_1|$, right: $|f_2 - \bar{f}_2|$).

TABLE I
THE OBTAINED ERROR FOR DIFFERENT VALUES OF μ WITH $\lambda = 0.000001$ ON γ_5 FOR SYSTEM 1

	$\mu = 1$	$\mu = 2$	$\mu = 3$	$\mu = 4$	$\mu = 5$	$\mu = 6$	$\mu = 7$	$\mu = 8$	$\mu = 9$
$\max_{t \in [0,1]} f_1(\gamma_5(t)) - \bar{f}_1(\gamma_5(t)) $	0.021	0.0021	0.0039	0.0054	0.0065	0.0070	0.0072	0.0074	0.0076
$\max_{t \in [0,1]} f_2(\gamma_5(t)) - \bar{f}_2(\gamma_5(t)) $	0.022	0.0020	0.0042	0.0060	0.0068	0.0075	0.0080	0.0081	0.0083

TABLE II
THE OBTAINED ERROR FOR DIFFERENT VALUES OF λ WITH $\mu = 2$ ON γ_5 FOR SYSTEM 1

	$\lambda = 0.1$	$\lambda = 0.01$	$\lambda = 0.001$	$\lambda = 0.00001$	$\lambda = 0.000001$	$\lambda = 0.0000001$
$\max_{t \in [0,1]} f_1(\gamma_5(t)) - \bar{f}_1(\gamma_5(t)) $	0.21	0.080	0.020	0.0052	0.0021	0.0021
$\max_{t \in [0,1]} f_2(\gamma_5(t)) - \bar{f}_2(\gamma_5(t)) $	0.18	0.074	0.020	0.0064	0.0020	0.0019

TABLE III
THE OBTAINED ERROR FOR DIFFERENT VALUES OF σ FOR SYSTEM 1

	No noise	$\sigma = 0.001$	$\sigma = 0.002$	$\sigma = 0.004$	$\sigma = 0.006$	$\sigma = 0.008$	$\sigma = 0.01$
$\max_{t \in [0,1]} f_1 - \bar{f}_1 $	0.00438	0.00553	0.00503	0.00735	0.00684	0.00798	0.00903
$\max_{t \in [0,1]} f_2 - \bar{f}_2 $	0.00508	0.00504	0.00673	0.00638	0.00828	0.01003	0.01522

System 2: The second system, which will be used in the subsequent experiments, is given as

$$D^{\frac{4}{5}}x = f(x) = \begin{pmatrix} f_1(x) \\ f_2(x) \end{pmatrix} = \begin{pmatrix} \sin x_2 + \cos x_1 \\ 0.6 \sin x_1 + 0.6 \end{pmatrix}.$$

Experiment 6 This experiment explores the error between f and the approximation, $\bar{f}$ over the region $[0, 1] \times [0, 1]$, and shows the effect of different parameters μ on System 2.

This experiment employed 16 trajectories, $\gamma_i, i = 1, 2, \dots, 16$, generated for this system over the time interval $[0, 1]$ with the initial points $\{(0, 0), (0, 0.3), (0, 0.6), (0, 0.9), (0.3, 0.3), (0.3, 0.6), (0.3, 0.9), (0.6, 0.6), (0.6, 0.9), (0.9, 0.9), (0.3, 0), (0.6, 0), (0.6, 0.3), (0.9, 0), (0.9, 0.3), (0.9, 0.6)\}$, the Gaussian RBF $K(x, y) = e^{-\frac{1}{\mu} \|x-y\|_2^2}$ with kernel width $\mu = 5$ and parameter $\lambda = 10^{-3}$, the absolute errors between $f_i(x)$ and obtained approximations $\bar{f}_i(x)$ on region $[0, 1] \times [0, 1]$ are shown in Figure 6. Table IV shows the impact of the addition of signal noise on the same experiment.

Experiment 7 This experiment explores the effect of selecting different kernel function for the approximation of System 2. Here the inverse multiquadrics (IMQ) RKF $K(x, y) = \frac{1}{\sqrt{\|x-y\|_2^2 + \mu^2}}$ is employed with $\mu = 0.6$. The obtained numerical results are given in Table V.

VIII. DISCUSSION

The purpose of this approach is to provide a method for system identification for fractional order dynamical systems. These systems do not admit estimations using local methods (such as those found in [23]), and thus, some form of non-local data interpretation must be utilized to estimate the dynamics. In a scenario where the fractional order of a system is unknown, the q parameter becomes a tuning parameter that can be used to better capture the dynamics of

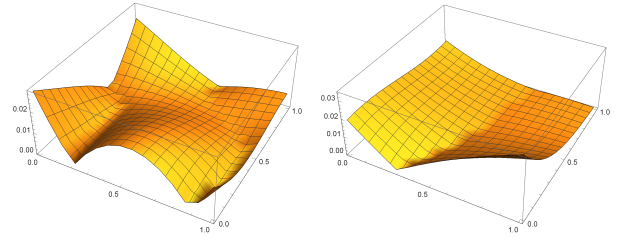


Fig. 6. Absolute errors between $f_i(x)$ and obtained approximations on $[0, 1] \times [0, 1]$ using sixteen trajectories for System 2 (left: $|f_1 - \bar{f}_1|$, right: $|f_2 - \bar{f}_2|$).

a system. It should be noted that even for the integer order case, with $q = 1$, this approach to system identification is still novel, and the inclusion of fractional order considerations broadens the scope of this method.

The experiments given in Section VII demonstrate the effectiveness of the method at obtaining an approximation of a fractional order dynamical system from observed trajectories. While approaches to system identification for integer order system frequently employ numerical differentiation locally to get samples of the dynamics (cf. [23]), local samples cannot be obtained for fractional order system in this manner. The approach taken in the present manuscript leverages integrals along trajectories as the fundamental unit of data, which follows the approach taken in [15]–[17].

In Section VII it was observed that using several small trajectories or segmented longer trajectories can be leveraged to improve the estimation of the dynamics for both System 1 and System 2. While occupation kernels are nonlocal, using short trajectories with initial points spread throughout a region can produce “local” estimate of fractional order dynamics. For longer trajectories, greater resolution may be obtained by segmenting a longer trajectory as $[0, 1/M], \dots, [0, M/M]$.

TABLE IV
THE OBTAINED ERROR FOR DIFFERENT VALUES OF σ FOR SYSTEM 2

	No noise	$\sigma = 0.001$	$\sigma = 0.002$	$\sigma = 0.004$	$\sigma = 0.006$	$\sigma = 0.008$	$\sigma = 0.01$
$\max_{t \in [0,1]} f_1 - \bar{f}_1 $	0.02913	0.03590	0.04043	0.04427	0.05176	0.05250	0.06709
$\max_{t \in [0,1]} f_2 - \bar{f}_2 $	0.01973	0.01919	0.02232	0.02198	0.02681	0.02582	0.02654

TABLE V
THE OBTAINED ERROR ALONG SEVERAL TRAJECTORIES USING IMQ RKF WITH $\mu = 2$ AND $\lambda = 0.00001$ FOR SYSTEM 2

	$\gamma_2(t)$	$\gamma_4(t)$	$\gamma_6(t)$	$\gamma_8(t)$	$\gamma_{10}(t)$
$\max_{t \in [0,1]} f_1(\gamma_i(t)) - \bar{f}_1(\gamma_i(t)) $	0.036	0.0252	0.030	0.032	0.020
$\max_{t \in [0,1]} f_2(\gamma_i(t)) - \bar{f}_2(\gamma_i(t)) $	0.032	0.0150	0.010	0.0086	0.0048

The effects of μ and λ are given in Tables I, II. The error between the function and its approximation deteriorates as λ gets larger. This dependence on λ is to be expected, as the regularization allows larger deviation of the approximation from the original function with larger λ . The value of μ that worked best in the experiments was $\mu = 2$, and this is similar to what is observed with standard reproducing kernels, where the best selection of μ is usually somewhere between the distances between sampled points and the diameter of the work space.

IX. CONCLUSION

The presented method for the estimation of nonlocal and nonlinear dynamics is a viable approach for nonlinear system identification of a nonlocal dynamical system. The particular systems explored in this manuscript are Caputo based dynamical systems, but the method can be easily adapted for a large number of other nonlocal operators. The use of occupation kernels allows for an individual trajectory to be treated as a single unit of data, and in this case, they provide nonlocal samples of fractional order dynamics. Leveraging the representer theorem, the regression problem 4 may be resolved through a linear combination of occupation kernels “centered” at the trajectories sampled from the system. This approach aligns with estimation methods using reproducing kernels (cf. [19]), and the effectiveness of the approach was demonstrated through a battery of numerical experiments presented in Section VII.

REFERENCES

- [1] J. A. Rosenfeld and W. E. Dixon, “Approximating the caputo fractional derivative through the mittag-leffler reproducing kernel hilbert space and the kernelized adams–bashforth–moulton method,” *SIAM Journal on Numerical Analysis*, vol. 55, no. 3, pp. 1201–1217, 2017.
- [2] K. Diethelm, *The analysis of fractional differential equations: An application-oriented exposition using differential operators of Caputo type*. Springer Science & Business Media, 2010.
- [3] G. Pang, L. Lu, and G. E. Karniadakis, “fpinns: Fractional physics-informed neural networks,” *SIAM Journal on Scientific Computing*, vol. 41, no. 4, pp. A2603–A2626, 2019.
- [4] R. L. Magin, *Fractional calculus in bioengineering*. Begell House Redding, 2006, vol. 2, no. 6.
- [5] S. Das and I. Pan, *Fractional Order System Identification*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, pp. 67–81.
- [6] M.-R. Rahmani and M. Farrokhi, “Nonlinear dynamic system identification using neuro-fractional-order hammerstein model,” *Transactions of the Institute of Measurement and Control*, vol. 40, no. 13, pp. 3872–3883, 2018.
- [7] L. Dorcák, E. A. Gonzalez, J. Terpak, J. Valsa, and L. Pivka, “Identification of fractional-order dynamical systems based on nonlinear function optimization,” *International Journal of Pure and Applied Mathematics*, vol. 89, no. 2, pp. 225–250, 2013.
- [8] H. Wendland, *Scattered data approximation*. Cambridge university press, 2004, vol. 17.
- [9] G. E. Fasshauer, *Meshfree approximation methods with MATLAB*. World Scientific, 2007, vol. 6.
- [10] 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*, vol. 50, no. 3, pp. 657–682, 2014.
- [11] —, “Kernel methods in system identification, machine learning and function estimation: A survey,” *Automatica*, vol. 50, no. 3, pp. 657–682, 2014.
- [12] F. P. Carli, T. Chen, and L. Ljung, “Maximum entropy kernels for system identification,” *IEEE Transactions on Automatic Control*, vol. 62, no. 3, pp. 1471–1477, 2016.
- [13] G. Birpoutsoukis, A. Marconato, J. Lataire, and J. Schoukens, “Regularized nonparametric Volterra kernel estimation,” *Automatica*, vol. 82, pp. 324–327, 2017.
- [14] J. Schoukens and L. Ljung, “Nonlinear system identification: A user-oriented road map,” *IEEE Control Syst. Mag.*, vol. 39, no. 6, pp. 28–99, 2019.
- [15] J. A. Rosenfeld, B. Russo, R. Kamalapurkar, and T. T. Johnson, “The occupation kernel method for nonlinear system identification,” 2019.
- [16] J. A. Rosenfeld, R. Kamalapurkar, L. F. Gruss, and T. T. Johnson, “Dynamic mode decomposition for continuous time systems with the Liouville operator,” *Under Review*.
- [17] J. A. Rosenfeld, R. Kamalapurkar, B. Russo, and T. T. Johnson, “Occupation kernels and densely defined liouville operators for system identification,” in *58th IEEE Conference on Decision and Control*. IEEE, 2013 (To Appear).
- [18] J. A. Rosenfeld and R. Kamalapurkar, “Dynamic mode decomposition with control liouville operators,” in *Proceedings of the 24th International Symposium on Mathematical Theory of Networks and Systems (MTNS 2021)*, To Appear.
- [19] I. Steinwart and A. Christmann, *Support vector machines*. Springer Science & Business Media, 2008.
- [20] B. C. Hall, *Quantum theory for mathematicians*. Springer, 2013.
- [21] K. Zhu, *Analysis on Fock spaces*. Springer Science & Business Media, 2012, vol. 263.
- [22] J. A. Rosenfeld, B. Russo, and W. E. Dixon, “The mittag leffler reproducing kernel hilbert spaces of entire and analytic functions,” *Journal of Mathematical Analysis and Applications*, vol. 463, no. 2, pp. 576–592, 2018.
- [23] 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.
- [24] K. Diethelm, *The Analysis of Fractional Differential Equations*. springer, 2004.
- [25] G. Fazhan and Q. Suping, “Piecewise reproducing kernel method for singularly perturbed delay initial value problems,” *Applied Mathematics Letters*, vol. 37, pp. 67–71, 2012.