Mathematics

http://www.aimspress.com/journal/Math

AIMS Mathematics, 5(4): 3284–3297.
DOI:10.3934/math.2020211
Received: 16 January 2020
Accepted: 26 March 2020
Published: 30 March 2020

Research article

A sigmoidal fractional derivative for regularization

Mostafa Rezapour, Adebowale Sijuwade* and Thomas Asaki

Department of Mathematics and Statistics, Washington State University, Pullman, WA, USA

The first two authors contributed equally to this work.

* **Correspondence:** Email: adebowale.sijuwade@wsu.edu.

Abstract: In this paper, we propose a new fractional derivative, which is based on a Caputo-type derivative with a smooth kernel. We show that the proposed fractional derivative reduces to the classical derivative and has a smoothing effect which is compatible with ℓ_1 regularization. Moreover, it satisfies some classical properties.

Keywords: fractional calculus; Caputo derivative; regularization
Mathematics Subject Classification: 26A33

1. Introduction

Fractional calculus has undergone significant developments in recent years and has found use in physics, engineering, economics, etc [1–3]. Classical results about the Riemann-Liouville and Caputo derivatives as well as fractional differential equations can be found in [4, 5]. In [11] and [33], Caputo and Fabrizio suggested a new fractional derivative, whose properties were investigated by Losada and Nieto [15]. This fractional derivative was utilized in various applications, including the fractional Nagumo equation in Alqahtani et al. [23], coupled systems of time-fractional differential problems in Alsaedi et al. [24] and Fischer's reaction-diffusion equation in Atangana et al [25]. More applications of the Caputo-Fabrizio fractional derivative can be found in Aydogan et al [26]. and Atangana et al [27].

For $0 \le \alpha \le 1$, $-\infty < a < t$, $f \in H^1(a,b)$ and $b > a$, the Caputo fractional derivative is defined by

$$ {}_a^C D_t^\alpha f(t) = \frac{1}{\Gamma(1-\alpha)} \int_a^t f'(s)(t-s)^{-\alpha} ds. \tag{1}$$

By replacing the term $\frac{1}{\Gamma(1-\alpha)}$ with the normalization constant $M(\alpha)$ such that $M(0) = M(1) = 1$ and adjusting the kernel $(t-s)^{-\alpha}$, we obtain the Caputo-Fabrizio fractional derivative defined by

$$ {}^{CF}_{a}D^{\alpha}_{t}f(t) = \frac{M(\alpha)}{\Gamma(1-\alpha)}\int_a^t f'(s)\exp\left(\frac{-\alpha(t-s)}{\alpha-1}\right)ds. \tag{2} $$

The Caputo-Fabrizio fractional derivative of a constant vanishes as does the usual Caputo derivative, however the new kernel $\exp\left(\frac{-\alpha}{\alpha-1}\right)$ is no longer singular for $s = t$. Caputo and Fabrizio try to extend their definition in [11] to functions in L^1 by

$$ {}^{CF}D^{(\alpha)}_{t}f(t) = \frac{M(\alpha)}{\Gamma(1-\alpha)}\int_{-\infty}^t (f(s)-f(t))\exp\left(\frac{-\alpha(t-s)}{\alpha-1}\right)ds. $$

Algahtani et al. [23] show that the nonlinear Nagumo equation given by

$$ {}^{CF}{}_{0}D^{\alpha}_{t}u(x,t) + \beta u(x,t)^n\partial_x u(x,t) = \partial_x(\alpha u(x,t)^n\partial_x u(x,t)) + \gamma u(x,t)(1-u^m)(u^m-\delta), \tag{3} $$

where $0 < \alpha < 1$ and β, γ, δ are constant, subject to the boundary conditions

$$ u(x,0) = f(x),\ u(0,t) = g(t) $$

has an exact solution. The authors show that this PDE can be reformulated in terms of a Lipschitz kernel. Existence of the exact solution is shown using a fixed point approach and uniqueness is provided, given that suitable assumptions are made about the Lipschitz constant. Their study claims that an exponential kernel is in some sense a better kernel than a power function, since the lack of a singularity provides a better filtration effect. In the context of fractional differential equation applications, since the associated functions are not defined in a Banach space, only approximate solutions to certain fractional differential equations can be investigated. The methods used to handle fractional differential problems such as ${}^{CF}D^{\alpha}f(t) = g(t, f(t))$, cannot be extended to the problems resembling ${}^{CF}D^{\alpha}f(t) = g(t, f(t), {}^{CF}D^{\alpha}f(t))$.

In Baleanu et al [14], the Caputo-Fabrizio fractional derivative on the Banach space $C_{\mathbb{R}}[0,1]$ is considered in the context of higher order series-type fractional integrodifferential equations. More precisely, an extended Caputo-Fabrizio type fractional derivative is provided of order $0 \leq \alpha < 1$ on $C_{\mathbb{R}}[0,1]$ for $b > 0$ by

$$ {}^{CF}{}_{N}D^{\alpha}f(t) = \frac{M(\alpha)}{1-\alpha}(f(t)-f(0))\exp\left(\frac{-\alpha t}{1-\alpha}\right) + \frac{\alpha M(\alpha)}{(1-\alpha)^2}\int_0^t (f(t)-f(s))\exp\left(\frac{-\alpha(t-s)}{1-\alpha}\right)ds. $$

These authors use a standard fixed point approach to establish uniqueness of solutions to fractional series-type differential problems such as

$$ {}^{CF}{}_{N}D^{\alpha}f(t) = \sum_{j=0}^{\infty}\frac{{}^{CF}{}_{N}D^{\rho^{[j]}}g(t, f(t), (\phi f)(t), h(t){}^{CF}_{N}D^{\gamma}f(t), g(t){}^{CF}_{N}D^{\delta}f(t))}{2^j}, $$

with initial condition $f(0) = 0$ and $\alpha, \gamma, \delta, \rho \in (0,1)$.
An extension of this type which is compatible with orders beyond $(0,1)$ has yet to be provided.

The Caputo-Fabrizio fractional derivative is discussed in the setting of distributions in [28]. Other types of fractional derivatives can be found in Katugampola [22] and Oliveira et al [6]. In de Oliveira

[12], it is shown that the choice of kernel in a Caputo-type fractional derivative is connected to the Laplace transform via convolution.

Let $\mathscr{I}$ denote the Schwarz class of smooth test functions whose derivatives decay at infinity. Moreover, let $\mathscr{I}'$ denote the space of continuous linear functionals on $\mathscr{I}$. The distributional derivative
{ T' } is defined as in [32]

$$\int_{\mathbb{R}} T'(t)\phi(t)dt = -\int_{\mathbb{R}} T(t)\phi(t)dt, \tag{4}$$

for all smooth compactly supported test functions ϕ on $\mathbb{R}$.The distributional Laplace transform is given by

$$F(s) = \mathscr{L}(\phi(t)) = \mathscr{F}(\phi(t)e^{-\sigma t})(\mu),$$

where $s = \sigma + i\mu$, $\mu < 0$ and $\phi(t)e^{-\sigma t} \in \mathscr{I}'$. Suppose that f is supported on $(0, \infty)$ such that $\sigma > 0$ and $f(t)e^{-\sigma t} \in \mathscr{I}'$. It follows that the Laplace transform of the derivative is given by

$$\mathscr{L}(\phi'(t))(s) = s\mathscr{L}(\phi(t))(s).$$

Let $\mathscr{L}$ denote the distributional Laplace transform defined by

$$\mathscr{L}(f'(x)) = \mathscr{L}^{-1}(s\mathscr{L}(f)).$$

One can define a more general fractional derivative as follows. Suppose that $\Phi(s, \alpha)$ is a fractional integrodifferential operator and $K(t, s) : \mathbb{R}^2 \to \mathbb{R}$ is a continuous kernel. Let the corresponding operator $\phi(s, \alpha)$ be defined for some fractional derivative D^α such that

$$\mathscr{L}(D^\alpha f(t)) = \Phi(s, \alpha)\mathscr{L}(f(t)),$$

where $\Phi(s, 1) = s, \Phi(s, -1) = \frac{1}{s}$ and $\Phi(s, 0) = 1$. Then, letting $\Phi(s, \alpha) = s\mathscr{L}(K(s, t, \alpha))$. Proceeding with the Convolution Theorem, we are left with a Caputo-type fractional operator of the form

$${}_aD^\alpha_K f(t) = \int_a^t K(t - s, \alpha) f'(s)ds, \tag{5}$$

which is dependent on the choice of kernel K. For $f \in H^1(a, b)$, and $n \in \mathbb{N}$, we can spot commonly used kernels such as the Caputo kernel $K_1 = \frac{1}{\Gamma(1-\alpha)}(t - s)^{\lceil\alpha\rceil-\alpha-1}$, the Caputo-Fabrizio kernel $K_2 = \frac{M(\alpha)}{1-\alpha}\exp\left(\frac{-\alpha(t-s)}{\alpha-1}\right)$ and the Gaussian kernel $K_3 = \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(\frac{-t^2}{2\sigma^2}\right)$ [4, 10, 13].

The memory principle for fractional derivatives describes the history of $f(t)$ near the terminal point $t = a$. Let L denote the memory length, satisfying $a + L \le t \le b$. Define the error in approximating the fractional derivative by

$$E_{L,\alpha,a}(t) = |{}_aD^\alpha_K f(t) -_{t-L} D^\alpha_K f(t)|,$$

where ${}_aD^\alpha_K f(t)$ is as in (5). If $f'(t) \le M$ for $a < t < b$ and $0 < \alpha < 1$, we have the following error estimate for the Caputo fractional derivative

$$E_{L,\alpha,a}(t) = \left| \frac{1}{\Gamma(1-\alpha)} \int_{t-L}^{t} f'(s)(t-s)^{-\alpha} ds \right| \leq \frac{ML^{1-\alpha}}{|\Gamma(2-\alpha)|}.$$

For all $\epsilon > 0$, if $E_{L,\alpha,a}(t) \leq \epsilon$ with $a + L \leq t \leq b$, we have

$$L \geq \left(\frac{M}{\epsilon|\Gamma(2-\alpha)|} \right)^{\frac{1}{\alpha-1}}. \tag{6}$$

Therefore, the Caputo fractional derivative with terminal a can be approximated by the corresponding fractional derivative with lower limit $t - L$, with the level of accuracy described above.

In this work, we propose a different fractional derivative that has a smooth kernel. Our primary interest in defining this fractional derivative is the improvement of machine learning algorithms. Caputo-type fractional derivatives have been applied in machine learning, such as in Pu et al [10]. In particular, fractional order gradient methods have been considered in order to improve the performance of the integer order methods. For example, suppose that $f : \mathbb{R}^n \to \mathbb{R}$ is convex and differentiable with a Lipschitz gradient, then the integer order gradient method defined by

$$x_{k+1} = x_k - \mu \nabla f(x_k)$$

has a linear convergence rate. Improving the performance of the integer-order gradient method is critical in optimization problems. In recent literature, fractional calculus has been thought to improve the integer order gradient method due to nonlocality and the memory principle. Fractional order gradient methods have been proposed based on the Caputo fractional derivative that offer competitive convergence rates. For example, in [19], a Caputo fractional gradient method is proposed that is shown to be monotone and exhibit strong convergence.

Fractional derivatives were used in the backpropagation algorithm for feedforward neural networks and convolutional neural networks in [20, 31]. In both studies, the rate of convergence was shown to exceed the rate of integer-order methods. Fractional-order methods have been used to investigate complex-valued neural networks in [17] and recurrent neural network models in [30]. In [19] and [16], gradients based on the Caputo fractional derivative are used to update parameters while integer order gradients are used to handle backpropagation allowing for simpler computation. The experiments therein are shown to improve the accuracy of the neural network's performance compared to integer-order methods while being equally costly.

In the training of machine learning models, one often needs to obtain weights of the features which optimize the training data. In the case of maximum likelihood training, regularization is typically needed so that the model does not overfit the training data. In ℓ_p regularization, the weight vector is penalized by its ℓ_p norm. While the case for $p = 1$ and $p = 2$ are very common and result in similar levels of accuracy, ℓ_1 regularization is much more practical. Due to its sparsity, ℓ_1 regularization is less memory intensive and more time-effective than ℓ_2 regularization. On the other hand, ℓ_1 regularization is problematic in that during the update process, the gradient of the regularization term is not differentiable at the origin as the error function given below

$$E_{\ell_1} = E + \lambda \sum_{k=1}^{N} |x_k| \tag{7}$$

has classical derivative

$$\frac{\partial E_{\ell_1}}{\partial x_j} = \frac{\partial E}{\partial x_j} + \lambda \text{sgn}(x_j).$$

A typical remedy to this problem is to use the stochastic gradient descent method, which approximates the gradient using the training data. Although time efficient for training, when the dimension of the feature space is large, the update process slows down significantly. Furthermore, the model becomes less sparse after training the data. The discontinuity induced by the regularizer proves to be problematic as it adjusts the direction of descent. The use of sigmoids in regularization problems has been previously explored as in Krutikov [29], but not in the context of fractional derivatives. Another remedy to the aforementioned problem is the use of fractional gradients over the classical descent methods. These methods are still in their infancy and problematic in that convergence to the local optimum is not always guaranteed, even when the algorithm converges. Furthermore, these methods often require an adjustment to the fractional derivative by truncation as in [9], variable order techniques as in [18], and methods based on the memory principle (6) due to the computational expense and the failure of the Caputo kernel to be smooth.

We would also like our operator to be nonlocal. In [13], it is shown that unlike the Caputo derivative, the Caputo-Fabrizio fractional derivative is not a nonlocal operator. The linear fractional differential equation

$$\lambda({}^{CF}_{a}D^{\alpha}_{t} f(t)) + \nu(t)g(t) + \eta(t, t_0)Y(t_0) = 0$$

is shown to reduce to a first-order ordinary differential equation. This means that the Caputo-Fabrizio derivative cannot sufficiently describe processes with nonlocality and memory. With the correct choice of kernel, this complication can be avoided.

2. Main results

In this section, we define a new left-sided fractional derivative. We show that the proposed fractional derivative reduces to the H^1 derivative as the order approaches 1. In the results to follow, for $0 < \alpha \leq 1$, we will let $C_1(\alpha)$ denote a normalization constant $\frac{C(\alpha)}{\Gamma(2-\alpha)}$ satisfying $C(\alpha)\Gamma(1-\alpha) \to \frac{1}{2}$ as $\alpha \to 1^-$.

Definition 2.1. (Left sigmoidal fractional derivative) Let $0 < \alpha \leq 1$, $f \in H^1((a, b))$, $t > a$ and $\{f(t)\}'$ denotes the H^1 distributional derivative as in (4). We define a new fractional derivative by

$$^{\sigma}D^{\alpha}_{a} f(t) = C_1(\alpha) \int_a^t \{f(s)\}' \operatorname{sech}^2\left(\frac{s-t}{1-\alpha}\right) ds. \tag{8}$$

Now, we show that the left sigmoidal fractional derivative reduces to the H^1 derivative.

Theorem 2.1. (Reduction to classical derivative) Suppose $f \in H^1(a, b)$, then

$$\lim_{\alpha \to 1^-} {}^{\sigma}D^{\alpha}_{a} f(t) = \{f(t)\}'. \tag{9}$$

Proof.

$$\lim_{\alpha\to1^-} {}^{\sigma}D_a^{\alpha}f(t) = \frac{C(\alpha)}{\Gamma(2-\alpha)}\lim_{\alpha\to1^-}\int_a^t \{f(s)\}'\,\mathrm{sech}^2\Big(\frac{s-t}{1-\alpha}\Big)ds$$

$$= \frac{2C(\alpha)}{\Gamma(1-\alpha)}\lim_{\alpha\to1^-}\int_a^t \{f(s)\}'\frac{\mathrm{sech}^2\,(\frac{s-t}{1-\alpha})}{2(1-\alpha)}ds = \frac{2C(\alpha)}{\Gamma(1-\alpha)}\lim_{\alpha\to1^-}\int_a^t \{f(s)\}'\frac{\mathrm{sech}^2\,(\frac{s-t}{1-\alpha})}{2(1-\alpha)}ds$$

$$= \lim_{\alpha\to1^-}\frac{2C(\alpha)}{\Gamma(1-\alpha)}\Big(\int_a^t \{f(s)\}'\lim_{\alpha\to1^-}\frac{\mathrm{sech}^2\Big(\frac{s-t}{1-\alpha}\Big)}{2(1-\alpha)}ds\Big) = \int_a^t \{f(s)\}'\delta(s-t)ds = \{f(t)\}',$$

where the last result follows from the observation that
$\delta(t)$ is the Dirac distribution.

□

In the following theorem, we show that this left sigmoidal fractional derivative is commutative with respect to the classical derivative.

Theorem 2.2. Suppose that f is at least twice continuously differentiable and ${}^{\sigma}D_a^{\alpha}f(t)$ is differentiable. If $f'(a) = 0$, then

$$ {}^{\sigma}D_a^{\alpha}({}^{\sigma}D_a^{1}f(t)) = {}^{\sigma}D_a^{1}({}^{\sigma}D_a^{\alpha}f(t)), \tag{10}$$

where $0 < \alpha < 1$.

Proof. From (8), integrating by parts yields

$$ {}^{\sigma}D_a^{\alpha}({}^{\sigma}D_a^{1}f(t)) = C_1(\alpha)\int_a^t f''(s)\,\mathrm{sech}^2\Big(\frac{s-t}{1-\alpha}\Big)ds$$

$$= \frac{f'(t)}{1-\alpha} + \frac{2C(\alpha)}{\Gamma(2-\alpha)(1-\alpha)}\int_a^t f''(s)\,\mathrm{sech}\Big(\frac{s-t}{1-\alpha}\Big)\tanh\Big(\frac{s-t}{1-\alpha}\Big)ds, \tag{11}$$

so we have

$$ {}^{\sigma}D_a^{1}({}^{\sigma}D_a^{\alpha}f(t)) = \lim_{\gamma\to1^-} {}^{\sigma}D_a^{\gamma}({}^{\sigma}D_a^{\alpha}f(t)) = \frac{d}{dt}({}^{\sigma}D_a^{\alpha}f(t)) = C_1(\alpha)\frac{d}{dt}\int_a^t f'(s)\,\mathrm{sech}^2\Big(\frac{s-t}{1-\alpha}\Big)ds$$

$$= \frac{f'(t)}{1-\alpha} + \frac{2C(\alpha)}{\Gamma(2-\alpha)(1-\alpha)}\int_a^t f''(s)\,\mathrm{sech}\Big(\frac{s-t}{1-\alpha}\Big)\tanh\Big(\frac{s-t}{1-\alpha}\Big)ds, \tag{12}$$

appealing to the Leibniz integral rule

$$\frac{d}{dt}\Big(\int_{a(t)}^{b(t)} f(t,s)ds\Big) = f(t,b(t))b'(t) - f(t,a(t))a'(t) + \int_{a(t)}^{b(t)}\frac{\partial}{\partial t}f(s,t)dt.$$

From (11) and (12), the desired result is obtained. □

In the next theorem, we show that the left sigmoidal fractional derivative does not satisfy the memory principle in the sense of (6). More precisely, the next theorem implies that we show that the left sigmoidal fractional derivative can be approximated by the corresponding fractional derivative with lower limit $t - L$ with increased accuracy for orders in which $C_1(\alpha)$ is large.

Theorem 2.3. (Memory principle) Suppose that f is differentiable on (a, b), $a + L \leq t \leq b$ and $0 < \alpha < 1$. For every $\epsilon > 0$, if there exists $C_0 > 0$ such that $f'(t) \leq C_0$, then

$$L \geq (1-\alpha)(|C_1(\alpha)|C_0\epsilon^{-1})^{\frac{1}{2}}. \tag{13}$$

Proof. Making use of the inequality

$$\cosh(s) \geq \sqrt{1+s^2},$$

we have

$$|{}^{\sigma}D_a^{\alpha} f(t) - {}^{\sigma}D_{t-L}^{\alpha} f(t)| = C_1(\alpha)\int_a^{t-L} f'(s)\,\mathrm{sech}^2\Big(\frac{s-t}{1-\alpha}\Big)ds \leq C_1(\alpha)C_0\int_a^{t-L}\frac{ds}{1+(\frac{s-t}{1-\alpha})^2} \leq \frac{C_1(\alpha)C_0}{1+(\frac{L}{1-\alpha})^2},$$

and the result follows.

□

In the theorem below, we show that our new fractional derivative provides a sigmoidal approximation to functions that have a piecewise linear H^1 distributional derivative. For instance, the proposed left sigmoidal fractional derivative is compatible with ℓ_1-regularization. In the case of the ℓ_1 norm, it can be used to define a fractional gradient, which approximates its classical gradient via a family of sigmoids as α approaches 1. This is promising in the context of gradient descent algorithms.

Theorem 2.4 (Norm-1 compatibility) ${}^{\sigma}D_a^{\alpha}$ provides a smooth approximation to the ℓ_1 norm defined by

$$||x||_1 = \sum_{k=1}^{n} |x_k|$$

as $\alpha \to 1$ in the sense that for the error function E given in (7), ${}^{\sigma}D_a^{\alpha}E_{\ell_1}(x_j)$ is given by

$${}^{\sigma}D_a^{\alpha}E(x_j) + \lambda C_1(\alpha)(\alpha-1)\tanh\Big(\frac{a-x_j}{1-\alpha}\Big),$$

where $a > 0$.

Proof. The result follows from the observation that

$$C_1(\alpha)\int_a^t \{|s|\}\,\mathrm{sech}^2\Big(\frac{s-t}{1-\alpha}\Big)ds = C_1(\alpha)\int_a^t H(t)\,\mathrm{sech}^2\Big(\frac{s-t}{1-\alpha}\Big)ds,$$

$$= C_1(\alpha)(\alpha-1)\tanh\Big(\frac{a-t}{1-\alpha}\Big)ds \to \frac{1}{2}(2H(t)-1) \text{ as } \alpha \to 1^-,$$

where $H(t)$ is the Heaviside function.

□

Theorem 2.5. (Mittag-Leffler function). Suppose that $\gamma, \eta > 0$ and $0 < a < t$. Then

$${}^{\sigma}D_a^{\alpha}E_{\gamma,\eta}(t) \leq C_1(\alpha)E_{\gamma,\eta}(t-a),$$

where $E_{\gamma,\eta}(z) = \sum_{k=0}^{\infty} \frac{z^k}{\Gamma(\gamma k+\eta)}$ is the two-parameter Mittag-Leffler function

Proof.

$$ {}^{\sigma}D_a^{\alpha} E_{\gamma,\eta}(t) = \int_a^t \operatorname{sech}^2\left(\frac{s-t}{1-\alpha}\right) \frac{d}{ds} \sum_{k=0}^{\infty} \frac{s^k}{\Gamma(\gamma k+\eta)} ds $$

$$ = \int_a^t \operatorname{sech}^2\left(\frac{s-t}{1-\alpha}\right) \sum_{k=0}^{\infty} \frac{k s^{k-1}}{\Gamma(\gamma k+\eta)} = \sum_{k=0}^{\infty} \frac{k}{\Gamma(\gamma k+\eta)} \int_a^t s^{k-1} \operatorname{sech}^2\left(\frac{s-t}{1-\alpha}\right) ds $$

$$ \leq \sum_{k=0}^{\infty} \frac{k}{\Gamma(\gamma k+\eta)} \int_a^t s^{k-1} ds = \sum_{k=1}^{\infty} \frac{(t-a)^k}{\Gamma(\gamma k+\eta)}. $$

□

Theorem 2.6. Suppose that $f \geq 0$, $1 < p < \infty$, $0 < \alpha < 1$ and $0 < t \leq T$. If $f \geq 0$ is differentiable with $f' \in L^p(\mathbb{R})$ and M is the maximal operator of f given by

$$ Mf(x) = \sup_{t \to 0} \frac{1}{2(a+x)} \int_{a-x}^{a+x} f(t) dt, $$

then

(a) ${}^{\sigma}D_{-t}^{\alpha} f(t) \leq 2TC_1(\alpha) M(|f'|)(0)$

(b) ${}^{\sigma}D_a^{\alpha} f(t)$ is integrable on $\mathbb{R}$.

Proof. (a) Since

$$ {}^{\sigma}D_{-t}^{\alpha} Mf(t) = C_1(\alpha) \int_{-t}^{t} f'(s) \operatorname{sech}^2\left(\frac{s-t}{1-\alpha}\right) ds \leq 2tC_1(\alpha) \cdot \frac{1}{2t} \int_{-t}^{t} f'(s) ds $$

$$ \leq 2TC_1(\alpha) \sup_{t>0} \frac{\int_{-t}^{t} |f'(s)| ds}{t} = 2TC_1(\alpha) M|f'|(0). $$

(b) From Young's convolution inequality, $\|f \star g\|_{L^r} \leq \|f\|_{L^p} \|g\|_{L^{\frac{pr}{p+r(p-1)}}}$.

$$ \int_{-\infty}^{\infty} {}^{\sigma}D_a^{\alpha} Mf(t) dt \leq C_1(\alpha) \int_{-\infty}^{\infty} \int_a^t |f'(s) \operatorname{sech}^2\left(\frac{s-t}{1-\alpha}\right)| ds $$

$$ = \left\| f'(t) \star \operatorname{sech}^2\left(\frac{t}{\alpha-1}\right) \right\|_{L^1(\mathbb{R})} \leq \|f'\|_{L^p(\mathbb{R})} \left\| \operatorname{sech}^2\left(\frac{t}{\alpha-1}\right) \right\|_{L^{\frac{p}{2p-1}}(\mathbb{R})} < \infty. $$

□

The next theorem describes the effect of the Laplace and Fourier transforms, which can extend to distributions as in de Oliveira [6]. The Convolution Theorem connects our choice of kernel as in (5) via the operator $\Phi(s,\alpha) = s\mathscr{L}(K(s,t,\alpha))$. In this case, $\Phi(s,\alpha)$ depends on the digamma function $\Psi(z) = \frac{\Gamma'(z)}{\Gamma(z)}$. This shows that the left-sigmoidal fractional derivative does not reduce to the left-sided Riemann-Liouville fractional derivative.

Theorem 2.7. (Transformations) Suppose that $0 < \alpha < 1$, $Re(s) > 0$, $\omega \in \mathbb{R}$, $a \in \mathbb{R}$ and f is a differentiable function of exponential order such that $f(0) = 0$. If $T_1(s), T_2(\omega)$ are defined by

$$T_1(s) = \left(1 + s\left(\frac{\Psi(\frac{2+s}{4}) - \Psi(\frac{s}{4})}{2}\right)\right),\ T_2(\omega) = \sqrt{\frac{\pi}{2}}\operatorname{csch}\left(\frac{\pi\omega}{2}\right),$$

then

(a) $\mathscr{L}({}^{\sigma}D_0^{\alpha} f(t))(s) = C_1(\alpha)(s(\alpha-1))^2 T_1((\alpha-1)s)\mathscr{L}(f)(s)$

(b) $\mathscr{F}({}^{\sigma}D_0^{\alpha})(\omega) = -C_1(\alpha)\omega^2|\alpha-1|(\alpha-1)T_2((\alpha-1)s)\mathscr{F}(f)(\omega)$,

where $\mathscr{L}(f)(s)$ denotes the Laplace transform of f and $\mathscr{F}(f)(\omega)$ denotes the Fourier transform of f.

Proof. (a) follows from a standard application of the Convolution theorem. By using the dilation property $\mathscr{L}(f(at)) = \frac{F(\frac{s}{a})}{a}$, we have

$$\frac{\mathscr{L}({}^{\sigma}D_0^{\alpha} f(t))(s)}{C_1(\alpha)} = \mathscr{L}\left(f' \star \operatorname{sech}^2\left(\frac{t}{\alpha-1}\right)\right) = \mathscr{L}(f')\mathscr{L}\left(\operatorname{sech}^2\left(\frac{t}{\alpha-1}\right)\right)$$

$$= s(\alpha-1)\mathscr{L}(f)(s)\mathscr{L}(\operatorname{sech}^2)(s(\alpha-1))$$

$$= ((\alpha-1)s)^2\mathscr{L}(f)(s)\mathscr{L}(\tanh)(s(\alpha-1)).$$

The transform $\mathscr{L}(\tanh t)$ is handled as follows

$$s^2\mathscr{L}(\tanh(t))(s) = s^2\mathscr{L}(\tanh(t)) = s^2\int_0^{\infty}\frac{e^{-st}(1-e^{-2t})}{1+e^{-2t}}dt$$

$$= s^2\int_0^{\infty} e^{-st}(1-e^{-2t})\sum_{k=0}^{\infty}(-e^{-2t})^k dt.$$

Because of the absolute convergence of the monotone decreasing sum $\sum_{k=0}^{\infty}(-1)^k e^{-2kt}dt$ and the nondecreasing nature of its partial sums, we can exchange integration and summation using the Lebesgue Monotone Convergence Theorem. Continuing, we have

$$s + 2s^2\sum_{k=1}^{\infty}(-1)^k\mathscr{L}(e^{-2kt}) = s + 2s^2\sum_{k=1}^{\infty}\frac{(-1)^k}{2k+s}$$

$$= s + 2s\sum_{k=1}^{\infty} = \frac{(-1)^k}{\frac{2k}{s}+1} = s\left(1 + s\left(\frac{\Psi(\frac{2+s}{4}) - \Psi(\frac{s}{4})}{2}\right)\right).$$

The identity

$$\sum_{k=0}^{\infty}\frac{(-1)^k}{sk+1} = \frac{\Psi(\frac{s+1}{2s}) - \Psi(\frac{1}{2s})}{2s}$$

used above comes from the Lerch transcendent, defined by

$$\Phi(s,z,a)=\sum_{k=0}^{\infty}\frac{z^k}{(a+k)^s},$$

where $|z|<1$, $a\neq 0,-1,-2,...$ and using the dilation property once more, the result follows.

(b) We proceed as in (a).

$$\mathscr{F}({}^{\sigma}D_0^{\alpha}f(t))=\int_{-\infty}^{\infty}({}^{\sigma}D_0^{\alpha}f(t))e^{i\omega t}dt=\mathscr{F}(f')\mathscr{F}\Big(\operatorname{sech}^2\Big(\frac{t}{\alpha-1}\Big)\Big)$$

$$=i\omega\mathscr{F}(f)\mathscr{F}\Big(\operatorname{sech}^2\Big(\frac{t}{\alpha-1}\Big)\Big)=i\omega|\alpha-1|\mathscr{F}(f)\mathscr{F}(\operatorname{sech}^2)((\alpha-1)\omega)$$

$$=-\omega^2|\alpha-1|(\alpha-1)\mathscr{F}(f)\mathscr{F}(\tanh(t))((\alpha-1)\omega).$$

To finish the proof, we recall the result

$$\mathscr{F}(\tanh(t))=i\omega\sqrt{\frac{\pi}{2}}\operatorname{csch}\Big(\frac{\pi\omega}{2}\Big).$$

□

Theorem 2.8. Suppose that f is differentiable and $0<\alpha<1$. Then

$$\int_a^t f'(s)e^{-(\frac{s-t}{1-\alpha})^2}ds\leq C_1(\alpha)^{-1}\,{}^{\sigma}D_a^{\alpha}(f(t))\leq\int_a^t\frac{(1-\alpha)^2f'(s)}{(1-\alpha)^2+(s-t)^2}ds\leq f(t)-f(a).$$

Proof. Using the inequality

$$\cosh x\leq e^{\frac{x^2}{2}},$$

we have

$$e^{-\frac{1}{2}(\frac{s-t}{1-\alpha})^2}\leq\operatorname{sech}\Big(\frac{s-t}{1-\alpha}\Big),$$

which results in the leftmost inequality. Noticing that $\cosh^2 x\geq 1+x^2$, we have that

$$\operatorname{sech}^2\Big(\frac{s-t}{1-\alpha}\Big)\leq\frac{(1-\alpha)^2}{(1-\alpha)^2+(s-t)^2}\leq 1,$$

finishing the last three inequalities.

□

Theorem 2.9. The problem

$${}^{\sigma}D_a^{\alpha}(f(t))=G(t),\ G(0)=0$$

has the solution

$$f(t)=\frac{g(t)}{C_1(\alpha)}+f(0),$$

where $G(t) = \int_0^t g(s)ds$.

Proof. Differentiating the differential equation above, the problem above reduces to

$$C_1(\alpha)f'(t) = g'(t),$$

which can be integrated to obtain the result.

□

Theorem 2.10. Let $0 < \alpha < 1$ and let $g : (a, b) \times \mathbb{R}^2$ be a continuous function such that there exists a constant $C_0 > 0$ satisfying

$$|g(t, x_1, y_1) - g(t, x_2, y_2)| \leq C_0(|x_1 - x_2| + |y_1 - y_2|)$$

for all $t \in (a, b)$ and $x_1, x_2, y_1, y_2 \in \mathbb{R}$ and $|(\alpha - 1)C(\alpha)C_0| < 1$. Then, the problem

$${}^{\sigma}D_a^{\alpha} f(t) = g(t, f(t), {}^{\sigma}D_a^{\alpha} f(t))$$

has a unique solution.

Proof.

$$|g(t, {}^{\sigma}D_a^{\alpha}(f_1(t))) - g(t, {}^{\sigma}D_a^{\alpha}(f_2(t))|$$

$$\leq |(\alpha - 1)C_1(\alpha) \tanh\left(\frac{a - t}{1 - \alpha}\right)|f_1 - f_2|$$

$$\leq |(\alpha - 1)C_1(\alpha)C_0||f_1 - f_2|.$$

Since $(\alpha - 1)C(\alpha)C_0 < 1$, the map $F : H^1(a, b) \to H^1(a, b)$ defined by

$$C_1(\alpha)^{-1} g(t, {}^{\sigma}D_a^{\alpha}(f_1(t)))$$

is a contraction. By the Banach fixed-point theorem, it has a unique fixed point, finishing the proof.

□

We note that this result is advantageous in that the analogous existence and uniqueness result as in fractional differential systems defined by the Caputo derivative is highly dependent on initial conditions imposed on the primary function of interest and its classical derivatives [4].

We now shift our attention to a gradient descent method. Suppose that $f(x)$ has a bounded derivative and unique critical point t^* such that $f'(t^*) = 0$. For $a \leq t \leq b$, $0 < \alpha < 1$, define the scalar left sigmoidal fractional gradient descent method by

$$t_{k+1} = t_k - \mu {}^{\sigma}D_{t_{k-1}}^{\alpha} f(t_k). \tag{14}$$

where $0 < \mu < 1$ is the learning rate.

Theorem 2.11 (Fractional Gradient Descent). Let f be as in (14). Then, the left-sigmoidal fractional-order gradient method (14) converges to the true critical point t^*.

Proof. Denote the Lipschitz constant of f by L. For $k \geq N$,

$$|t_k - t_{k+1}| = \mu\, {}^{\sigma}D^{\alpha}_{t_{K-1}} f(t_k) = C_1(\alpha)\mu \left| \int_{t_{k-1}}^{t_k} f'(s)\operatorname{sech}^2\left(\frac{s-t_k}{1-\alpha}\right) ds \right|$$

$$\leq C_1(\alpha)\mu L|\alpha - 1| \tanh\left(\frac{s-t_k}{1-\alpha}\right) \leq C_1(\alpha)\mu L|t_k - t_{k-1}|.$$

Repeating this process, it follows that the t_k form a Cauchy sequence, guaranteeing convergence. To show that the sequence converges to the critical point, suppose for contradiction that the sequence $(t_k)_{k=0}^{\infty}$ converges to a point $\hat{t} \neq t^*$. Then, for every $\epsilon > 0$, there exists $N \in \mathbb{N}$ such that for all $k \geq N$, $|f'(t_k)| > 0$ and

$$|t_{k-1} - \hat{t}| < \epsilon < |t^* - \hat{t}|.$$

As a consequence of (14) and Theorem 2.8, we have

$$|t_{k+1} - t_k| = C_1(\alpha)\mu \left| \int_{t_{k-1}}^{t_k} f'(s)\operatorname{sech}^2\left(\frac{s-a}{1-\alpha}\right) ds \right| \geq C_1(\alpha)\mu \inf_{k>N} \int_{t_{k-1}}^{t_k} f'(s) e^{-(\frac{s-t}{1-\alpha})^2} ds$$

$$\geq C_1(\alpha)\mu \inf_{k>N} |f'(t_{k-1})| \int_{t_{k-1}}^{t_k} 1 - \left(\frac{s-t}{1-\alpha}\right)^2 ds \geq M_1|t_k - t_{k-1}|\left(1 + \frac{|t_k - t_{k-1}|}{(1-\alpha)^3}\right) \geq M_1 M_2 |t_k - t_{k-1}|^{\frac{3}{2}},$$

where

$$M_1 = C_1(\alpha)\mu \inf_{k>N} |f'(t_{k-1})|\,, M_2 \leq \frac{1}{\sqrt{3(1-\alpha)^3}}.$$

On the other hand, we have the inequality

$$|t_{k+1} - t_{k-1}| \leq |t_{k+1} - t^*| + |t^* - t_{k-1}| < 2\epsilon.$$

Choosing $\epsilon < \frac{1}{2(M_1M_2)^2}$ yields $M_1M_2 > |t_{k+1} - t_k|^{\frac{-1}{2}}$, which implies that $|t_{k+1} - t_k| > |t_k - t_{k-1}|$, contradicting the assumption that the sequence (t_k) is convergent.

□

3. Conclusion

In this paper, we defined a new sigmoidal fractional derivative, which is compatible with certain weakly differentiable functions. We showed that this fractional derivative satisfies some forms of classical properties and is compatible with the ℓ_1 norm by a sigmoidal approximation. For further research, we will investigate this operator in optimization and machine learning. We note that the left-sigmoidal fractional derivative can be applied in the context of gradient descent, which has applications in optimization and machine learning [7, 8]. Recently, backpropagation and convolution neural networks have been studied in the context of fractional derivatives, typically of the Caputo-type are being used for gradient descent. This idea is still novel and needs to see improvements. For example, the gradient descent method has been handled by Sheng et al. [20, 21], Wang et al. [19], Wei et al. [9] and Bao et al [16]. These methods are still early in development. The following topics still need to be fully addressed: convergence to an extreme point, extending the available range of fractional order, more complicated neural networks, loss function compatibility and the usage of the chain rule.

Conflict of interest

The authors declare that there is no conflict of interest.

References

1. R. Gorenflo, F. Mainardi, 223–276. *Fractional calculus: integral and differential equations of fractional order*. In: Fractals and Fractional Calculus in Continuum Mechanics, Springer Verlag, Wien and New York 1997.
2. A. N. Kochubei, *General fractional calculus, evolution equations, and renewal processes*, Integr. Equat. Oper. Th., **71** (2011), 583–600.
3. V. Kiryakova, *Generalised Fractional Calculus and Applications*, Pitman Research Notes in Mathematics, CRC Press, 1993.
4. I. Podlubny, *Fractional Differential Equations*, Academic Press, New York, 2009.
5. M. Caputo, *Elasticità e Dissipazione*, Zanichelli, Bologna, 1965.
6. E. C. de Oliveira, J. A. T. Machado, *A review of definitions for fractional derivatives and integral*, Math. Prob. Ing., **2014** (2014), 238459.
7. J. S. Zeng and W. T. Yin, *On nonconvex decentralized gradient descent,* IEEE T. Signal Proces., **66** (2018), 2834–2848.
8. Y. LeCun, Y. Bengio, and G. Hinton, *Deep learning*, Nature, **521** (2015), 436–444.
9. Y. Wei, Y. Kang, W. Yin, et al. *Design of generalized fractional order gradient descent method*, preprint, 2018.
10. Y. F. Pu, G. L. Zhou, Y. Zhang, et al. *Fractional Extreme Value Adaptive Training Method: Fractional Steepest Descent Approach*, IEEE T. Neur. Net. Lear., **26** (2013), 653–662.
11. M. Caputo, M. Fabrizio, *A new Definition of Fractional Derivative without Singular Kernel*, Progr. Fract. Differ. Appl., **1** (2015), 1–13.
12. E. C. de Oliveira, S. Jarosz, J. Vaz Jr., *Fractional Calculus via Laplace Transform and its Application in Relaxation Processes*, Commun. Nonlinear Sci., **69** (2019), 58–72.
13. V. E. Tarasov, *No nonlocality. no fractional derivative*, Commun. Nonlinear Sci., **62** (2018), 15–163.
14. D. Baleanu, A. Mousalou, S. Rezapour, (2018). *The extended fractional Caputo–Fabrizio derivative of order* $0 \leq \sigma$, Adv. Differ. Equ-NY, **2018** (2018), 255.
15. J. Losada, J. J. Nieto, *Properties of a new fractional derivative without singular Kernel*, Prog. Fract. Differ. Appl., **1** (2015), 87–92.
16. C. Bao, Y. PU, Y. Zhang, *Fractional-Order Deep Backpropagation Neural Network*, Comput. Intel. Neurosc., **2018** (2018), 1–10.
17. J. Wang, G. Yang, B. Zhang, et al. *Convergence Analysis of Caputo-Type Fractional Order Complex-Valued Neural Networks*, IEEE Access, **5** (2017), 14560–14571.

18. S. Cheng, Y. Wei, Y. Chen, et al. *An innovative fractional order LMS based on variable initial value and gradient order*, Signal Process., **133** (2017), 260–269.
19. J. Wang, Y. Wen, Y. Gou, et al. (2017). *Fractional-order gradient descent learning of BP neural networks with Caputo derivative*, Neural Networks, **89** (2017), 19–30.
20. D. Sheng, Y. Wei, Y. Chen, et al. *Convolutional neural networks with fractional order gradient method*, Neurocomputing, 2019.
21. Y. Q. Chen, Q. Gao, Y. H. Wei, et al. *Study on fractional order gradient methods*, Appl. Math. Comput., **314** (2017), 310–321.
22. U. N. Katugampola, *A New Fractional Derivative with Classical Properties*, arXiv preprint arXiv:1410.6535, 2014.
23. R. T. Alqahtani, *Fixed-point theorem for Caputo–Fabrizio fractional Nagumo equation with nonlinear diffusion and convection*, J. Nonlinear Sci. Appl., **9** (2016), 1991–1999.
24. A. Alsaedi, D. Baleanu, S. Etemad, et al. *On Coupled Systems of Time-Fractional Differential Problems by Using a New Fractional Derivative*, **2016** (2016), 1–8.
25. A. Atangana, *On the new fractional derivative and application to nonlinear Fisher's reaction–diffusion equation*, Appl. Math. Comput., **273** (2016), 948–956.
26. S. M. Aydogan, D. Baleanu, A. Mousalou, et al. *On approximate solutions for two higher-order Caputo-Fabrizio fractional integro-differential equations*, Adv. Differ. Equ-NY, **2017** (2017), 221.
27. A. Atangana, J. F. Gómez-Aguilar, *Decolonisation of fractional calculus rules: Breaking commutativity and associativity to capture more natural phenomena*, Eur. Phys. J. Plus, **133** (2018), 166.
28. T. M. Atanacković, S. Pilipović, D. Zorica, *Properties of the Caputo-Fabrizio fractional derivative and its distributional settings*, Fract. Calc. Appl. Anal., **21** (2018), 29–44.
29. V. N. Krutikov, L. A. Kazakovtsev, G. Shkaberina, et al. *New method of training two-layer sigmoid neural networks using regularization*, IOP Conference Series: Materials Science and Engineering, **537** (2019), 042055.
30. R. Rakkiyappan, R. Sivaranjani, G. Velmurugan, et al. *Analysis of global $o(t^{-\alpha})$ stability and global asymptotical periodicity for a class of fractional- order complex-valued neural networks with time varying delays*, Neural Networks, **77** (2016), 51–69.
31. X. Chen, *Application of fractional calculus in bp neural networks*, (Ph.D. thesis), Nanjing, Jiangsu: Nanjing Forestry University, 2013.
32. A. H.Zemanian, *Distribution theory and transform analysis*, New York: Dover Publ. Inc., 1987.
33. M. Caputo, M. Fabrizio, *Applications of new time and spatial fractional derivatives with exponential kernels*, Progr. Fract. Differ. Appl., **2** (2016), 1–11.

AIMS Press

© 2020 the Author(s), licensee AIMS Press. This is an open access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0)