SIAM J. NUMER. ANAL.
Vol. 55, No. 3, pp. 1201–1217

APPROXIMATING THE CAPUTO FRACTIONAL DERIVATIVE THROUGH THE MITTAG-LEFFLER REPRODUCING KERNEL HILBERT SPACE AND THE KERNELIZED ADAMS–BASHFORTH–MOULTON METHOD*

JOEL A. ROSENFELD† AND WARREN E. DIXON†

Abstract. This paper introduces techniques for the estimation of solutions to fractional order differential equations (FODEs) and the approximation of a function's Caputo fractional derivative. These techniques are based on scattered data interpolation via reproducing kernel Hilbert spaces (RKHSs). Specifically, an RKHS is generated for the purpose of estimating fractional derivatives from the Mittag-Leffler function. The RKHS, called the Mittag-Leffler RKHS, as well as others are utilized to estimate Caputo fractional derivatives and to introduce a modified Adams–Bashforth–Moulton method for the estimation of the solution to FODEs.

Key words. Mittag-Leffler function, reproducing kernel Hilbert spaces, fractional order differential equations, Adams–Bashforth–Moulton method, scattered data interpolation

AMS subject classifications. 46C02, 65D25, 33F02

DOI. 10.1137/16M1056894

1. Introduction.

In recent decades, fractional order calculus has seen a wide array of applications. Fractional order calculus has been employed to model viscoelastic materials [5, 6, 7] and diffusion processes [21]. Moreover, it has found many applications in control theory, such as analogues of PID controllers [33], fractional order optimal control [2, 3], and other applications [23]. Of the variety of fractional order derivatives available for scientific applications, the Caputo fractional derivative is the most widely used. This is because fractional order differential equations (FODEs) arising from the Caputo fractional derivative require initial conditions of integer order, which can be related to physical quantities. This differs from FODEs arising from the Riemann–Liouville fractional derivative, where the initial conditions rely on fractional derivatives of the state.

Given $T \in \mathbb{R}_+$, for a sufficiently regular function, $f : [0, T] \to \mathbb{C}$, and $m \in \mathbb{N}$, the Caputo fractional derivative of order $q \in (m-1, m)$ at $s \in [0, T]$ is given by

$$D_*^q f(s) := \frac{1}{\Gamma(m-q)} \int_0^s (s-t)^{m-q-1} f^{(m)}(t) dt \tag{1}$$

and arises from the application of the Riemann–Liouville fractional integration operator of order $m-q$, J_0^{m-q}, to the mth order derivative of f. Further details concerning the definition can be found in [8].

The challenge of approximating the Caputo fractional derivative of a function is the nonlocal property of the fractional derivative. Unlike integer order derivatives, the

*Received by the editors January 15, 2016; accepted for publication (in revised form) January 19, 2017; published electronically May 23, 2017.
http://www.siam.org/journals/sinum/55-3/M105689.html
Funding: The work of the authors was supported in part by NSF award 1509516 and Office of Naval Research grant N00014-13-1-0151. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the sponsoring agencies.
†Nonlinear Controls and Robotics (NCR) Laboratory, Department of Mechanical and Aerospace Engineering, University of Florida, Gainesville, FL 32611 (joelar@ufl.edu, wdixon@ufl.edu).

Caputo fractional derivative cannot be calculated using only local information about a point [8, 18]. The integral in (1) necessitates the utilization of the entire history of a function to compute the fractional derivative.

Various approaches to approximating Caputo fractional derivatives and to approximating solutions to FODEs arising from the Caputo fractional derivative have been developed. Previous efforts toward the approximation of the Caputo fractional derivative have investigated approximations through Grünwald–Letnikov methods [8] as well as approximations by Chebychev polynomials [29] and collocation with Bessel functions [30]. In [11], the approximation of the Caputo fractional derivative was performed by using methods of Gaussian quadrature. In [16], a method was introduced to approximate the fractional order derivative of a signal with noise via Jacobi orthogonal polynomials. Other significant works toward the approximations of the Caputo fractional derivative and related FODEs can be found in [8, 9, 10].

To the authors' knowledge, the theory surrounding reproducing kernel Hilbert spaces (RKHSs), a common tool in approximation theory [22, 28], has not been utilized to estimate a function's fractional derivative, nor has it been employed to estimate a numerical solution of a FODE. The focus of this paper is the introduction of new RKHSs and algorithms for the estimate of Caputo fractional derivatives and solutions of FODEs related to the Caputo fractional derivative.

To this end, this paper introduces the Mittag-Leffler RKHS (of order q) for the purpose of approximating a function's Caputo fractional derivative. In particular, the Mittag-Leffler RKHS leverages the fact that the Mittag-Leffler function is an eigenfunction for the Caputo fractional derivative (cf. [8, 12]) to simplify the estimation of a function's Caputo fractional derivative.

The Mittag-Leffler RKHS is a universal RKHS (see the discussion in section 3) with the kernel function given by $K_q(\lambda, t) := E_q(\lambda^q t^q)$, where $q > 0$ and $E_q(t) := \sum_{n=0}^{\infty} \Gamma(nq+1)^{-1} t^n$ is the Mittag-Leffler function [8, 12]. The Mittag-Leffler kernel function satisfies an eigenvalue equation for the Caputo fractional derivative, $D_*^q E_q(\lambda^q t^q) = \lambda^q E_q(\lambda^q t^q)$, where $\lambda \in \mathbb{R}_+$ [12]. While the Mittag-Leffler function was introduced by Gosta Mittag-Leffler in a series of papers that appeared in the beginning of the 20th century, it has been an often neglected special function until the past couple of decades [12]. The recent interest in this function is due to the growing number of applications for which fractional calculus can be applied to yield new insights [15, 17, 19]. Moreover, several investigations have been performed toward the computation of the Mittag-Leffler function itself [13, 20, 24, 26]. For this reason, the introduction of a kernel function based on the Mittag-Leffler function is timely.

Through the method of scattered data interpolation, a universal RKHS can be used to approximate a given continuous function uniformly over a compact subset of the input space. Once strict definiteness is established for the Gram matrix in (2), any finite set of points can be interpolated by a linear combination of functions of the form $K(x, x_i)$. In addition, if the function to be approximated is contained in the Hilbert space itself, the method of scattered data interpolation will converge to the sampled function uniformly as more samples are interpolated. Theorem 2 in section 3 demonstrates that approximation of a function in the norm of the Mittag-Leffler RKHS leads to uniform approximation of its Caputo fractional derivative over a compact set.

Section 2 demonstrates the approximation properties of the Mittag-Leffler RKHS by approximating the Caputo fractional derivative of two generic functions. This is justified by the universality of the Mittag-Leffler RKHS. A universal RKHS is dense in the space of real valued continuous functions over a compact set. Thus, it is

often assumed that a sampled continuous function is in fact contained within the Hilbert space. The assumption that a sampled function is contained in an RKHS is reasonable, since up to sample noise a function in the RKHS is contained within the envelope of the sampled function determined by the noise threshold. Thus, universal RKHSs are often used for function approximation, and in particular, Gaussian radial basis functions (RBFs) correspond to a universal RKHS that is the modal choice for function approximation in the statistical learning community [14, 28].

Related to the problem of approximating a function's Caputo fractional derivative is the problem of estimating the solution of a FODE governed by the Caputo fractional derivative, $D_*^q y(t) = f(t, y(t))$. Diethelm generalized the Adams–Bashforth–Moulton (ABM) method to the fractional order case in [8, 10]. The technique uses a piecewise linear interpolation of estimated points $\{(t_i, f(t_i, y_i))\}_{i=1}^k$ to approximate $y(t_{k+1})$. Since accurate approximations of a function can be achieved through interpolation via RKHSs, a natural adjustment of the ABM algorithm is to replace piecewise linear interpolation by interpolation with reproducing kernels, a so-called kernelized ABM method presented in section 4. The framework is presented for general kernel functions, not just the Mittag-Leffler kernel, and as such, numerical experiments performed in section 5.2 use two different kernel functions for computing numerical solutions of FODEs. Moreover, when $q \in \mathbb{N}$, the kernelized ABM method also applies to ordinary differential equations.

Section 2 of this paper discusses approximations in RKHSs via scattered data interpolation. Section 3 introduces the Mittag-Leffler RKHS and implications of approximating with linear combinations of the Mittag-Leffler kernel. Section 5.1 presents several examples of the approximation method introduced, and in particular the approximation of the Caputo fractional derivative of several well-known functions will be examined.

2. Approximation with RKHSs. An RKHS over a set X is a Hilbert space of functions, $f : X \to \mathbb{C}$ (or $f : X \to \mathbb{R}$), for which given each $x \in X$ the evaluation functional $E_x f = f(x)$ is bounded [1, 28]. Consequently the Riesz theorem guarantees that for each $x \in X$ there exists a function $k_x \in H$ for which $\langle f, k_x \rangle_H = f(x)$ for all $f \in H$ [1].

Associated with each RKHS is an associated kernel function given by $K(x, y) = \langle k_y, k_x \rangle_H$. For real valued kernel functions with domain $\mathbb{R}^n$, it was demonstrated in [22] that if a kernel function is of the form $K(x, y) = g(x^T y)$, where $g(t) := \sum_{m=0}^{\infty} a_m t^m$ converges for all t with $a_m > 0$ for all m, then the Gram matrix, $(K(x_i, x_j))_{i,j=1}^M$, is strictly positive definite. Moreover, a kernel function of the form $K(x, y) = g(x^T y)$ is universal [28], which means that given any compact subset $D \subset \mathbb{R}^n$, the associated Hilbert space is dense inside of the space of real valued continuous functions over D (with respect to the supremum norm).

RKHSs are ideal for function approximation, because interpolation in an RKHS, H, leads to uniform convergence of the estimates, whereas interpolation with polynomials can result in a divergent sequence of functions [31]. Given a collection of samples, $\{(x_i, f(x_i))\}_{i=1}^m$, of a function $f \in H$, computing the solution to the interpolation matrix equation with a real kernel function $K : X \times X \to \mathbb{R}$,

$$(K(x_i, x_j))_{i,j=1}^m \mathbf{w} = \mathbf{f}, \tag{2}$$

is equivalent to computing the projection of f onto the span of $\{K(\cdot, x_i)\}_{i=1}^m$, where $\mathbf{w} \in \mathbb{R}^m$ and $\mathbf{f} = (f(x_1), \ldots, f(x_m))^T \in \mathbb{R}^m$. If the sequence $\{x_i\}_{i=1}^\infty \subset X$ is selected appropriately, then $\{K(\cdot, x_i)\}_{i=1}^\infty$ can be a basis for an RKHS. Thus the interpolation

of a function leads to a sequence of functions that converge to f in the Hilbert space norm. For an RKHS associated with real entire functions, such as $ML^2(\mathbb{R}_+; q)$ (introduced in the next section), a sequence of distinct points $\{x_i\}_{i=1}^{\infty} \subset \mathbb{R}$ that has an accumulation point is sufficient for establishing $\{K(\cdot, x_i)\}_{i=1}^{\infty}$ as a basis.

Interpolating a function via arbitrary samples is called scattered data interpolation, and the determination of a unique solution to (2) relies upon the ability to establish the invertibility of the matrix $(K(x_i, x_j))_{i,j=1}^{m}$. The matrix is guaranteed to be positive definite for any kernel; however, for some kernels, it may have a nontrivial nullspace. For a kernel function that is *strictly* positive definite the matrix in (2) is guaranteed to be invertible. The following theorem gives a sufficient condition to demonstrate that a kernel function is strictly positive definite.

THEOREM 2.1 (see [22]). *Let $g : \mathbb{R} \to \mathbb{R}$ given by $g(t) = \sum_{m=1}^{\infty} a_m t^m$ be a real entire function for which $a_m \geq 0$ for all m, and define a real valued kernel function $K : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ as $K(x, y) = g(x^T y)$. Consider the set $A = \{m : a_m > 0\}$; then K is strictly positive definite iff A contains zero as well as an infinite number of even integers and an infinite number of odd integers.*

If a kernel function K is known to be strictly positive definite, then the following proposition can be established.

PROPOSITION 2.2. *If $K : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ is strictly positive definite, then so is $\tilde{K} : X \times X \to \mathbb{R}$ given by $\tilde{K}(x, y) := K(\phi(x), \phi(y))$ for any injective function $\phi : X \to \mathbb{R}$.*

Proof. Since ϕ is injective, for any sequence of distinct points $x_1, \ldots, x_M \in X$, $\phi(x_1), \ldots, \phi(x_M) \in \mathbb{R}$ are also distinct. Relabeling $y_i = \phi(x_i)$ for $i = 1, \ldots, M$ it can be seen that $(K(y_i, y_j))_{i,j=1}^{M}$ is strictly positive definite by the hypothesis. Thus the kernel function $\tilde{K}(x, y) = K(\phi(x), \phi(y))$ is strictly positive definite. ∎

As a consequence of approximating a function in an RKHS, the Hilbert space norm dominates the supremum norm over compact sets. Specifically, the following proposition holds.

PROPOSITION 2.3 (see [28]). *Let X be a metric space and $Y \subset X$ a compact subset. Further, suppose that H is an RKHS over X with a continuous kernel function, $K : X \times X \to \mathbb{R}$, and $f \in H$; then there exists a $C > 0$ such that $\sup_{x \in Y} |f(x)| \leq C \|f\|_H$.*

A similar bound as in Proposition 2.3 can be established for a function's derivative, if it is in the RKHS.

THEOREM 2.4 (see [28]). *Let $X \subset \mathbb{R}$ be an open subset, $m \geq 0$, and $K : X \times X \to \mathbb{C}$ be an m-times continuously differentiable kernel on X with RKHS H. Then every $f \in H$ is m-times continuously differentiable, and for $\alpha \in \mathbb{N}$ with $\alpha \leq m$ and $x \in X$ we have*

$$\left| \frac{d^{\alpha}}{dx^{\alpha}} f(x) \right| \leq \|f\|_H \left(\frac{d^{2\alpha}}{dz^{\alpha} dy^{\alpha}} K(z, y) \Big|_{(z,y)=(x,x)} \right)^{1/2}.$$

Theorem 2.5 generalizes Theorem 2.4 to Caputo fractional derivatives. The approximation of a function in a RKHS thus leads to the uniform approximation of its Caputo fractional derivative over compact sets. This allows the use of generic RKHSs for the purposes of estimating Caputo fractional order derivatives.

THEOREM 2.5. *Let $K : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ be a kernel function that is m-times continuously differentiable in each variable, $q \in (m-1, m)$, and H be the RKHS associated*

with K. Norm convergence in the RKHS implies uniform convergence of the Caputo fractional derivative of order q over compact subsets of $\mathbb{R}$.

Proof. Let $f, g \in H$, and set I as a compact subset of $\mathbb{R}$. Since K is m-times continuously differentiable, f and g are m-times continuously differentiable by Theorem 2.4. Now consider for $s \in I$,

$$\begin{aligned}|D_*^q f(s) - D_*^q g(s)| &= \left| \frac{1}{\Gamma(m-q)} \left(\int_0^s \frac{f^{(m)}(\tau)}{(s-\tau)^{m-1-q}} d\tau - \int_0^s \frac{g^{(m)}(\tau)}{(s-\tau)^{m-1-q}} d\tau \right) \right| \\ &\leq \frac{1}{\Gamma(m-q)} \sup_{0\leq\tau\leq s} |f^{(m)}(\tau) - g^{(m)}(\tau)| \left| \int_0^s \frac{1}{(s-\tau)^{m-1-q}} d\tau \right| \\ &\leq \frac{C_s \cdot s^{m-q}}{\Gamma(m-q)} \|f-g\|_H.\end{aligned}$$

The second inequality follows when Theorem 2.4 is applied to $\sup_{0\leq\tau\leq s} |f^{(m)}(\tau) - g^{(m)}(\tau)|$. Specifically,

$$C_s := \sup_{\tau\in[0,s]} \left(\frac{d^{2\alpha}}{dz^\alpha dy^\alpha} K(z,y) \Big|_{(z,y)=(\tau,\tau)} \right)^{1/2},$$

and the finiteness of C_s follows from the continuous differentiability assumption on $K(z,y)$. Moreover, C_s increases with respect to s, and it follows that a uniform bound over $[0,s]$ can be achieved when C_s is replaced by a suitable upper bound for the compact set I. ∎

Note that the kernel functions in Theorem 2.5 are assumed to be continuously differentiable. In section 3, the Mittag-Leffler function is used to generate a kernel function that is continuously differentiable everywhere but the origin; therefore a different method is used to demonstrate the result of Theorem 2.5 for the Mittag-Leffler RKHS.

3. The Mittag-Leffler RKHS. In this section, the Mittag-Leffler kernel function and its corresponding RKHS is introduced. Given $q > 0$, the Mittag-Leffler kernel function of order q (or simply the Mittag-Leffler kernel function when q is understood) is the function, $K_q : \mathbb{R}_+ \times \mathbb{R}_+ \to \mathbb{R}$, given by

$$K_q(t,\lambda) = E_q(\lambda^q t^q). \tag{3}$$

For convenience $\mathbb{R}_+$ will be taken as $[0,\infty)$ throughout this paper. By Theorem 2.1 and Proposition 2.2, the Mittag-Leffler kernel function is a strictly positive definite kernel function. Each positive definite kernel function corresponds uniquely with a RKHS. Moreover, by the Aronszajn–Moore theorem (cf. [4]), the RKHS is given by

$$ML^2(\mathbb{R}_+;q) := \left\{ f(t) = \sum_{n=0}^\infty a_n t^{qn} \,\middle|\, \sum_{n=0}^\infty \Gamma(qn+1)|a_n|^2 < \infty \right\},$$

with the Hilbert space norm given by

$$\|f\|_{ML^2(\mathbb{R}_+;q)} = \left(\sum_{n=0}^\infty \Gamma(qn+1)|a_n|^2 \right)^{1/2}. \tag{4}$$

The RKHS $ML^2(\mathbb{R}_+;q)$ will be called the Mittag-Leffler RKHS of order q and is a space of functions infinitely differentiable on $(0,\infty)$. In particular, $ML^2(\mathbb{R}_+;q)$ contains the functions t^{qn} for $n=0,1,2,\ldots$, which form an orthogonal system for which $\|t^{qn}\|_{ML^2(\mathbb{R}_+;q)}=\sqrt{\Gamma(qn+1)}$ (i.e., t^{qn} corresponds to $a_j=0$ unless $j=n$ for which $a_n=1$). Since $ML^2(\mathbb{R}_+;q)$ contains all linear combinations of t^{qn} (for $n=0,1,2,\ldots$), an application of the Muntz–Szasz theorem (cf. Theorem 15.26 in [25]) demonstrates that $ML^2(\mathbb{R}_+;q)$ is universal over $\mathbb{R}^+$. Specific RKHS properties of $ML^2(\mathbb{R}_+;q)$ are provided in the following proposition for convenience.

PROPOSITION 3.1. *The $ML^2(\mathbb{R}_+;q)$ is a real valued RKHS. In particular, letting $\lambda\geq 0$, $f(t)=\sum_{n=0}^{\infty}a_nt^{qn}$, and $g(t)=\sum_{n=0}^{\infty}b_nt^{qn}$ be functions in $ML^2(\mathbb{R}_+;q)$, then $\langle f,g\rangle_{ML^2(\mathbb{R}_+;q)}=\sum_{n=0}^{\infty}\Gamma(qn+1)a_nb_n$, $\langle f(\cdot),K_q(\cdot,\lambda)\rangle_{ML^2(\mathbb{R}_+;q)}=f(\lambda)$, and $K_q(\cdot,\lambda)\in ML^2(\mathbb{R}_+;q)$. Moreover, $\{t^{qn}\}_{n=0}^{\infty}$ is a complete system of orthogonal functions contained in $ML^2(\mathbb{R}_+;q)$.*

Proof. The quantity $\langle f,g\rangle_{ML^2(\mathbb{R}_+;q)}=\sum_{n=0}^{\infty}\Gamma(qn+1)a_nb_n$ can be shown to be finite via the Cauchy–Schwarz inequality,

$$\begin{aligned}|\langle f,g\rangle_{ML^2(\mathbb{R}_+;q)}| &= \left|\sum_{n=0}^{\infty}\Gamma(qn+1)a_nb_n\right| \\ &\leq \left(\sum_{n=0}^{\infty}\Gamma(qn+1)|a_n|^2\right)^{1/2}\cdot\left(\sum_{n=0}^{\infty}\Gamma(qn+1)|b_n|^2\right)^{1/2} \\ &= \|f\|_{ML^2(\mathbb{R}_+;q)}\|g\|_{ML^2(\mathbb{R}_+;q)}.\end{aligned}$$

Thus, $\langle\cdot,\cdot\rangle_{ML^2(\mathbb{R}_+;q)}$ is a valid inner product for $ML^2(\mathbb{R}_+;q)$ and corresponds with (4) through $\langle f,f\rangle_{ML^2(\mathbb{R}_+;q)}=\|f\|^2_{ML^2(\mathbb{R}_+;q)}$ making $ML^2(\mathbb{R}_+;q)$ a Hilbert space.

For $\lambda\geq 0$, the function $K_q(t,\lambda)=\sum_{n=0}^{\infty}\frac{t^{qn}\lambda^{qn}}{\Gamma(qn+1)}$ has the $ML^2(\mathbb{R}_+;q)$ norm

$$\|K_q(\cdot,\lambda)\|^2_{ML^2(\mathbb{R}_+;q)}=\sum_{n=0}^{\infty}\frac{|\lambda|^{2qn}}{\Gamma(qn+1)},$$

which converges, since asymptotically $\Gamma(qn+1)\sim\sqrt{2\pi(qn+1)}\left(\frac{qn+1}{e}\right)^{qn+1}$ by Stirling's formula. Hence, $K_q(\cdot,\lambda)\in ML^2(\mathbb{R}_+;q)$ for all $\lambda\geq 0$. To verify the reproducing property consider

$$\langle f,K_q(\cdot,\lambda)\rangle_{ML^2(\mathbb{R}_+;q)}=\sum_{n=0}^{\infty}\Gamma(qn+1)a_n\frac{\lambda^{qn}}{\Gamma(qn+1)}=\sum_{n=0}^{\infty}a_n\lambda^{qn}=f(\lambda).$$

Therefore it is established that $ML^2(\mathbb{R}_+;q)$ is an RKHS over $\mathbb{R}_+$.

For the system $\{t^{qn}\}_{n=0}^{\infty}$, note that $\langle f,t^{qn}\rangle_{ML^2(\mathbb{R}^n;q)}=\Gamma(qn+1)a_n$. Therefore,

$$\langle t^{qm},t^{qn}\rangle_{ML^2(\mathbb{R}^n;q)}=\begin{cases}0 & \text{if } m\neq n,\\ \Gamma(qn+1) & \text{if m=n.}\end{cases}$$

The system can be seen to be complete, since $f\perp\{t^{qn}\}_{n=0}^{\infty}$ iff $a_n=0$ for all n, which means $f\equiv 0$. ∎

Using the methods presented in section 2, a continuous function can be approximated through the interpolation of its samples by solving (2). Moreover, by Corollary 4.36 in [28], given a compact set, $I\subset(0,\infty)$, the derivative of a function $f\in ML^2(\mathbb{R}_+;q)$ can be uniformly approximated over I. More generally, the following theorem can be established.

THEOREM 3.2. *If $f \in ML^2(\mathbb{R}_+;q)$, $I \subset \mathbb{R}_+$ is compact, $q > 0$, and $\{t_i\}_{i=1}^{\infty} \subset \mathbb{R}_+$ is a sequence of distinct points that contains an accumulation point, then the sequence of functions f_m obtained by interpolating $\{(t_i, f(t_i))\}_{i=1}^{m}$ via (2), with the Mittag-Leffler kernel function of order q, converges uniformly to f over I, and $D_*^q f_m$ converges uniformly to $D_*^q f$ over I. Moreover, for each $h \in ML^2(\mathbb{R}_+;q)$ its Caputo fractional derivative, $D_*^q h$, is in an RKHS, denoted $D_*^q ML^2(\mathbb{R}_+;q)$, and $\|D_*^q h\|_{D_*^q ML^2(\mathbb{R}_+;q)} \le \|h\|_{ML^2(\mathbb{R}_+;q)}$.*

Proof. Since each $h \in ML^2(\mathbb{R}_+;q)$ can be represented by $h(t) = g(t^q)$, where g is a real entire function, the functions in $ML^2(\mathbb{R}_+;q)$ are uniquely determined by their values on a sequence with an accumulation point. In particular, the Hilbert space norm $\|f_m - f\|_{ML^2(\mathbb{R}_+;q)}$ converges to zero by the discussion in section 2. Thus, by Proposition 2.3, the sequence of functions converge uniformly to f over the compact set I.

Through the application of the operator D_*^q term by term to t^{qn} for each $n = 0, 1, 2, \ldots$, each function $h(t) = \sum_{n=0}^{\infty} a_n t^{qn} \in ML^2(\mathbb{R}_+;q)$ is sent to

$$D_*^q h(t) = \sum_{n=0}^{\infty} a_{n+1} \frac{\Gamma((n+1)q+1)}{\Gamma(nq+1)} t^{qn},$$

by Theorem 3.10 in [8], which lies in the Hilbert space $D_*^q ML^2(\mathbb{R}_+;q)$ containing the orthonormal basis

$$\left\{ e_n(t) = \frac{\sqrt{\Gamma((n+1)q+1)}}{\Gamma(qn+1)} t^{qn} \right\}_{n=0}^{\infty}.$$

Again by the Aronszajn–Moore theorem the kernel function for $D_*^q ML^2(\mathbb{R}_+;q)$ is given by

$$\tilde{K}_q(\lambda, t) = \sum_{n=0}^{\infty} e_n(\lambda) e_n(t) = \sum_{n=0}^{\infty} \frac{\Gamma(q(n+1)+1)}{\Gamma(qn+1)^2} \lambda^{qn} t^{qn}.$$

When viewed as an operator from $ML^2(\mathbb{R}_+;q)$ to $D_*^q ML^2(\mathbb{R}_+;q)$, the operator D_*^q is norm decreasing. Specifically,

$$\|D_*^q h\|_{D_*^q ML^2(\mathbb{R}_+;q)} = \|h(\cdot) - a_0\|_{ML^2(\mathbb{R}_+;q)} \le \|h\|_{ML^2(\mathbb{R}_+;q)}$$

for all $h \in ML^2(\mathbb{R}_+;q)$. It should be noted that $\tilde{K}_q$ is continuous in both arguments.

Thus, by Proposition 2.3,

$$|D_*^q(f_n - f)(s)| \le C \|D_*^q(f_n - f)\|_{D_*^q ML^2(\mathbb{R}_+;q)} \le C \|f_n - f\|_{ML^2(\mathbb{R}_+;q)},$$

where C only depends on I. Thus, the convergence of $D_*^q f_n \to D_*^q f$ is uniform over I. ∎

In Theorem 3.2 the functions $f_m(t)$ are of the form $f_m(t) = \sum_{i=1}^{m} w_i K_q(t, t_i)$. The advantage of approximating f with functions of this form is that the Caputo fractional derivative can be computed for f_m as

$$D_*^q f_m(s) = \sum_{i=1}^{m} w_i t_i^q K_q(t, t_i). \tag{5}$$

The approximation of functions by scattered data interpolation with the Mittag-Leffler kernel function is explored in the next section with numerical results.

Theorem 3.2 guarantees uniform convergence of the Caputo fractional derivative over a compact set of $\mathbb{R}_+$. However, since the function $F(t) = K_q(\lambda, t)$ has a vertical slope at the origin, the approximation has a high sensitivity to error at $t = 0$. This sensitivity is observed in the numerical experiments of section 5.1.

4. A kernelized fractional Adams–Bashforth–Moulton method. As an example application of the approximation of functions by linear combinations of kernel functions for the purposes of fractional calculus, the present section examines a kernelized ABM method for approximating solutions to FODEs governed by the Caputo fractional derivative. In [10, 8], the classical ABM method was modified for applications to FODEs, and the methods were used to examine viscoplastic materials in [10]. Given $q \in (m-1, m)$, $m \in \mathbb{N}$, and a fractional order initial value problem,

$$D_*^q y(s) = f(s, y(s)) \text{ with } \frac{d^k}{ds^k} y(0) = y_0^{(k)}$$

for $k = 0, 1, \ldots, m-1$, the solution $y(s)$ can be written as a Volterra integral equation:

$$y(s) = \sum_{k=0}^{m-1} \frac{s^k}{k!} y_0^{(k)} + \frac{1}{\Gamma(q)} \int_0^s (s-\tau)^{q-1} f(\tau, y(\tau))\, d\tau. \tag{6}$$

The objective of the ABM method is to compute an approximation of $y(s)$ by leveraging the Volterra integral equation in (6).

Let $t_1 = 0 < t_2 < \cdots < t_N = T$ be regularly spaced points in the interval $[0, T]$. Loosely speaking, the ABM method presented in [8] interpolates the points $\{(t_i, f(t_i, y(t_i))\}_{i=1}^N$ by a piecewise linear function, and then uses the interpolating function to predict y_{k+1}. As a result, under certain smoothness assumptions on f, an error bound for $|y(t_i) - y_i|$ was obtained when t_i are equally spaced. In particular, the convergence rate $O(h^2)$ was obtained in Theorem C.4 of [8] for $q \geq 1$.

The modification of the ABM method presented here exchanges piecewise linear interpolation for scattered data interpolation by RKHSs. Since convergence rates based on the spacing of the data points are difficult to obtain explicitly for functions approximated through scattered data interpolation, a convergence rate depending on norms from RKHSs are obtained instead.

The first term y_1 is given by the initial condition

$$y_1 = \sum_{j=0}^{m-1} \frac{t_1^j}{j!} y_0^{(j)}.$$

For $k > 1$, the determination of the approximation, y_{k+1}, of $y(t_{k+1})$ is computed through two calculations. First, a *predictor* y_{k+1}^P is computed via

$$y_{k+1}^P = \sum_{j=0}^{m-1} \frac{t_{k+1}^j}{j!} y_0^{(j)} + \frac{1}{\Gamma(q)} \sum_{j=0}^{k} b_{j,k+1} f(t_j, y_j), \tag{7}$$

where $b_{j,k+1} = \frac{h^q}{q}((k+1-j)^q - (k-j)^q)$ arises from a piecewise constant approximation of the function $f(t, y(t))$ under the integral in (6) with equally spaced nodes $t_i = t_1 + (i-1)h$ and $h > 0$. The predictor given here is unchanged from (C.19) in [8].

The final estimation of $y(t_{k+1})$, also known as the *corrector*, for the kernelized ABM method is given by

$$y_{k+1} = \sum_{j=0}^{m-1} \frac{t_{k+1}^j}{j!} y_0^{(j)} + \frac{1}{\Gamma(q)} \int_0^{t_{k+1}} (s-\tau)^{q-1} \sum_{j=1}^{k+1} w_{j,k+1} K(\tau, t_j)\, d\tau, \tag{8}$$

where

$$\sum_{j=0}^{k+1} w_{j,k+1} K(t_i, t_j) = \begin{cases} f(t_i, y_i), & i = 1, \dots, k, \\ f(t_{k+1}, y_{k+1}^P), & i = k+1, \end{cases} \tag{9}$$

and K is some appropriately selected kernel function over $\mathbb{R}_+$. When $K(\lambda, t) = K_q(\lambda, t)$ the integral can be evaluated explicitly in (8) and

$$y_{k+1} = \sum_{j=0}^{m-1} \frac{t_{k+1}^j}{j!} y_0^{(j)} + \frac{w_{1,k+1}}{\Gamma(q+1)} t_{k+1}^q + \sum_{j=2}^{k+1} \frac{w_{i,k+1}}{t_j^q} \left(K_q(t_{k+1}, t_j) - 1 \right). \tag{10}$$

The corrector in (8) replaces (C.15) in [8], where instead of a piecewise linear interpolation of $\{f(t_i, y_i)\}_i$, scattered data interpolation is used with a kernel function.

As discussed in section 2, $\mathbf{w}_{k+1} = (w_{1,k+1}, \dots, w_{k+1,k+1})^T \in \mathbb{R}^{k+1}$ is the solution of

$$\begin{pmatrix} K(t_1,t_1) & \cdots & K(t_{k+1},t_1) \\ \vdots & \ddots & \vdots \\ K(t_1,t_k) & \cdots & K(t_{k+1},t_k) \\ K(t_1,t_{k+1}) & \cdots & K(t_{k+1},t_{k+1}) \end{pmatrix} \begin{pmatrix} w_{1,k+1} \\ \vdots \\ w_{k,k+1} \\ w_{k+1,k+1} \end{pmatrix} = \begin{pmatrix} f(t_1, y_1) \\ \vdots \\ f(t_k, y_k) \\ f(t_{k+1}, y_{k+1}^P) \end{pmatrix}. \tag{11}$$

In other words $\mathbf{w}_{k+1} := M_{k+1}^{-1} \mathbf{f}_{k+1}$, where M_{k+1} is the matrix in (11) and

$$\mathbf{f}_{k+1} := (f(t_1, y_1), \dots, f(t_k, y_k), f(t_{k+1}, y_{k+1}^P))^T.$$

THEOREM 4.1. *Let K be a continuous strictly positive definite kernel function over $\mathbb{R}_+$ with corresponding RKHS H, and let $q > 0$. Suppose that $f(t, y(t)) \in H$ where $D_*^q y(t) = f(t, y(t))$. For $k \in \mathbb{N}$, the error $|y(t_{k+1}) - y_{k+1}|$ is bounded by*

$$C \left\| f(t, y(t)) - P_{S_{k+1}} f(t, y(t)) \right\|_{[0,t_k]} + D \left\| P_{S_{k+1}} f(t, y(t)) - \sum_{i=1}^{k+1} w_{i,k+1} K(t, t_i) \right\|_{[0,t_k]}, \tag{12}$$

where $C, D > 0$ can be selected independent of k over a compact interval, $P_{S_{k+1}}$ is the projection onto $\operatorname{span}\{K(t, t_i)\}_{i=1}^{k+1}$, $(w_{1,k+1}, \dots, w_{k+1,k+1})^T$ is the solution to (11), and $\|\cdot\|_{[0,t_k]}$ is the norm arising from the restriction of the RKHS H to the set $[0, t_k]$.

Proof. Let $\sum_{i=1}^{k+1} w_{i,k+1} K(t, t_i)$ be as in (9), and let $P_{S_{k+1}} f(\cdot, y(\cdot))$ be the projection of $f(\cdot, y(\cdot))$ onto the subspace $S_{k+1} = \operatorname{span}\{K(t, t_i)\}_{i=1}^{k+1}$ of the RKHS H.

Consider the quantity $|y(t_{k+1}) - y_{k+1}|$ and replace $y(t_{k+1})$ with the fractional integral of its differential equation, and replace y_{k+1} by its definition:

$$\begin{aligned} &|y(t_{k+1}) - y_{k+1}| \\ &\quad = \left| \frac{1}{\Gamma(q)} \int_0^{t_{k+1}} (t_{k+1} - \tau)^{q-1} \left(f(\tau, y(\tau)) - \sum_{i=1}^{k+1} w_{i,k+1} K(\tau, t_i) \right) d\tau \right| \\ &\quad \le \left| \frac{1}{\Gamma(q)} \int_0^{t_{k+1}} (t_{k+1} - \tau)^{q-1} \left(f(\tau, y(\tau)) - P_{S_{k+1}} f(t, y(t)) \right) d\tau \right| \\ &\qquad + \left| \frac{1}{\Gamma(q)} \int_0^{t_{k+1}} (t_{k+1} - \tau)^{q-1} \left(P_{S_{k+1}} f(t, y(t)) - \sum_{i=1}^{k+1} w_{i,k+1} K(\tau, t_i) \right) d\tau \right| \end{aligned}$$

$$\leq \frac{t_{k+1}^q}{\Gamma(q+1)} \sup_{\tau\in[0,t_{k+1}]} |f(\tau,y(\tau)) - P_{S_{k+1}} f(\tau,y(\tau))|$$
$$+ \frac{t_{k+1}^q}{\Gamma(q+1)} \sup_{\tau\in[0,t_{k+1}]} \left| P_{S_{k+1}} f(\tau,y(\tau)) - \sum_{i=1}^{k+1} w_{i,k+1} K(\tau,t_i) \right|$$
$$\leq C\|f(\tau,y(\tau)) - P_{S_{k+1}} f(\tau,y(\tau))\|_{[0,t_{k+1}]}$$
$$+ D \left\| P_{S_{k+1}} f(\tau,y(\tau)) - \sum_{i=1}^{k+1} w_{i,k+1} K(\tau,t_i) \right\|_{[0,t_{k+1}]}.$$

The last inequality follows, since over a compact set I, $\sup_{y\in I}|f| \leq C\|f\|$ for some $C > 0$ when $f \in H$ and the kernel is continuous. ∎

It is necessary to require that $f(\cdot, y(\cdot)) \in H$ to establish the bound in (12). In the context of the Mittag-Leffler RKHS and the RKHS arising from the exponential kernel, this requirement may seem restrictive, since both spaces consist of infinitely differentiable functions. However, this bound holds no matter what RKHS is used. The usual requirement for other numerical methods is that $f(\cdot, y(\cdot))$ is twice continuously differentiable (cf. Theorem C.4 and the following discussion in [8]). The bound in (12) can be satisfied for functions in broader classes than C^2, for example, the native RKHSs for the Wendland RBFs correspond to Sobolev spaces, and the degree of smoothness of functions in those spaces corresponds to the smoothness of the chosen Wendland RBF [32]. Therefore, the assumption $f(\cdot, y(\cdot)) \in H$ is constrained only by the smoothness of the kernel function.

The norm contained in (12) is the norm obtained from the restriction of functions in H to $[0, t_k]$. The norm of the restricted RKHS is itself an RKHS norm, and as such the projection, $P_{S_{k+1}}$, can be realized through interpolation using the kernel functions as $\sum_{i=1}^{k+1} \tilde{w}_{i,k+1} K(t,t_i) = P_{S_{k+1}} f(t,y(t))$. Thus, $P_{S_{k+1}} f(t_i,y(t_i)) = f(t_i,y(t_i))$ for each $i = 1,\ldots,k+1$. By increasing the number of samples, the distance between $f(\cdot,y(\cdot))$ and $P_{S_{k+1}} f(\cdot,y(\cdot))$ can be made small in the Hilbert space norm, and in turn, the supremum norm of their difference will be correspondingly small. Since perfect information of $f(\cdot,y(\cdot))$ is not generally available to the numerical method, $\sum_{i=1}^{k+1} w_{i,k+1} K(\cdot,t_i)$ estimates $f(\cdot,y(\cdot))$ by interpolating $f(t_i,y_i)$ instead, as is done in other numerical methods [8]. Heuristically, if the time steps remain small enough, $f(t_i,y_i)$ is close to $f(t_i,y(t_i))$. Thus, $P_{S_{k+1}} f(\cdot,y(\cdot)))$ is close to $\sum_{i=1}^{k+1} w_{i,k+1} K(\cdot,t_i)$ in the Hilbert space norm contained in (12).

5. Numerical results.

5.1. Numerical results for the approximation of the Caputo fractional derivative. Implementation of the approximations detailed in sections 2 and 3 requires a selection of the number of sample points and the estimation of the Mittag-Leffler kernels. To evaluate of $K_q(t,t_i)$ in (5), the MATLAB routine from [24] was used to approximate the two-parameter Mittag-Leffler function. The collocation points were chosen to be uniformly spaced in the interval $[0,1]$.

Similar to [29], the functions $h_1(t) = (t-1/2)^4 + 1$ and $h_2(t) = \sin(2\pi t)$ were used to examine the performance of the developed approximation of the 1/2 and 1/8th order Caputo fractional derivative over the interval $[0,1]$. Since Theorem 3.2 provides convergence guarantees for functions in $ML^2(\mathbb{R}_+;q)$, the 1/2 and 1/8th Caputo fractional derivative of $h_3(t) = \sin(2\pi t^q)$ over $[0,1]$ was also used to examine the developed approximation method.

TABLE 1
Error bounds for Caputo estimation.

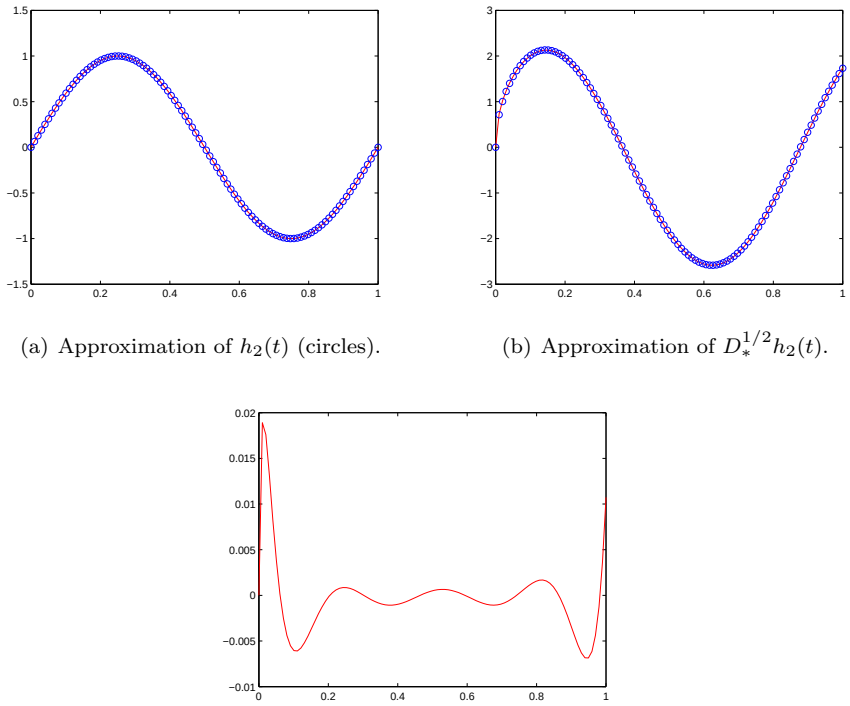
Function	q	N	E_1	E_2	E_3
h_1	1/2	5	$2.75E-3$	$1.37E-2$	$3.03E-3$
		7	$3.14E-4$	$1.86E-3$	$4.41E-4$
		10	$6.09E-5$	$3.26E-4$	$3.98E-5$
		50	$2.65E-5$	$2.93E-5$	$2.57E-5$
		100	$1.34E-5$	$9.98E-6$	$8.56E-6$
h_1	1/8	5	$4.06E-2$	$6.61E-2$	$2.26E-2$
		7	$2.68E-2$	$4.74E-2$	$1.55E-3$
		10	$3.70E-3$	$6.02E-3$	$4.40E-4$
		50	$2.08E-4$	$1.70E-4$	$1.61E-4$
		100	$1.19E-5$	$1.31E-5$	$1.76E-6$
h_2	1/2	5	$2.39E-1$	$8.93E-1$	$5.81E-1$
		7	$6.22E-2$	$3.44E-1$	$9.66E-2$
		10	$3.06E-3$	$1.90E-2$	$6.12E-3$
		50	$1.06E-3$	$1.36E-2$	$2.22E-3$
		100	$7.72E-4$	$1.31E-2$	$2.73E-3$
h_2	1/8	5	$4.21E-1$	$5.83E-1$	$4.52E-1$
		7	$9.90E-1$	$1.73E+0$	$4.65E-2$
		10	$7.71E-1$	$1.31E+0$	$3.90E-2$
		50	$7.46E-2$	$7.64E-2$	$2.49E-3$
		100	$1.05E-2$	$7.96E-3$	$7.60E-4$
h_3	1/2	5	$7.90E-2$	$2.03E+0$	$2.40E-1$
		7	$1.65E-1$	$2.91E+0$	$1.58E-1$
		10	$1.45E-2$	$4.25E-1$	$7.96E-3$
		50	$1.16E-4$	$1.15E-2$	$6.86E-5$
		100	$9.90E-5$	$8.18E-3$	$6.39E-5$
h_3	1/8	5	$5.22E-1$	$7.18E+0$	$2.49E-2$
		7	$4.80E-1$	$6.90E+0$	$2.27E-2$
		10	$9.41E-2$	$1.77E+0$	$5.76E-3$
		50	$4.22E-4$	$3.84E+0$	$1.51E-4$
		100	$1.20E-5$	$3.39E+0$	$6.94E-5$

Since h_1 and h_2 are continuously differentiable at the origin, $D_q^* h_1(0) = 0$ and $D_q^* h_2(0) = 0$. Therefore, a modified Mittag-Leffler kernel function was used for the approximation of h_1 and h_2. In particular, the kernel given by

$$K_{q,1}(\lambda, t) := K_q(\lambda, t) - \frac{\lambda^q t^q}{\Gamma(q+1)} = 1 + \sum_{n=2}^{\infty} \frac{t^{qn} \lambda^{qn}}{\Gamma(qn+1)}$$

was used for scattered data interpolation. The kernel $K_{q,1}$ is the kernel function for the subspace $\{t^q\}^\perp \subset ML^2(\mathbb{R}_+; q)$. For each $\lambda \in \mathbb{R}_+$, $D_*^q K_{q,1}(\lambda, 0) = 0$, which enables accurate approximation of $D_*^q f(0)$ when the function to be approximated is continuously differentiable near the origin. Note that this kernel is universal by the Muntz–Szasz theorem [25] and is strictly positive definite by Theorem 2.1.

Table 1 enumerates the results of the numerical experiments. The variable q represents the order of the Caputo fractional derivative applied to the function, as well as the order of the Mittag-Leffler space the approximation was performed under. The quantity N is the number of equally spaced collocation nodes in the unit interval $[0, 1]$, including the time $t = 0$ and $t = 1$. E_1 is the supremum norm of the error of approximation to the original function, while E_2 is the supremum norm over the interval $[0, 1]$ of the error in approximation of the Caputo fractional derivative of the respective order. Since it is known that the approximation may have poor performance near the origin, E_3 is the supremum norm of the error in approximation of the Caputo fractional derivative over the subinterval $[0.1, 0.9]$. Independent of the number of



(a) Approximation of $h_2(t)$ (circles).

(b) Approximation of $D_*^{1/2}h_2(t)$.

(c) Approximation error of $D_*^{1/2}h_2(t)$.

FIG. 1. *A representative approximation of the Caputo fractional derivative of order $q = 1/2$ of $h_2(t)$. This approximation was performed using $N = 10$. Figure* 1(a) *presents the approximation (solid line) to the function $h_2(t)$ (circles) with $N = 10$ and $q = 1/2$. Figure* 1(b) *presents the approximation (solid line) of $D_*^{1/2}h_2(t)$ (circles) with $N = 10$ and $q = 1/2$. Figure* 1(c) *demonstrates the approximation error of the approximation of $D_*^{1/2}h_2(t)$ with $N = 10$ and $q = 1/2$ shown in Figure* 1(b).

collocation points, the errors were estimated by sampling 200 regularly spaced points in the interval $[0, 1]$. The larger number of samples for the error functions guarantees that the error is estimated at points that do not coincide with the collocation points.

Figures 1(a)–1(c) present the results from a representative numerical experiment: the approximation of h_2's Caputo fractional derivative of order $1/2$ with 10 collocation nodes.

5.2. Numerical results for the kernelized fractional order adams–bashforth–moulton method. The execution of the kernelized ABM method was performed using two kernels, the exponential kernel $K_1(t, \lambda) = \exp(\lambda t)$ and the modified Mittag-Leffler kernel of the previous section, $K_{q,1}(t, \lambda)$. For both cases, the fractional integral in (6) can be explicitly evaluated in terms of the two-parameter Mittag-Leffler function

$$E_{\alpha,\beta}(z) := \sum_{n=0}^{\infty} \frac{z^n}{\Gamma(\alpha n + \beta)}.$$

Denoting the Riemann–Liouville fractional integral as

$$J_0^q f(s) := \frac{1}{\Gamma(q)} \int_0^s (s - \tau)^{q-1} f(\tau)\, d\tau$$

TABLE 2
Error bounds for kernelized ABM.

FODE	Kernel or method	q	N	E_4
(13)	K_1	1/2	5	$5.57E-1$
			10	$1.80E-1$
			50	$1.92E-2$
			100	$6.78E-3$
(13)	$K_{1/2,1}$	1/2	5	$5.61E-1$
			10	$1.85E-1$
			50	$1.83E-2$
			100	$6.29E-3$
(13)	ABM	1/2	5	$2.44E+0$
			10	$3.08E+0$
			50	$2.22E+0$
			100	$1.66E+0$
(14)	K_1	1/2	5	$3.10E-2$
			10	$8.06E-3$
			50	$2.25E-3$
			100	$1.69E-3$
(14)	$K_{1/2,1}$	1/2	5	$2.72E-2$
			10	$8.16E-3$
			50	$1.14E-3$
			100	$1.95E-3$
(14)	ABM	1/2	5	$2.45E-1$
			10	$1.82E-1$
			50	$9.34E-2$
			100	$6.91E-2$
(14)	K_1	1/8	5	$2.89E-1$
			10	$2.21E-1$
			50	$1.23E-1$
			100	$9.60E-2$
(14)	$K_{1/8,1}$	1/8	5	$3.25E-1$
			10	$2.56E-1$
			50	$1.57E-1$
			100	$1.19E-1$
(14)	ABM	1/8	5	$4.19E-1$
			10	$3.06E-1$
			50	$2.03E-1$
			100	$1.78E-1$

the following hold:

$$J_0^q K_1(t,\lambda) = J_0^q \exp(\lambda t) = t^q E_{1,q+1}(\lambda t),$$

$$J_0^q K_{q,1}(t,\lambda) = J_0^q \left(E_q(\lambda^q t^q) - \frac{\lambda^q t^q}{\Gamma(q+1)} \right) = \lambda^{-q} \left(E_q(\lambda^q t^q) - 1 \right) - \frac{\lambda^q t^{2q}}{\Gamma(2q+1)}.$$

Thus, the computation of the fractional integral in the kernelized ABM method, as in (8), can be evaluated analytically in terms of the Mittag-Leffler function by replacing the fractional integral of the kernel functions by the above formulas.

The results detailed in Table 2 show the errors of approximation of the solution to two FODEs,

$$D_*^q y(s) = y(s), y(0) = 1 \tag{13}$$

as well as

$$D_*^q y(s) = -y(s), y(0) = 1. \tag{14}$$

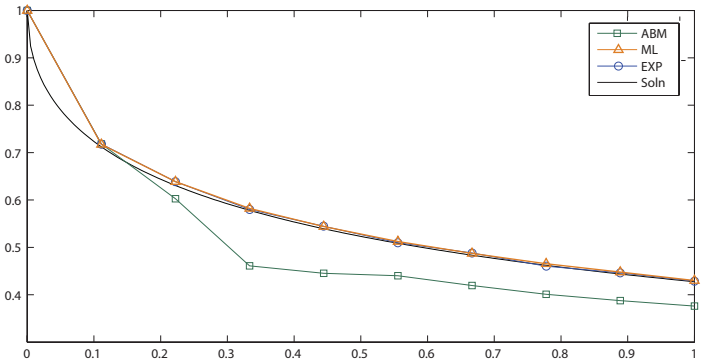


FIG. 2. *The solid unmarked line represents the solution to* (14) *with* $q = 1/2$. *The kernelized methods are the curves marked with circles and triangles, representing the numerical solution generated by the exponential kernel and the modified Mittag-Leffler kernel,* $K_{q,1}$, *respectively (these two curves are overlapping). The curve marked with squares represents the ABM method given in* [10]. *Each method was executed using* 10 *time steps.*

The solutions for these FODEs are $y(s) = E_q(s^q)$ and $y(s) = E_q(-s^q)$, respectively. The column labeled "FODE" indicates the FODE whose solution was estimated. The "Kernel or method" column indicates which kernel function was used to perform the estimation. The order of the derivative is given by q, and the number of collocation points is given by N. Finally, $E_4 = \max_{i=1,\dots,N} |y(t_i) - y_i|$. For comparison, the results of using the fractional order ABM method given in [10] is also listed. Figure 2 is a plot of the solution to (14) with $q = 1/2$, as well as the approximations of the solution produced by each method with 10 time steps.

6. Discussion.

6.1. Discussion of the estimation of the Caputo fractional derivative.

Table 1 indicates that as N increases, E_1 becomes small. Moreover, the approximation of the Caputo fractional derivative of the polynomial $h_1(t) = (t-1/2)^4+1$ tended to be more accurate than that of the sinusoid $h_2(t) = \sin(2\pi t)$ or of $h_3(t) = \sin(2\pi t^q)$. Heuristically, the establishment of an accurate approximation requires more collocation nodes given the greater variability of the function to be approximated; this may characterize the higher accuracy achieved for h_1.

The approximation of h_2 was not as accurate as that of h_3. The functions h_1 and h_2 are not explicitly in $ML^2(\mathbb{R}_+; q)$, whereas $h_3 \in ML^2(\mathbb{R}_+; q)$. However, since $ML^2(\mathbb{R}_+; q)$ is universal over $\mathbb{R}_+$, as is $\{t^q\}^\perp \subset ML^2(\mathbb{R}_+; q)$, there are functions in $ML^2(\mathbb{R}_+; q)$ that are arbitrarily close (in the uniform metric) to h_1 and h_2, and this justifies the approximation.

The E_2 column of Table 1 indicates that the Caputo fractional derivative estimate is poor at some points for h_1, h_2, and h_3. The maximum error was 3.39 when 100 nodes were used to approximate h_3 over the space $ML^2(\mathbb{R}_+; 1/8)$. However, when a neighborhood of the origin is excluded, the approximation error maximum was reduced to 6.94×10^{-5} over the interval $[0.1, 0.9]$. The discrepancy between columns E_2 and E_3 in Table 1 expresses the sensitivity of $K_q(\lambda, t)$ at $t = 0$ and was anticipated at the end of section 3.

Until future research can reduce the sensitivity of $K_q(\lambda, t)$ at $t = 0$, a composite approach can be used for approximating the fractional derivative of a function. Since functions having a Caputo fractional derivative for each point in the interval $[0, 1]$ are assumed to be differentiable over the interval, a linearization of a given function can be used near the origin, $[0, \delta]$, where the developed approximation of the Caputo fractional derivative is sensitive. Given enough collocation points, the approximation of the Caputo fractional derivative of a function will achieve any desired accuracy over $[\delta', 1]$ where $0 < \delta' < \delta$ by using scattered data interpolation. Thus, once the time variable leaves $[0, \delta]$, the approximation by Mittag-Leffler kernels may be used.

6.2. Discussion of the kernelized ABM method. Table 2 indicates that the kernels K_1 and $K_{q,1}$ performed comparably in the estimation of the solution to the FODEs given by (13) and (14). The minimum approximation error of the solution to (14) was achieved by $K_{1/2,1}$ with 50 time steps. The approximations degraded when the order of the FODE was decreased from 1/2 to 1/8. The results of Table 2 indicate that the kernelized ABM method can be used to produce accurate numerical solutions to FODEs.

Moreover, Table 2 indicates that the kernelized method outperforms the ABM method given in [10] for these particular systems. A heuristic explanation for the effectiveness of the kernelized method in this circumstance is that the kernels used to generate the numerical approximation have growth rates similar to the solution, which is nonlinear.

The kernelized ABM method requires more computation time than the ABM method in [8]. The difference in computation time is explained by the computational complexity of the ABM method described in [8] and the kernelized ABM presented in section 4. In particular, to perform the interpolation step the system in (11) must be solved. In [8], the algorithm for performing the piecewise linear interpolation is $O(k)$ in computational complexity. For algorithms that seek to solve (11) directly through matrix inversion, the computational complexity is typically between $O(k^2)$ and $O(k^3)$, which is the ultimate performance bottleneck. Exploiting the symmetry and positivity of the matrix, the conjugate gradient method can be used to improve performance. For kernels such as the Mittag-Leffler kernel, the matrix in (11) is dense; this means that the operation of matrix multiplication in the conjugate gradient method will be $O(k^2)$, resulting in an overall performance of $O(k^3)$ for the conjugate gradient method. However, if the matrix is sparse, such as when a compactly supported RBF is used, the operation of multiplication of a vector by a sparse matrix is $O(m)$, where m is the number of nonzero elements of the sparse matrix [27]. In particular, the Gram matrix for a compactly supported RBF is banded and has $O(k)$ nonzero entries. Therefore, the computational complexity may be improved to $O(k^2)$ by utilizing a compactly supported RBF in the kernelized ABM method. These arguments are general, and the complexities listed can be refined for particular algorithms for matrix inversion and multiplication.

7. Conclusion. This paper introduced a new RKHS (i.e., $ML^2(\mathbb{R}_+; q)$) for the purpose of approximating the Caputo fractional derivative of a function and for producing numerical solutions of FODEs. Theorem 3.2 established that for $f \in ML^2(\mathbb{R}_+; q)$, the Caputo fractional derivative of order q of f can be uniformly approximated over compact subsets of the positive real numbers by linear combinations of the Mittag-Leffler kernel functions. Numerical experiments demonstrated the effectiveness of the approximation methods over the interval $[0, 1]$. However, it was

observed that the approximations of f's Caputo fractional derivative near the origin exhibit sensitivity issues to be addressed in future work.

This paper also introduced the kernelized ABM method for producing numerical solutions to FODEs. The kernelized ABM method introduces a new approach to numerical estimation of solutions to FODEs through the use of kernel functions and scattered data interpolation. The method can achieve greater accuracy than piecewise linear approaches (cf. [10]) in certain cases, and this was demonstrated through numerical experiments detailed in Table 2.

REFERENCES

[1] J. Agler and J. E. McCarthy, *Pick Interpolation and Hilbert Function Spaces*, Grad. Stud. Math., AMS, Providence, RI, 2002.

[2] O. P. Agrawal and D. Baleanu, *A Hamiltonian formulation and a direct numerical scheme for fractional optimal control problems*, J. Vibration Control, 13 (2007), pp. 1269–1281, doi:10.1177/1077546307077467.

[3] O. P. Agrawal, O. Defterli, and D. Baleanu, *Fractional optimal control problems with several state and control variables*, J. Vibration Control, 16 (2010), pp. 1967–1976, doi:10.1177/1077546309353361.

[4] N. Aronszajn, *Theory of reproducing kernels*, Trans. Amer. Math. Soc., 68 (1950), pp. 337–404.

[5] R. Bagley and P. Torvik, *A theoretical basis for the application of fractional calculus to viscoelasticity*, J. Rheology, 27 (1983), pp. 201–210.

[6] R. Bagley and P. Torvik, *On the fractional calculus model of viscoelastic behavior*, J. Rheology, 30 (1986), pp. 133–155.

[7] M. Caputo and F. Mainardi, *Linear models of dissipation in anelastic solids*, Riv. Nuovo Cimento (1971-1977), 1 (1971), pp. 161–198, doi:10.1007/BF02820620.

[8] K. Diethelm, *The Analysis of Fractional Differential Equations*: An Application-Oriented Exposition using Differential Operators of Caputo Type, vol. 2004 of Lecture Notes in Math. 2004, Springer-Verlag, Berlin, 2010, https://dx.doi.org/10.1007/978-3-642-14574-2.

[9] K. Diethelm, N. Ford, A. Freed, and Y. Luchko, *Algorithms for the fractional calculus: A selection of numerical methods*, Comput. Methods Appl. Mech. Engrg., 194 (2005), pp. 743–773, doi:dx.doi.org/10.1016/j.cma.2004.06.006.

[10] K. Diethelm and A. Freed, *On the solution of nonlinear fractional-order differential equations used in the modeling of viscoplasticity*, in Scientific Computing in Chemical Engineering II, F. Keil, W. Mackens, H. Voss, and J. Werther, eds., Springer, Berlin, 1999, pp. 217–224.

[11] S. Esmaeili and G. V. Milovanovic, *Nonstandard Gauss-Lobatto quadrature approximation to fractional derivatives*, Fract. Calc. Appl. Anal., 17 (2014), pp. 1075–1099, https://dx.doi.org/10.2478/s13540-014-0215-z.

[12] R. Gorenflo, A. A. Kilbas, F. Mainardi, and S. V. Rogosin, *Mittag-Leffler Functions, Related Topics and Applications*, Springer Monogr. Math., Springer, Heidelberg, 2014, https://dx.doi.org/10.1007/978-3-662-43930-2.

[13] R. Gorenflo, J. Loutchko, Y. Luchko, and D. T. F. Mainardi, *Computation of the Mittag-Leffler function $E(\alpha, \beta)(z)$ and its derivative*, Fract. Calc. Appl. Anal., 5 (2002).

[14] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*, 2nd ed., Springer Ser. Statist., Springer, New York, 2009.

[15] C. G. Koh and J. M. Kelly, *Application of fractional derivatives to seismic analysis of base-isolated models*, Earthquake Eng. Struct. Dynam., 19 (1990), pp. 229–241, https://dx.doi.org/10.1002/eqe.4290190207.

[16] D. Liu, O. Gibaru, W. Perruquetti, and T. Laleg-Kirati, *Fractional order differentiation by integration and error analysis in noisy environment*, IEEE Trans. Automat. Control, 60 (2015), pp. 2945–2960, doi:10.1109/TAC.2015.2417852.

[17] R. L. Magin, *Fractional calculus in bioengineering: A tool to model complex dynamics*, in proceeding of the 13th International Carpathian Control Conference (ICCC), 2012, pp. 464–469, doi:10.1109/CarpathianCC.2012.6228688.

[18] A. B. Malinowska and D. F. M. Torres, *Introduction to the Fractional Calculus of Variations*, Imperial College Press, London, 2012, https://dx.doi.org/10.1142/p871.

[19] X. Moreau, O. Altet, and A. Oustaloup, *The CRONE suspension: Management of the dilemma comfort-road holding*, Nonlinear Dynam., 38 (2004), pp. 461–484, https://dx.doi.org/10.1007/s11071-004-3772-6.
[20] I. Moret and P. Novati, *On the convergence of Krylov subspace methods for matrix Mittag-Leffler functions*, SIAM J. Numer. Anal., 49 (2011), pp. 2144–2164, https://dx.doi.org/10.1137/080738374.
[21] K. B. Oldham and J. Spanier, *The Fractional Calculus*: *Theory and Applications of Differentiation and Integration to Arbitrary Order*, Harcourt Brace Jovanovich, New York, 1974.
[22] A. Pinkus, *Strictly positive definite functions on a real inner product space*, Adv. Comput. Math., 20 (2004), pp. 263–271.
[23] I. Podlubny, *Fractional-order systems and fractional-order controllers*, in Proceedings of the Conference Internationale Francophone dAutomatique, 1994, p. 2002.
[24] I. Podlubny, *Mittag-Leffler function*, MATLAB Central, 2005, http://www.mathworks.com/matlabcentral/fileexchange/8738-mittag-leffler-function.
[25] W. Rudin, *Real and Complex Analysis*, 3rd, ed., McGraw-Hill, New York, 1987.
[26] H. Seybold and R. Hilfer, *Numerical algorithm for calculating the generalized Mittag-Leffler function*, SIAM J. Numer. Anal., 47 (2008), pp. 69–88, https://www.jstor.org/stable/25663114.
[27] J. R. Shewchuk, *An Introduction to the Conjugate Gradient Method Without the Agonizing Pain*, Tech. report, Carnegie Mellon University, Pittsburgh, PA, 1994.
[28] I. Steinwart and A. Christmann, *Support Vector Machines*, Inform. Sci. Statist., Springer, New York, 2008.
[29] H. Sugiura and T. Hasegawa, *Quadrature rule for Abel's equations: Uniformly approximating fractional derivatives*, J. Comput. Appl. Math., 223 (2009), pp. 459–468, doi:http://dx.doi.org/10.1016/j.cam.2008.01.019.
[30] E. Tohidi and H. Saberi Nik, *A Bessel collocation method for solving fractional optimal control problems*, Appl. Math. Model., 39 (2015), pp. 455–465, https://dx.doi.org/10.1016/j.apm.2014.06.003.
[31] L. N. Trefethen, *Approximation Theory and Approximation Practice*, SIAM, Philadelphia, 2013.
[32] H. Wendland, *Scattered Data Approximation*, Cambridge Monogr. Appl. Comput. Math. 17, Cambridge University Press, Cambridge, UK, 2005.
[33] M. Zamani, M. Karimi-Ghartemani, N. Sadati, and M. Parniani, *Design of a fractional order {PID} controller for an {AVR} using particle swarm optimization*, Control Engineering Practice, 17 (2009), pp. 1380–1387, doi:http://dx.doi.org/10.1016/j.conengprac.2009.07005.