



Article

Fractional Gradient Methods via ψ -Hilfer Derivative

Nelson Vieira ^{1,*} , M. Manuela Rodrigues ¹ and Milton Ferreira ^{2,3}

- ¹ CIDMA — Center for Research and Development in Mathematics and Applications, Department of Mathematics, University of Aveiro, Campus Universitário de Santiago, 3810-193 Aveiro, Portugal; mrodrigues@ua.pt
- ² School of Technology and Management, Polytechnic of Leiria, 2411-901 Leiria, Portugal; milton.ferreira@ipleiria.pt
- ³ CIDMA—Center for Research and Development in Mathematics and Applications, University of Aveiro, 3810-193 Aveiro, Portugal
- * Correspondence: nloureirovieira@gmail.com

Abstract: Motivated by the increase in practical applications of fractional calculus, we study the classical gradient method under the perspective of the ψ -Hilfer derivative. This allows us to cover several definitions of fractional derivatives that are found in the literature in our study. The convergence of the ψ -Hilfer continuous fractional gradient method was studied both for strongly and non-strongly convex cases. Using a series representation of the target function, we developed an algorithm for the ψ -Hilfer fractional order gradient method. The numerical method obtained by truncating higher-order terms was tested and analyzed using benchmark functions. Considering variable order differentiation and step size optimization, the ψ -Hilfer fractional gradient method showed better results in terms of speed and accuracy. Our results generalize previous works in the literature.

Keywords: fractional calculus; ψ -Hilfer fractional derivative; fractional gradient method; optimization

MSC: 26A33; 65K99; 35R11; 49M99



Citation: Vieira, N.; Rodrigues, M.M.; Ferreira, M. Fractional Gradient Methods via ψ -Hilfer Derivative. *Fractal Fract.* **2023**, *7*, 275. <https://doi.org/10.3390/fractalfract7030275>

Academic Editor: Rodica Luca

Received: 30 January 2023

Revised: 9 March 2023

Accepted: 14 March 2023

Published: 21 March 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The gradient descent method is a classical convex optimization method. It is widely used in many areas of computer science, such as in image processing [1,2], machine learning [3–5], and control systems [6]. Its use on a large scale is essentially due to its intuitive structure, ease of implementation, and accuracy. In recent years, there has been an increase in interest in the application of fractional calculus techniques to develop and implement fractional gradient methods (FGM). The first work dealing with such methods is in [2,7], which addresses problems in the fields of signal processing and adaptive learning. The design of fractional least mean squares algorithms is another example of the application of FGM [8–10]. Recently, some applications of FGM have focused on artificial intelligence subjects such as machine learning, deep learning, and neural networks (see [11–13] and references therein).

Replacing the first-order integer derivative with a fractional derivative in a gradient can improve its convergence, because long-term information can be included. However, there are some convergence issues in the numerical implementation of the FGM because the real extreme value of the target function is not always the same as the fractional extreme value.

In [14], the authors propose a new FGM to overcome this problem, considering an iterative update of the lower limit of integration in the fractional derivative to shorten the memory characteristic presented in the fractional derivative and truncating the higher order terms of the series expansion associated with the target function. Afterward, Wei et al. [15] designed another method involving variable fractional order to solve the convergence problem.

In the field of fractional calculation, several definitions of fractional and derivative integrals varying in their kernel size can be found. This diversity allows certain problems to be tackled with specific fractional operators. To establish a general operator, a ψ -fractional integral operator with respect to function ψ was proposed in [16,17], where the kernel depends on a function ψ with specific properties. To incorporate as many fractional derivative definitions as possible into a single formulation, the concept of a fractional derivative of a function with respect to the function ψ was introduced. In 2017, Almeida [18] proposed the ψ -Caputo fractional derivative and studied its main properties. A similar approach can be used to define the ψ -Riemann–Liouville fractional derivative. In 2018, Sousa and Oliveira [19] unified both definitions using Hilfer’s concept and introduced the ψ -Hilfer fractional derivative. This approach offers the flexibility of choosing the differentiation type, as Hilfer’s definition interpolates smoothly between fractional derivatives of Caputo and Riemann–Liouville types. Additionally, by choosing the function ψ , we obtain well-known fractional derivatives, such as Caputo, Riemann–Liouville, Hadamard, Katugampola, Chen, Jumarie, Prabhakar, Erdélyi-Kober, and Weyl, among others (see Section 5 in [19]).

The aim of this work is to propose a FGM with a ψ -Hilfer fractional derivative. Using this type of general derivative allows us to deal with several fractional derivatives in the literature at the same time. It also allows us to study cases where the target function is a composition of functions. In the first section, we show some auxiliary results concerning the chain rule and solutions of some fractional partial differential equations to study the convergence of the continuous ψ -fractional gradient method for strongly and non-strongly convex target functions. In the second section, we introduce and implement numerical algorithms for the ψ -Hilfer FGM in one- and two-dimensional cases, generalizing the ideas presented in [14,15]. The proposed algorithms were tested using benchmark functions. The numerical results had better performance compared with the classical gradient in terms of accuracy and number of iterations.

In summary, this paper is organized as follows: in Section 2, we recall some basic concepts about the ψ -Hilfer derivative and the two-parameter Mittag–Leffler function. We present some auxiliary results in Section 3, which are then used to analyze the continuous gradient method for strongly and non-strongly convex target functions in Section 4. In the last section of the paper, we design and implement numerical algorithms for the ψ -Hilfer FGM by replacing the lower limit of the fractional integral with the last iterate and by using the variable order of differentiation with the optimization of the step size in each iteration. The convergence, accuracy, and speed of the algorithms are analyzed using different examples.

2. General Fractional Derivatives and Special Functions

In this section, we recall some concepts related to fractional integrals and derivatives of a function with respect to another function ψ (for more details, see [16,18,19]).

Definition 1. (cf. [19], Def. 4) Let $[a, b]$ be a finite or infinite interval on the real line \mathbb{R} and $\alpha > 0$. In addition, let ψ be an increasing and positive monotone function on (a, b) . The left Riemann–Liouville fractional integral of a function f with respect to another function ψ on $[a, b]$ is defined by

$$\left(I_{a+}^{\alpha; \psi} f\right)(t) = \frac{1}{\Gamma(\alpha)} \int_a^t \psi'(w) (\psi(t) - \psi(w))^{\alpha-1} f(w) dw, \quad t > a. \quad (1)$$

Now, we introduce the definition of the so-called ψ -Hilfer fractional derivative of a function f with respect to another function.

Definition 2. (cf. [19], Def. 7) Let $\alpha > 0$, $m = \lfloor \alpha \rfloor + 1$, and $I = [a, b]$ be a finite or infinite interval on the real line and $f, \psi \in C^m[a, b]$ of two functions such that ψ is a positive monotone

increasing function and $\psi'(t) \neq 0$, for all $t \in I$. The ψ -Hilfer left fractional derivative ${}^H\mathbb{D}_{t,a^+}^{\alpha,\mu;\psi}$ of order α and type $\mu \in [0, 1]$ is defined by

$$\left({}^H D_{a^+}^{\alpha,\mu;\psi} f\right)(t) = I_{a^+}^{\mu(m-\alpha);\psi} \left(\frac{1}{\psi'(t)} \frac{d}{dt} \right)^m I_{a^+}^{(1-\mu)(m-\alpha);\psi} f(t). \quad (2)$$

We observe that when $\mu = 0$, we recover the left fractional derivative ψ -Riemann–Liouville (see Definition 5 in [19]) and when $\mu = 1$, we obtain the left ψ -Caputo fractional derivative (see Definition 6 in [19]). The following list shows some fractional derivatives that are encompassed in Definition 2 for specific choices of the function ψ and parameter μ :

- Riemann–Liouville: $\psi(t) = t$, $I = \mathbb{R}^+$, and $\mu = 0$;
- Caputo: $\psi(t) = t$, $I = \mathbb{R}^+$, and $\mu = 1$;
- Katugampola: $\psi(t) = t^\rho$, with $\rho \in \mathbb{R}^+$, $I = \mathbb{R}^+$, and $\mu = 0$;
- Caputo–Katugampola: $\psi(t) = t^\rho$, with $\rho \in \mathbb{R}^+$, $I = \mathbb{R}^+$, and $\mu = 1$;
- Hadamard: $\psi(t) = \ln(t)$, $I =]1, +\infty[$, and $\mu = 0$;
- Caputo–Hadamard: $\psi(t) = \ln(t)$, $I =]1, +\infty[$, and $\mu = 1$.

For a more complete list, please refer to Section 5 in [19]. By considering partial fractional integrals and derivatives, previous definitions can be defined for higher dimensions (see Chapter 5 in [16]). Furthermore, the ψ -Hilfer fractional derivative of an n -dimensional vector function $f(t) = (f_1(t), \dots, f_n(t))$ is defined component-wise as

$${}^H D_{a^+}^{\alpha,\mu;\psi} f(t) = \left({}^H D_{a^+}^{\alpha,\mu;\psi} f_1(t), \dots, {}^H D_{a^+}^{\alpha,\mu;\psi} f_n(t) \right). \quad (3)$$

Next, we present some technical results related to previously introduced operators.

Theorem 1. (cf. [19], Thm. 5) If $f \in C^m[a, b]$, $\alpha > 0$, $m = \lfloor \alpha \rfloor + 1$, and $\mu \in [0, 1]$, then

$$I_{a^+}^{\alpha;\psi} {}^H D_{a^+}^{\alpha,\mu;\psi} f(t) = f(t) - \sum_{k=1}^m \frac{(\psi(t) - \psi(a))^{\gamma-k}}{\Gamma(\gamma-k+1)} f_\psi^{[m-k]} I_{a^+}^{(1-\mu)(m-\alpha);\psi} f(a), \quad (4)$$

where $\gamma = \alpha + \mu(k - \alpha)$ and $f_\psi^{[m]} f(t) = \left(\frac{1}{\psi'(t)} \frac{d}{dt} \right)^m f(t)$.

Lemma 1. (cf. [19], Lem. 5) Given $\delta \in \mathbb{R}$, consider the function $f(t) = (\psi(t) - \psi(a))^{\delta-1}$, where $\delta > m$. Then, for $m = \lfloor \alpha \rfloor + 1$ and $\mu \in [0, 1]$, we have

$${}^H D_{a^+}^{\alpha,\mu;\psi} f(x) = \frac{\Gamma(\delta)}{\Gamma(\delta - \alpha)} (\psi(t) - \psi(a))^{\delta-\alpha-1}. \quad (5)$$

Some results of the paper are given in terms of the two-parameter Mittag–Leffler function, which is defined by the following power series (see [20])

$$E_{\beta_1, \beta_2}(z) = \sum_{n=0}^{+\infty} \frac{z^n}{\Gamma(\beta_1 n + \beta_2)}, \quad \operatorname{Re}(\beta_1) > 0, \quad \beta_2 \in \mathbb{C}, \quad z \in \mathbb{C}.$$

For $z = -x$, with $x \in \mathbb{R}^+$, $0 < \beta_1 < 2$, and $\beta_2 \in \mathbb{C}$, the two-parameter Mittag–Leffler function has the following asymptotic expansion (see Equation (4.7.36) in [20]):

$$E_{\beta_1, \beta_2}(-x) = - \sum_{k=1}^p \frac{(-x)^{-k}}{\Gamma(\beta_2 - \beta_1 k)} + O(|x|^{-1-p}), \quad p \in \mathbb{N}, \quad x \rightarrow +\infty. \quad (6)$$

3. Auxiliary Results

In this section, we present some auxiliary results needed for our work. These extend some results presented in [21] to the ψ -Hilfer derivative of arbitrary type $\mu \in [0, 1]$.

We start by presenting a representation formula for the solution of a Cauchy problem involving the ψ -Hilfer derivative. Let us consider $h : \mathbb{R}_0^+ \times \mathbb{R}^n \rightarrow \mathbb{R}^n$ and $f : \mathbb{R} \rightarrow \mathbb{R}^n$. We obtain the following results.

Proposition 1. Let $\alpha \in [0, 1]$ and $\mu \in [0, 1]$. A continuous function f is a solution of the problem

$$\begin{cases} {}^H D_{a^+}^{\alpha, \mu; \psi} f(t) = h(t, f(t)), & t \geq a, \\ I_{a^+}^{(1-\alpha)(1-\mu); \psi} f(a^+) = f_a \end{cases} \quad (7)$$

if and only if f is given by

$$\begin{aligned} f(t) &= \frac{f_a}{\Gamma(\alpha + \mu(1-\alpha))} (\psi(t) - \psi(a))^{\alpha + \mu(1-\alpha) - 1} \\ &+ \frac{1}{\Gamma(\alpha)} \int_a^t (\psi(t) - \psi(s))^{\alpha-1} \psi'(s) h(s, f(s)) ds, \quad t \geq a. \end{aligned} \quad (8)$$

Proof. Applying $I_{a^+}^{\alpha; \psi}$ to both sides of the fractional differential equation in (7), and taking (4) with $m = 1$, we have

$$I_{a^+}^{\alpha; \psi} {}^H D_{a^+}^{\alpha, \mu; \psi} f(t) = f(t) - \frac{f_a}{\Gamma(\alpha + \mu(1-\alpha))} (\psi(t) - \psi(a))^{\alpha + \mu(1-\alpha) - 1}$$

which is equivalent to

$$f(t) = \frac{f_a}{\Gamma(\alpha + \mu(1-\alpha))} (\psi(t) - \psi(a))^{\alpha + \mu(1-\alpha) - 1} + I_{a^+}^{\alpha; \psi} {}^H D_{a^+}^{\alpha, \mu; \psi} f(t).$$

From (1) and (7), we obtain these results. \square

Now, we present some results concerning the fractional derivative of a composite function. Let us consider $g : \mathbb{R}^n \rightarrow \mathbb{R}$, $f : \mathbb{R} \rightarrow \mathbb{R}^n$, ∇ as the classical gradient operator, and the ψ -Hilfer fractional derivative of an n -dimensional vector function as (3).

Theorem 2. Let $\alpha \in [0, 1]$, $\mu \in [0, 1]$, $a \geq 0$, and the function ψ to be in the conditions of Definition 2. For $t > a$, let us define the function ζ_t by setting

$$\zeta_t(s) = g(f(s)) - g(f(t)) - \langle (\nabla g)(f(t)), f(s) - f(t) \rangle, \quad (9)$$

where

$$(\nabla g)(f(t)) = \left(\frac{\partial g}{\partial x_1}(f(t)), \dots, \frac{\partial g}{\partial x_n}(f(t)) \right).$$

The following identity holds

$$\begin{aligned} &\Gamma(1-\alpha) \left({}^H D_{a^+}^{\alpha, \mu; \psi} [g(f(t))] - \langle (\nabla g)(f(t)), {}^H D_{a^+}^{\alpha, \mu; \psi} f(t) \rangle \right) \\ &= [\psi(t) - \psi(a)]^{-\alpha} [g(f(t)) - \langle (\nabla g)(f(t)), f(t) \rangle] \\ &\quad - \alpha \int_a^t (\psi(t) - \psi(s))^{-\alpha-1} \psi'(s) \zeta_t(s) ds. \end{aligned} \quad (10)$$

Proof. By the Newton–Leibniz formula, for each component of the function f , one has

$$f_i(t) = f_i(a) + \int_a^t f'_i(s) ds = f_i(a) + I_{a^+}^{1;\psi} \left(\frac{f'_i}{\psi'} \right) (t), \quad i = 1, \dots, n.$$

From (3), for $i = 1, \dots, n$, we have

$$\begin{aligned} {}^H D_{a^+}^{\alpha, \mu; \psi} f_i(t) &= {}^H D_{a^+}^{\alpha, \mu; \psi} [f_i(a)](t) + {}^H D_{a^+}^{\alpha, \mu; \psi} \circ I_{a^+}^{1;\psi} \left(\frac{f'_i}{\psi'} \right) (t) \\ &= I_{a^+}^{\mu(1-\alpha); \psi} \left[\frac{1}{\psi'(t)} \frac{d}{dt} \right] I_{a^+}^{(1-\mu)(1-\alpha); \psi} [f_i(a)](t) \\ &\quad + I_{a^+}^{\mu(1-\alpha); \psi} \left[\frac{1}{\psi'(t)} \frac{d}{dt} \right] I_{a^+}^{(1-\mu)(1-\alpha); \psi} I_{a^+}^{1;\psi} \left(\frac{f'_i}{\psi'} \right) (t) \\ &= f_i(a) I_{a^+}^{\mu(1-\alpha); \psi} \left[\frac{1}{\psi'(t)} \frac{d}{dt} \right] I_{a^+}^{(1-\mu)(1-\alpha); \psi} [1](t) \\ &\quad + I_{a^+}^{\mu(1-\alpha); \psi} \left[\frac{1}{\psi'(t)} \frac{d}{dt} \right] I_{a^+}^{(1-\mu)(1-\alpha)+1; \psi} \left(\frac{f'_i}{\psi'} \right) (t). \end{aligned} \quad (11)$$

Taking (1) into account, the first term in (11) is

$$\begin{aligned} &f_i(a) I_{a^+}^{\mu(1-\alpha); \psi} \left[\frac{1}{\psi'(t)} \frac{d}{dt} \right] I_{a^+}^{(1-\mu)(1-\alpha); \psi} [1](t) \\ &= f_i(a) I_{a^+}^{\mu(1-\alpha); \psi} \left[\frac{1}{\psi'(t)} \frac{d}{dt} \right] \frac{1}{\Gamma((1-\mu)(1-\alpha))} \int_a^t (\psi(t) - \psi(s))^{(1-\mu)(1-\alpha)-1} \psi'(s) ds \\ &= f_i(t) I_{a^+}^{\mu(1-\alpha); \psi} \left[\frac{1}{\psi'(t)} \frac{d}{dt} \right] \left[\frac{(\psi(t) - \psi(a))^{(1-\mu)(1-\alpha)}}{\Gamma((1-\mu)(1-\alpha) + 1)} \right] \\ &= \frac{f_i(a)}{\Gamma(\mu(1-\alpha))} \int_a^t (\psi(t) - \psi(s))^{\mu(1-\alpha)-1} \frac{(\psi(s) - \psi(a))^{(1-\mu)(1-\alpha)-1}}{\Gamma((1-\mu)(1-\alpha))} \psi'(s) ds \\ &= \frac{f_i(a)}{\Gamma(1-\alpha)} (\psi(t) - \psi(a))^{-\alpha}. \end{aligned} \quad (12)$$

For the second term in (11), taking (1) and the Leibniz rule for differentiation under the integral sign into account, we obtain

$$\begin{aligned} &I_{a^+}^{\mu(1-\alpha); \psi} \left[\frac{1}{\psi'(t)} \frac{d}{dt} \right] I_{a^+}^{(1-\mu)(1-\alpha)+1; \psi} \left(\frac{f'_i}{\psi'} \right) (t) \\ &= I_{a^+}^{\mu(1-\alpha); \psi} \left[\frac{1}{\psi'(t)} \frac{d}{dt} \right] \frac{1}{\Gamma((1-\mu)(1-\alpha) + 1)} \int_a^t (\psi(t) - \psi(s))^{(1-\mu)(1-\alpha)} f'_i(s) ds \\ &= I_{a^+}^{\mu(1-\alpha); \psi} \frac{1}{\Gamma((1-\mu)(1-\alpha))} \int_a^t (\psi(t) - \psi(s))^{(1-\mu)(1-\alpha)-1} f'_i(s) ds \\ &= I_{a^+}^{\mu(1-\alpha); \psi} I_{a^+}^{(1-\mu)(1-\alpha); \psi} \left(\frac{f'_i}{\psi'} \right) (t) \\ &= \frac{1}{\Gamma(1-\alpha)} \int_a^t (\psi(t) - \psi(s))^{-\alpha} f'_i(s) ds. \end{aligned} \quad (13)$$

From (12) and (13), expression (11) simplifies to

$$\begin{aligned} & {}^H D_{a^+}^{\alpha, \mu; \psi} f_i(t) \\ &= \frac{1}{\Gamma(1-\alpha)} \left[f_i(a)(\psi(t) - \psi(a))^{-\alpha} + \int_a^t (\psi(t) - \psi(s))^{-\alpha} f'_i(s) ds \right], \quad i = 1, \dots, n \end{aligned}$$

and therefore,

$${}^H D_{a^+}^{\alpha, \mu; \psi} f(t) = \frac{1}{\Gamma(1-\alpha)} \left[f(a)(\psi(t) - \psi(a))^{-\alpha} + \int_a^t (\psi(t) - \psi(s))^{-\alpha} f'(s) ds \right]. \quad (14)$$

Hence, we can write

$$\begin{aligned} & \Gamma(1-\alpha) \left({}^H D_{a^+}^{\alpha, \mu; \psi} [g(f(t))] - \langle (\nabla g)(f(t)), {}^H D_{a^+}^{\alpha, \mu; \psi} f(t) \rangle \right) \\ &= [\psi(t) - \psi(a)]^{-\alpha} [g(f(a)) - \langle (\nabla g)(f(t)), f(a) \rangle] \\ &\quad + \int_a^t (\psi(t) - \psi(s))^{-\alpha} \langle (\nabla g)(f(s)) - (\nabla g)(f(t)), f'(s) \rangle ds \\ &= [\psi(t) - \psi(a)]^{-\alpha} [g(f(a)) - \langle (\nabla g)(f(t)), f(a) \rangle] \\ &\quad + \int_a^t (\psi(t) - \psi(s))^{-\alpha} d\zeta_t(s) \end{aligned} \quad (15)$$

which implies, by integrating by parts, that

$$\begin{aligned} & \Gamma(1-\alpha) \left({}^H D_{a^+}^{\alpha, \mu; \psi} [g(f(t))] - \langle (\nabla g)(f(t)), {}^H D_{a^+}^{\alpha, \mu; \psi} f(t) \rangle \right) \\ &= [\psi(t) - \psi(a)]^{-\alpha} [g(f(a)) - \langle (\nabla g)(f(t)), f(a) \rangle] \\ &\quad + \lim_{s \rightarrow t} [\psi(t) - \psi(s)]^{-\alpha} \zeta_t(s) - [\psi(t) - \psi(a)]^{-\alpha} \zeta_t(a) \\ &\quad - \alpha \int_a^t (\psi(t) - \psi(s))^{-\alpha-1} \psi'(s) \zeta_t(s) ds. \end{aligned} \quad (16)$$

As $\alpha \in [0, 1]$, we have by L'Hôpital's rule that

$$\lim_{s \rightarrow t} [\psi(t) - \psi(s)]^{-\alpha} \zeta_t(s) = \lim_{s \rightarrow t} \frac{\zeta_t(s)}{[\psi(t) - \psi(s)]^\alpha} = \lim_{s \rightarrow t} \frac{\zeta'_t(s) [\psi(t) - \psi(s)]^{1-\alpha}}{-\alpha \psi'(s)} = 0. \quad (17)$$

Finally, from (17) and (9), we obtain the following from (16)

$$\begin{aligned} & \Gamma(1-\alpha) \left({}^H D_{a^+}^{\alpha, \mu; \psi} [g(f(t))] - \langle (\nabla g)(f(t)), {}^H D_{a^+}^{\alpha, \mu; \psi} f(t) \rangle \right) \\ &= [\psi(t) - \psi(a)]^{-\alpha} [\zeta_t(a) - g(f(t)) - \langle (\nabla g)(f(t)), f(t) \rangle - \zeta_t(a)] \\ &\quad - \alpha \int_a^t (\psi(t) - \psi(s))^{-\alpha-1} \psi'(s) \zeta_t(s) ds, \end{aligned}$$

which gives the desired result. \square

Corollary 1. Let $\alpha \in [0, 1]$, $\mu \in [0, 1]$, $a \geq 0$, and the function ψ have the conditions of Definition 2. If $g : \mathbb{R}^n \rightarrow \mathbb{R}$ is of class C^1 and convex, i.e.,

$$g(x) \geq g(y) + \langle \nabla g(x), x - y \rangle, \quad \text{for all } x, y \in \mathbb{R}^n, \quad (18)$$

then

$${}^H D_{a^+}^{\alpha, \mu; \psi} g(f(t)) \leq \left\langle (\nabla g)(f(t)), {}^H D_{a^+}^{\alpha, \mu; \psi} f(t) \right\rangle + \frac{g(0)}{\Gamma(1-\alpha)} (\psi(t) - \psi(a))^{-\alpha}. \quad (19)$$

Proof. From (18), it follows that for $x = 0$,

$$g(f(t)) - \langle (\nabla g)(f(t)), f(t) \rangle \leq g(0) - \langle (\nabla g)(f(t)), 0 \rangle = g(0).$$

On the other hand, based on Theorem 2, we have the following for $x = 0$

$$\begin{aligned} & g(f(t)) - \langle (\nabla g)(f(t)), f(t) \rangle \\ &= \Gamma(1-\alpha) [\psi(t) - \psi(a)]^\alpha \left({}^H D_{a^+}^{\alpha, \mu; \psi} [g(f(t))] - \left\langle (\nabla g)(f(t)), {}^H D_{a^+}^{\alpha, \mu; \psi} f(t) \right\rangle \right). \end{aligned}$$

Combining the two previous expressions, we obtain our results. \square

If we consider $\mu = 0$, which corresponds to the ψ -Riemann–Liouville case, the previous results reduce to Proposition 3.3 and Corollary 3.4 in [21], respectively. Moreover, the correspondent results for the ψ -Caputo case ($\mu = 1$) are presented in Proposition 3.1 of [21].

Now, we present an auxiliary result involving the two-parameter Mittag–Leffler function.

Proposition 2. Let $\alpha \in [0, 1]$, $\mu \in [0, 1]$, and $a \geq 0$. Moreover, let ψ be in the conditions of Definition 2, such that $\sup\{\psi(t) : t \geq a\} = +\infty$. Then, the following limit holds

$$\lim_{t \rightarrow +\infty} \int_a^t (\psi(t) - \psi(a))^{\alpha-1} \psi'(s) E_{\alpha, \alpha+\mu(1-\alpha)}(-\lambda(\psi(s) - \psi(a))^\alpha) ds = 0.$$

Proof. Taking into account Theorem 5.1 in [22] for the case of a homogeneous equation, the solution of the initial value problem

$$\begin{cases} {}^H D_{a^+}^{\alpha, \mu; \psi} u(t) = -\lambda u(t); \quad \lambda \in \mathbb{R}^+, \alpha \in [0, 1], \mu \in [0, 1], t \geq a \\ I_{a^+}^{1-\alpha-\mu(1-\alpha); \psi} u(a) = u_a \end{cases}$$

is given by

$$u(t) = (\psi(t) - \psi(a))^{\alpha-\mu(1-\alpha)-1} E_{\alpha, \alpha+\mu(1-\alpha)}(-\lambda(\psi(t) - \psi(a))^\alpha).$$

Hence, by Proposition 1, we have

$$\begin{aligned} & E_{\alpha, \alpha+\mu(1-\alpha)}(-\lambda(\psi(t) - \psi(a))^\alpha) \\ &= \frac{u_a}{\Gamma(\alpha + \mu(1-\alpha))} (\psi(t) - \psi(a))^{\alpha+\mu(1-\alpha)-1} \\ &\quad - \frac{\lambda}{\Gamma(\alpha)} \int_a^t (\psi(t) - \psi(s))^{\alpha-1} \psi'(s) E_{\alpha, \alpha+\mu(1-\alpha)}(-\lambda(\psi(s) - \psi(a))^\alpha) ds. \end{aligned}$$

Taking the limit when $t \rightarrow +\infty$ on both sides and considering the asymptotic expansion (6), we conclude that the left-hand side tends to approach zero and the first term of the right-hand side also tends to approach zero. Hence, we obtain

$$0 = 0 - \frac{\lambda}{\Gamma(\alpha)} \int_a^t (\psi(t) - \psi(s))^{\alpha-1} \psi'(s) E_{\alpha, \alpha+\mu(1-\alpha)}(-\lambda(\psi(s) - \psi(a))^\alpha) ds$$

which leads to our results. \square

The case when $\mu = 1$, i.e., the ψ -Caputo case, was already studied in [21] and corresponds to Lemma 3.7.

4. Continuous Gradient Method via the ψ -Hilfer Derivative

Assume that we aim to determine the minimum of a function $f : \mathbb{R}^n \rightarrow \mathbb{R}$. To achieve this, the gradient descent method is used, starting with an initial prediction x_0 of the local minimum and producing a sequence x_0, x_1, x_2, \dots based on the following recurrence relation:

$$x_{k+1} = x_k - \theta_k \nabla f(x_k), \quad (20)$$

where the step size $\theta > 0$ is either constant or varying at each iteration k . The sequence $\{x_k\}_{k=0}^{+\infty}$ generated by the gradient descent method is monotonic, i.e., $f(x_0) > f(x_1) > f(x_2) > \dots$, and is expected to converge to a local minimum of f . Typically, the stopping criterion is in the form $\|\nabla f(x)\| \leq \epsilon$, where $\epsilon > 0$. By expressing (20) as

$$\frac{x_{k+1} - x_k}{\theta_k} = -\nabla f(x_k), \quad (21)$$

we can interpret (21) as the discretization of the initial value problem

$$y'(t) = -(\nabla f)(y(t)), \quad y(0) = y_0 \in \mathbb{R}^n, \quad (22)$$

using the explicit Euler scheme with step size θ_k . The system (22) is known as the *continuous gradient method* (see [21]). Assuming that f is both strongly convex and smooth, the solutions of (20) and (22) converge to the unique stationary point at an exponential rate. In general, if a convergence result is shown for a continuous method, then we can construct various finite difference schemes for the solution of the associated Cauchy problem. Let us now consider the following ψ -fractional version of (22)

$${}^H D_{a^+}^{\alpha, \mu; \psi} z(t) = -\theta (\nabla f)(z(t)), \quad t \geq a, \quad (23)$$

such that $z : \mathbb{R} \rightarrow \mathbb{R}^n$, $f : \mathbb{R}^n \rightarrow \mathbb{R}$, $\alpha \in [0, 1]$, $\mu \in [0, 1]$, and $I_{a^+}^{1-\alpha; \psi} z(a^+) = z_a \in \mathbb{R}^n$, where the last expression is evaluated at the limit $t \rightarrow a^+$. For $y^* \in S(f) = \{z \in \mathbb{R}^n : \nabla f(z) = 0\}$, let us define the following sum of squares error function:

$$\varphi(t) = \frac{1}{2} \|z(t) - y^*\|^2, \quad t \geq a. \quad (24)$$

4.1. The Convex Case

Here, we investigate (23) under the assumption of non-strongly convexity of f .

Theorem 3. Let $\alpha \in [0, 1]$ and $\mu \in [0, 1]$. Suppose that the function $f : \mathbb{R}^n \rightarrow \mathbb{R}$ is of class C^1 and convex, i.e., f satisfies (18). For the ψ -fractional differential Equation (23), with step size θ constant, the solution $z(\cdot)$ converges to y^* with the upper bound

$$\|z(t) - y^*\|^2 \leq C |\psi(t) - \psi(a)|^{-\mu(1-\alpha)}, \quad \text{for all } t \geq a. \quad (25)$$

Proof. Based on Corollary 1 applied to (24) and the fact that f is of class C^1 and convex, we have

$$\begin{aligned} {}^H D_{a^+}^{\alpha, \mu; \psi} \varphi(t) &\leq -\theta \langle z(t) - y^*, (\nabla f)(z(t)) \rangle \\ &\leq \theta [f(y^*(t)) - f(z(t))] \leq 0, \quad \text{for all } t \geq a. \end{aligned}$$

By the properties of ψ (see Definition 2), the previous expression is equivalent to

$$I_{a^+}^{\alpha;\psi} {}^H D_{a^+}^{\alpha,\mu;\psi} \varphi(t) \leq \theta I_{a^+}^{\alpha;\psi} [f(y^*(t)) - f(z(t))] \leq 0.$$

Using the composition rule (4), with $m = 1$, we have

$$\varphi(t) - \frac{(\psi(t) - \psi(a))^{-\mu(1-\alpha)}}{\Gamma(1-\mu(1-\alpha))} I_{a^+}^{(1-\mu)(1-\alpha);\psi} \varphi(a^+) \leq 0.$$

From (24) and considering $C = \frac{1}{\Gamma(1-\mu(1-\alpha))} I_{a^+}^{(1-\mu)(1-\alpha);\psi} \varphi(a^+)$, we obtain our results. \square

If we consider $\mu = 0$ in the previous result, we recover Theorem 4.2 in [21].

4.2. The Strongly Convex Case

Here, we show that under the assumption of strong convexity of the function f , the solution of (23) admits a Mittag–Leffler convergence, which is a general type of exponential convergence to the stationary point. Recall the definition of a strongly convex function.

Definition 3. (cf. [21]) A function $f \in C^1$ is strongly convex with parameter $m_f > 0$ if

$$f(x) \geq f(y) + \langle \nabla f(y), x - y \rangle + \frac{m_f}{2} \|x - y\|^2, \quad \text{for all } x, y \in \mathbb{R}^n,$$

where ∇ stands for the gradient operator.

Theorem 4. Let $\alpha \in [0, 1]$ and $\mu \in [0, 1]$. Suppose that f is of class C^2 and is strongly convex. Considering the ψ -fractional differential Equation (23), where the step size θ is a constant, then the solution $z(\cdot)$ converges to y^* , with the upper bound

$$\|z(t) - y^*\|^2 \leq \varphi_a [\psi(t) - \psi(a)]^{(\alpha-1)(1-\mu)} E_{\alpha, \alpha+\mu(1-\alpha)} \left(-\theta m_f (\psi(t) - \psi(a))^\alpha \right), \quad t \geq a,$$

where $\varphi_a = \frac{1}{2} I_{a^+}^{(1-\mu)(1-\alpha);\psi} \|z(t) - y^*\|^2(a^+)$.

Proof. In Definition 3, if we consider $y = z(t)$ and $x = y^*$, we have for $t \geq a$ and $m_f > 0$ that

$$f(y^*) \geq f(z(t)) - \langle (\nabla f)(z(t)), y^* - z(t) \rangle + \frac{m_f}{2} \|y^* - z(t)\|^2,$$

which is equivalent to

$$\langle (\nabla f)(z(t)), y^* - z(t) \rangle \geq f(z(t)) - f(y^*) + \frac{m_f}{2} \|y^* - z(t)\|^2 \geq \frac{m_f}{2} \|y^* - z(t)\|^2, \quad (26)$$

where the last inequality holds, as $y^* \in \arg(\min_{x \in \mathbb{R}^n} f(x))$. From (23), Corollary 1, and (26), we have

$$\begin{aligned} {}^H D_{a^+}^{\alpha,\mu;\psi} \varphi(t) &\leq \left\langle z(t) - y^*, {}^H D_{a^+}^{\alpha,\mu;\psi} z(t) \right\rangle \\ &= -\theta \langle z(t) - y^*, (\nabla f)(z(t)) \rangle \\ &\leq -\frac{\theta m_f}{2} \|z(t) - y^*\|^2. \end{aligned} \quad (27)$$

Setting

$$h(t) = -\frac{\theta m_f}{2} \|z(t) - y^*\|^2 - {}^H D_{a^+}^{\alpha,\mu;\psi} \varphi(t), \quad (28)$$

we have from (27) that $h(t) \geq 0$ for all $t \geq a$, and moreover, from (28) and recalling (24), the previous expression is equivalent to

$${}^H D_{a^+}^{\alpha, \mu; \psi} \varphi(t) = -\theta m_f \varphi(t) - h(t). \quad (29)$$

By Theorem 5.1 in [22], we have that the solution of (29) is

$$\begin{aligned} \varphi(t) &= \varphi_a [\psi(t) - \psi(a)]^{\alpha + \mu(1-\alpha) - 1} E_{\alpha, \alpha + \mu(1-\alpha)} \left(-\theta m_f (\psi(t) - \psi(a))^\alpha \right) \\ &\quad - \int_a^t (\psi(t) - \psi(a))^{\alpha-1} \psi'(w) E_{\alpha, \alpha} \left(-\theta m_f (\psi(t) - \psi(a))^\alpha \right) h(w) dw \end{aligned} \quad (30)$$

$$\leq \varphi_a [\psi(t) - \psi(a)]^{\alpha + \mu(1-\alpha) - 1} E_{\alpha, \alpha + \mu(1-\alpha)} \left(-\theta m_f (\psi(t) - \psi(a))^\alpha \right), \quad (31)$$

where the last inequality holds, as the integrand function in (30) is positive and $0 \leq a < t$. \square

4.3. Convergence at an Exponential Rate

Theorem 4 establishes the Mittag-Leffler convergence rate for the solution of (23) to a stationary point. Specifically, when $\alpha = 1$, the exponential rate $O(e^{-\theta, m_f \psi(t)})$ of the continuous gradient method (22) is recovered for any $\mu \in [0, 1]$.

Theorem 5. Let $\alpha \in [0, 1]$, $\mu \in [0, 1]$, and ψ satisfy the conditions of Definition 2 with $\sup\{\psi(t) : t \geq a\} = +\infty$. Let $f : \mathbb{R}^n \rightarrow \mathbb{R}$ be a function that is C^1 , convex, and Lipschitz smooth with constant L_f , that is,

$$\|\nabla f(x) - \nabla f(y)\| \leq L_f \|x - y\|, \quad \text{for all } x, y \in \mathbb{R}^n. \quad (32)$$

If the solution $z(\cdot)$ of (23) converges to y^* at the exponential rate $O(e^{-\omega \psi(t)})$, then $y^* = 0$.

Proof. Let $z(\cdot)$ be a solution of (23) converging to the stationary point y^* at the rate $O(e^{-\omega \psi(t)})$. Then, there exists a t_1 greater or equal to a , such that

$$\|z(t) - y^*\| \leq e^{-\omega \psi(t)}, \quad \text{for all } t \geq t_1. \quad (33)$$

By contradiction, let us assume that $y^* \neq 0$. We can then set

$$k = \frac{\theta}{\|y^*\|} + 1. \quad (34)$$

From Formula (4.11.4b) in [20],

$$E_{\alpha, \beta}(z) = 2E_{2\alpha, \beta}(z^2) - E_{\alpha, \beta}(-z), \quad \text{Re}(\alpha) > 0, \beta \in \mathbb{C}, \quad (35)$$

we can find $t_2 \geq t_1$ with the property that

$$E_{\alpha, \beta}(-L_f(\psi(t) - \psi(a))^\alpha) \geq k e^{-\omega \psi(t)}, \quad \text{for all } t \geq t_2. \quad (36)$$

By Proposition 1, $z(\cdot)$ is of the following form

$$\begin{aligned} z(t) &= \frac{z_a}{\Gamma(\alpha + \mu(1-\alpha))} (\psi(t) - \psi(a))^{\alpha + \mu(1-\alpha) - 1} \\ &\quad - \frac{\theta}{\Gamma(\alpha)} \int_a^t (\psi(t) - \psi(s))^{\alpha-1} \psi'(s) (\nabla f)(z(s)) ds, \end{aligned}$$

which is equivalent to

$$\begin{aligned} & z_a(\psi(t) - \psi(a))^{\alpha+\mu(1-\alpha)-1} \\ &= \Gamma(\alpha + \mu(1 - \alpha)) z(t) \\ &+ \frac{\theta \Gamma(\alpha + \mu(1 - \alpha))}{\Gamma(\alpha)} \int_a^t (\psi(t) - \psi(s))^{\alpha-1} \psi'(s) (\nabla f)(z(s)) ds. \end{aligned} \quad (37)$$

Setting

$$u(t) = z_a(\psi(t) - \psi(a))^{\alpha+\mu(1-\alpha)-1} - \Gamma(\alpha + \mu(1 - \alpha)) y^*,$$

then, by (37), we obtain

$$\begin{aligned} \|u(t)\| &= \left\| z_a(\psi(t) - \psi(a))^{\alpha+\mu(1-\alpha)-1} - \Gamma(\alpha + \mu(1 - \alpha)) y^* \right\| \\ &\leq \Gamma(\alpha + \mu(1 - \alpha)) \|z(t) - y^*\| \\ &+ \frac{\theta \Gamma(\alpha + \mu(1 - \alpha))}{\Gamma(\alpha)} \int_a^t (\psi(t) - \psi(s))^{\alpha-1} \psi'(s) \|(\nabla f)(z(s))\| ds. \end{aligned}$$

By assumption (32), we obtain

$$\begin{aligned} & \|u(t)\| \\ &\leq \Gamma(\alpha + \mu(1 - \alpha)) \|z(t) - y^*\| \\ &+ \frac{\theta L_f \Gamma(\alpha + \mu(1 - \alpha))}{\Gamma(\alpha)} \int_a^t (\psi(t) - \psi(s))^{\alpha-1} \psi'(s) \|z(t) - y^*\| ds \\ &= \Gamma(\alpha + \mu(1 - \alpha)) \|z(t) - y^*\| \\ &+ \frac{\theta L_f \Gamma(\alpha + \mu(1 - \alpha))}{\Gamma(\alpha)} \int_a^{t_2} (\psi(t) - \psi(s))^{\alpha-1} \psi'(s) \|z(t) - y^*\| ds \\ &+ \frac{\theta L_f \Gamma(\alpha + \mu(1 - \alpha))}{\Gamma(\alpha)} \int_{t_2}^t (\psi(t) - \psi(s))^{\alpha-1} \psi'(s) \|z(t) - y^*\| ds. \end{aligned} \quad (38)$$

Now, denoting $Q = \sup\{\|z(\cdot) - y^*\| : t \in [a, t_2]\}$ and using (36), we obtain

$$\begin{aligned} & \|u(t)\| \\ &\leq \Gamma(\alpha + \mu(1 - \alpha)) \|z(t) - y^*\| \\ &+ \frac{\theta L_f Q \Gamma(\alpha + \mu(1 - \alpha))}{\Gamma(\alpha)} \int_a^{t_2} (\psi(t) - \psi(s))^{\alpha-1} \psi'(s) ds \\ &+ \frac{\theta L_f \Gamma(\alpha + \mu(1 - \alpha))}{\Gamma(\alpha)} \int_{t_2}^t (\psi(t) - \psi(s))^{\alpha-1} \psi'(s) E_{\alpha, \alpha+\mu(1-\alpha)} \left(-L_f (\psi(s) - \psi(a))^\alpha \right) ds \\ &= \Gamma(\alpha + \mu(1 - \alpha)) \|z(t) - y^*\| \\ &+ \frac{\theta L_f Q \Gamma(\alpha + \mu(1 - \alpha))}{\alpha \Gamma(\alpha)} [(\psi(t) - \psi(a))^\alpha - (\psi(t) - \psi(t_2))^\alpha] \\ &+ \frac{\theta L_f \Gamma(\alpha + \mu(1 - \alpha))}{\Gamma(\alpha)} \int_{t_2}^t (\psi(t) - \psi(s))^{\alpha-1} \psi'(s) E_{\alpha, \alpha+\mu(1-\alpha)} \left(-L_f (\psi(s) - \psi(a))^\alpha \right) ds. \end{aligned}$$

By the mean value theorem (applied to the function $(\psi(t) - \psi(\cdot))^\alpha$), there exists $\delta \in]a, t_2[$ such that

$$(\psi(t) - \psi(a))^\alpha - (\psi(t) - \psi(t_2))^\alpha = -\alpha(\psi(t) - \psi(\delta))^{\alpha-1} \psi'(\delta)$$

which implies that, as $\alpha \in [0, 1]$ and $\sup\{\psi(t) : t \geq a\} = +\infty$,

$$\begin{aligned} & \lim_{t \rightarrow +\infty} [(\psi(t) - \psi(a))^\alpha - (\psi(t) - \psi(t_2))^\alpha] \\ &= -\alpha \lim_{t \rightarrow +\infty} [(\psi(t) - \psi(\delta))^{\alpha-1} \psi'(\delta)] \\ &= -\alpha \lim_{t \rightarrow +\infty} \frac{\psi'(\delta)}{(\psi(t) - \psi(\delta))^{1-\alpha}} = 0. \end{aligned} \quad (39)$$

Hence, taking the limit of $\|u(t)\|$ when $t \rightarrow +\infty$, we obtain

$$\lim_{t \rightarrow +\infty} \|u(t)\| = 0 \Leftrightarrow \Gamma(\alpha + \mu(1 - \alpha)) \|y^*\| = 0.$$

This implies that $\|y^*\| = 0$, which is a contradiction. \square

5. ψ -Hilfer Fractional Gradient Method

The aim of this section is to construct and implement a numerical method for the ψ -Hilfer FGM in one- and two-dimensional cases. For both cases, we perform numerical simulations using benchmark functions.

5.1. The One-Dimensional Case

5.1.1. Design of the Numerical Method

The gradient descent method typically takes steps proportional to the negative gradient (or approximate gradient) of a function at the current iteration, that is, x_{k+1} is updated by the following law

$$x_{k+1} = x_k - \theta f'(x_k), \quad k = 0, 1, 2, \dots, \quad (40)$$

where $\theta > 0$ is the step size or learning rate, and $f'(x_k)$ is the first derivative of f evaluated at $x = x_k$. We assume that $f : \mathbb{R} \rightarrow \mathbb{R}$ admits a local minimum at the point x^* in $\mathbb{D}_\rho(x^*) = \{x \in \mathbb{R} : |x - x^*| < \rho\}$, for some $\rho > 0$, and f admits a Taylor series expansion centered at $x_0 = a$,

$$f(x) = \sum_{p=0}^{+\infty} \frac{f^{(p)}(a)}{p!} (x - a)^p, \quad (41)$$

with domain of convergence $D \subseteq \mathbb{R}$ such that $X^* \in D$. As we want to consider the fractional gradient in the ψ -Hilfer sense, our first (and natural attempt) is to consider the iterative method

$$x_{k+1} = x_k - \theta {}^H D_{a^+}^{\alpha, \mu; \psi} f(\psi(x_k)), \quad k = 0, 1, 2, \dots, \quad (42)$$

where ${}^H D_{a^+}^{\alpha, \mu; \psi}$ is the ψ -Hilfer derivative of order $\alpha \in]0, 1[$ and type $\mu \in [0, 1]$, given by (2), and the function ψ is in the conditions of Definition 2. However, a simple example shows that (42) is not the correct approach. In fact, let us consider the quadratic function $f(x) = (x - h)^2$ with a minimum at $x^* = h$. For this function, we have that

$${}^H D_{a^+}^{\alpha, \mu; \psi} f(x) = \sum_{p=p_0}^2 \frac{f^{(p)}(a)}{\Gamma(p + 1 - \alpha)} (x - a)^{p-\alpha}, \quad (43)$$

where $p_0 = 0$ if $\mu \in [0, 1]$ and $p_0 = 1$ if $\mu = 1$. As ${}^H D_{a^+}^{\alpha, \mu; \psi} f(h) \neq 0$, the iterative method (42) does not converge to the real minimum point. This example shows that the ψ -Hilfer FGM with a fixed lower limit of integration does not converge to the minimum point. This is due to the influence of long-time memory terms, which is an intrinsic feature of fractional derivatives. In order to address this problem and inspired by the ideas presented in [14,15], we replace the starting point a in the fractional derivative by the term x_{k-1} of the previous iteration, that is,

$$x_{k+1} = x_k - \theta {}^H D_{x_{k-1}^+}^{\alpha, \mu; \psi} f(\psi(x_k)), \quad k = 1, 2, \dots \quad (44)$$

where $\alpha \in]0, 1[$, $\mu \in [0, 1]$, and $\theta > 0$. This eliminates the long-time memory effect during the iteration procedure. In this sense, and taking into account the series representation (41) and differentiation rule (5), we obtain

$${}^H D_{x_{k-1}^+}^{\alpha, \mu; \psi} f(\psi(x_k)) = \sum_{p=p_0}^{+\infty} \frac{f^{(p)}(\psi(x_{k-1}))}{\Gamma(p+1-\alpha)} (\psi(x_k) - \psi(x_{k-1}))^{p-\alpha}, \quad (45)$$

where $p_0 = 0$ if $\mu \in [0, 1]$ or $p_0 = 1$ if $\mu = 1$. Thus, the representation formula (45) depends only on $\mu = 1$ or $\mu \neq 1$. With this modification in the ψ -Hilfer FGM, we obtain the following convergence results.

Theorem 6. *If the algorithm (44) is convergent, with the fractional gradient given by (45), then it converges to the minimum point of $f(\psi(\cdot))$.*

Proof. Let x^* be the minimum point of $f(\psi(\cdot))$. We prove that the sequence $(x_k)_{k \in \mathbb{N}}$ converges to x^* by contradiction. Assume that x_k converges to a different $x \neq x^*$ and $f'(\psi(x)) \neq 0$. As the algorithm is convergent, we have that $\lim_{k \rightarrow +\infty} \|x_k - x\| = 0$. Moreover, for any small positive ϵ , there exists a sufficiently large number $N \in \mathbb{N}$, such that

$$|\psi(x_{k-1}) - \psi(x)| < \epsilon < |\psi(x^*) - \psi(x)| \quad (46)$$

for any $k > N$. Thus,

$$\delta = \inf_{p \geq N} |f^{(p_0)}(\psi(x_{k-1}))| > 0$$

must hold. From (45) we have

$$\begin{aligned} |x_{k+1} - x_k| &= \theta \left| {}^H D_{x_{k-1}^+}^{\alpha, \mu; \psi} f(\psi(x_k)) \right| \\ &= \theta \left| \sum_{p=0}^{+\infty} \frac{f^{(p+p_0)}(\psi(x_{k-1}))}{\Gamma(p+p_0+1-\alpha)} (\psi(x_k) - \psi(x_{k-1}))^{p+p_0-\alpha} \right| \\ &\geq \theta \left| \frac{f^{(p_0)}(\psi(x_{k-1}))}{\Gamma(p_0+1-\alpha)} (\psi(x_k) - \psi(x_{k-1}))^{p_0-\alpha} \right| \\ &\quad - \theta \left| \sum_{p=1}^{+\infty} \frac{f^{(p+p_0)}(\psi(x_{k-1}))}{\Gamma(p+p_0+1-\alpha)} (\psi(x_k) - \psi(x_{k-1}))^{p+p_0-\alpha} \right|. \end{aligned}$$

Considering

$$C = \sup_{p \geq N} \frac{|f^{(p_0)}(\psi(x_{k-1}))|}{\Gamma(p_0+1-\alpha)}, \quad (47)$$

we have, from the previous expression,

$$|x_{k+1} - x_k| \geq \theta \left| \frac{f^{(p_0)}(\psi(x_{k-1}))}{\Gamma(p_0 + 1 - \alpha)} (\psi(x_k) - \psi(x_{k-1}))^{p_0 - \alpha} \right| - \theta C \sum_{p=1}^{+\infty} |\psi(x_k) - \psi(x_{k-1})|^{p+p_0-\alpha}.$$

The geometric series in the previous expression is convergent for sufficiently large k . Hence, we obtain

$$|x_{k+1} - x_k| \geq \theta \left| \frac{f^{(p_0)}(\psi(x_{k-1}))}{\Gamma(p_0 + 1 - \alpha)} (\psi(x_k) - \psi(x_{k-1}))^{p_0 - \alpha} \right| - \theta C \frac{|\psi(x_k) - \psi(x_{k-1})|^{1+p_0-\alpha}}{1 - |\psi(x_k) - \psi(x_{k-1})|},$$

which is equivalent to

$$\begin{aligned} & |x_{k+1} - x_k| \\ & \geq \theta \left[\left| \frac{f^{(p_0)}(\psi(x_{k-1}))}{\Gamma(p_0 + 1 - \alpha)} \right| - C \frac{|\psi(x_k) - \psi(x_{k-1})|}{1 - |\psi(x_k) - \psi(x_{k-1})|} \right] |\psi(x_k) - \psi(x_{k-1})|^{p_0 - \alpha} \\ & \geq d |\psi(x_k) - \psi(x_{k-1})|^{p_0 - \alpha}, \end{aligned} \quad (48)$$

where

$$d = d(\epsilon) = \theta \left[\frac{\delta}{\Gamma(p_0 + 1 - \alpha)} - \frac{2C\epsilon}{1 - \epsilon} \right]. \quad (49)$$

One can always find ϵ sufficiently small, such that

$$\frac{\delta}{\Gamma(p_0 + 1 - \alpha)} - \frac{2C\epsilon}{1 - \epsilon} > \frac{2\epsilon^\alpha}{\theta} \Leftrightarrow \frac{\delta}{\Gamma(p_0 + 1 - \alpha)} > \frac{2\epsilon^\alpha}{\theta} + \frac{2C\epsilon}{1 - \epsilon} \quad (50)$$

because the function $g(\epsilon) = \frac{2\epsilon^\alpha}{\theta} + \frac{2C\epsilon}{1 - \epsilon}$ is positively increasing for $\alpha \in [0, 1]$, $\theta \in [0, 1]$, and $\epsilon \in [0, 1]$. Hence, from (48) and taking into account (50), we obtain

$$|x_{k+1} - x_k| > 2\epsilon^{p_0} = \begin{cases} 2\epsilon, & \text{if } \mu = 1 \\ 2, & \text{if } \mu \in [0, 1] \end{cases}. \quad (51)$$

On the other hand, from the assumption (46), we have

$$|x_{k+1} - x_k| \leq |\psi(x_{k+1}) - \psi(x_k)| \leq |\psi(x_{k+1}) - \psi(x)| + |\psi(x) - \psi(x_k)| = 2\epsilon,$$

which contradicts (51). This completes the proof. \square

Sometimes, the function f is not smooth enough to admit a series representation in the form (41), and therefore, the implementation of (44) using the series (45) is not possible. For implementation in practice, we need to truncate the series. In our first approach, we

consider only the term of the series containing $f'(\psi(x_k))$, as it is the most relevant for the gradient method. Thus, the ψ -Hilfer FGM (44) simplifies to

$$x_{k+1} = x_k - \theta \frac{f'(\psi(x_{k-1}))}{\Gamma(2-\alpha)} (\psi(x_k) - \psi(x_{k-1}))^{1-\alpha}. \quad (52)$$

Furthermore, in order to avoid the appearance of complex numbers, we introduce the modulus in the expression (52), that is,

$$x_{k+1} = x_k - \theta \frac{f'(\psi(x_{k-1}))}{\Gamma(2-\alpha)} |\psi(x_k) - \psi(x_{k-1})|^{1-\alpha}. \quad (53)$$

The pseudocode associated to (53) is presented in Algorithm 1. As (53) is independent of the μ parameter, from now on we call the method ψ -FGM. Following the same arguments as in the proof of Theorem 6, we obtain the following results.

Theorem 7. *If the algorithm (44) is convergent with the fractional gradient given by (53), then it converges to the minimum point of $f(\psi(\cdot))$.*

In the following pseudocode, we describe the implementation of the algorithm (53).

Algorithm 1: ψ -FGM with higher order truncation

Inputs:

Functions: $\psi(x), f'(\psi(x))$

Fixed parameters: $\alpha, a, \theta, \epsilon$

Initial guess: x_0

Output: k -iteration: x_k

Initialization:

$k = 2, x_1 = a$

while $|f'(\psi(x_k))| \geq \epsilon$ **do**

$x_k = x_{k-1} - \theta \frac{f'(\psi(x_{k-2}))}{\Gamma(2-\alpha)} |\psi(x_{k-1}) - \psi(x_{k-2})|^{1-\alpha}$

$k = k + 1$

end

As we have seen, it is possible to construct a ψ -FGM that converges to the minimum point of a function. To improve the convergence of the proposed method, we can consider variable order differentiation $\alpha(x)$ in each iteration. Some examples of $\alpha(x)$ are given by (see [15]):

$$\alpha(x) = \frac{1}{1 + \beta J(x)}, \quad (54)$$

$$\alpha(x) = \frac{2}{1 + e^{\beta J(x)}}, \quad (55)$$

$$\alpha(x) = \frac{1}{\cos(\beta J(x))}, \quad (56)$$

$$\alpha(x) = 1 - \frac{2}{\pi} \arctan(\beta J(x)), \quad (57)$$

$$\alpha(x) = 1 - \tanh(\beta J(x)), \quad (58)$$

where $\beta > 0$ and we consider the loss function $J(x) = (f'(\psi(x)))^2$ to be minimized in each iteration. The consideration of the square in the loss function guarantees its non-negativity. All examples given satisfy

$$\lim_{\beta \rightarrow 0} \alpha(x) = 1 = \lim_{x \rightarrow x^*} \alpha(x), \quad (59)$$

where the second limit results from the fact that $J(x) \rightarrow 0$ as $x \rightarrow x^*$. Variable order differentiation turns the ψ -FGM into a learning method, because as x gradually approaches x^* , $\alpha(x) \approx 1$. The ψ -FGM with variable order is given by

$$x_{k+1} = x_k - \theta \frac{f'(\psi(x_{k-1}))}{\Gamma(2 - \alpha(x_k))} |\psi(x_k) - \psi(x_{k-1})|^{1 - \alpha(x_k)}.$$

Theorem 6 remains valid for this variation of Algorithm 1.

5.1.2. Numerical Simulations

Now, we provide some examples that show the validity of the previous algorithms applied to the quadratic function $f(x) = (x - 1)^2$, which is one of the simplest benchmark functions. This function is a convex function with a unique global minimum at $x^* = 1$. For the ψ -derivative, we consider the following cases:

- *Caputo and Riemann–Liouville fractional derivatives*: $\psi_1(x) = x$, $I = [0, +\infty[$, and $a = 0$;
- *Hadamard fractional derivative*: $\psi_2(x) = \ln(x)$, $I = [1, +\infty[$, and $a = 1$;
- *Katugampola fractional derivative*: $\psi_3(x) = x^{0.5}$, $I = [0, +\infty[$, and $a = 0.5$. In this case, a cannot coincide with the lower limit of the interval I because ψ' is not defined at $x = 0$.

Figures 1–3 show the numerical results of Algorithm 1 applied to the composite functions $f(\psi_1(\cdot))$, $f(\psi_2(\cdot))$, and $f(\psi_3(\cdot))$, choosing different parameters. In Figure 1, we consider $\theta = 0.1$, $x_0 = 2$, $\epsilon = 10^{-9}$, and different orders of differentiation $\alpha = 0.4, 0.6, 0.8, 1.0$.

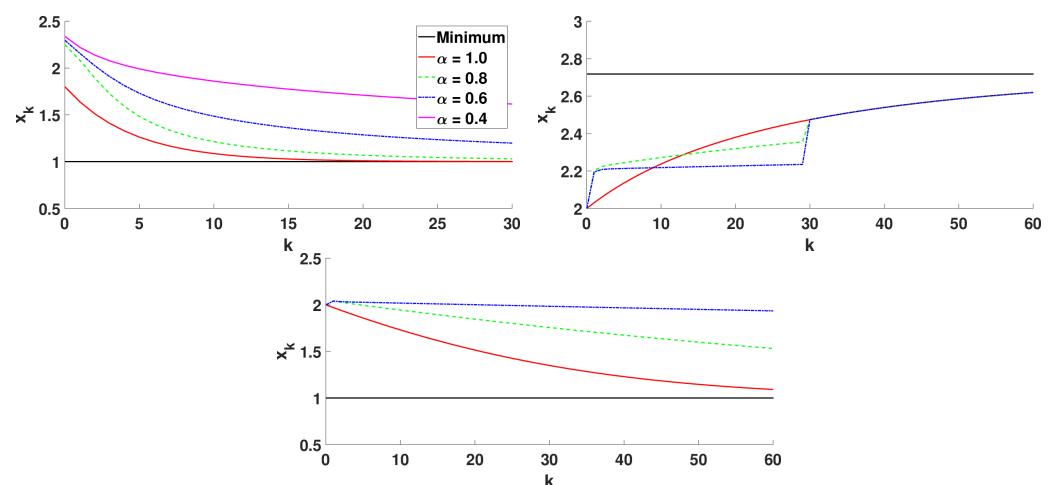


Figure 1. Algorithm 1 for different orders of differentiation, α .

Analyzing the plots in Figure 1, we see that the convergence of the ψ -FGM in the non-integer case is slower, in general, than in the integer case ($\alpha = 1.0$). In Figure 2, we consider $x_0 = 2$, $\alpha = 0.75$, $\epsilon = 10^{-9}$, and different step sizes $\theta = 0.05, 0.1, 0.2, 0.3$.

The plots in Figure 2 show that the increment of the step size makes convergence faster. The optimization of the step size in each iteration would lead to the optimal convergence of the method in a few iterations. In Figure 3, we consider different initial approximations x_0 , and the values $\alpha = 0.75$, $\theta = 0.1$, and $\epsilon = 10^{-9}$. As expected, convergence becomes faster as x_0 approaches the minimum point.

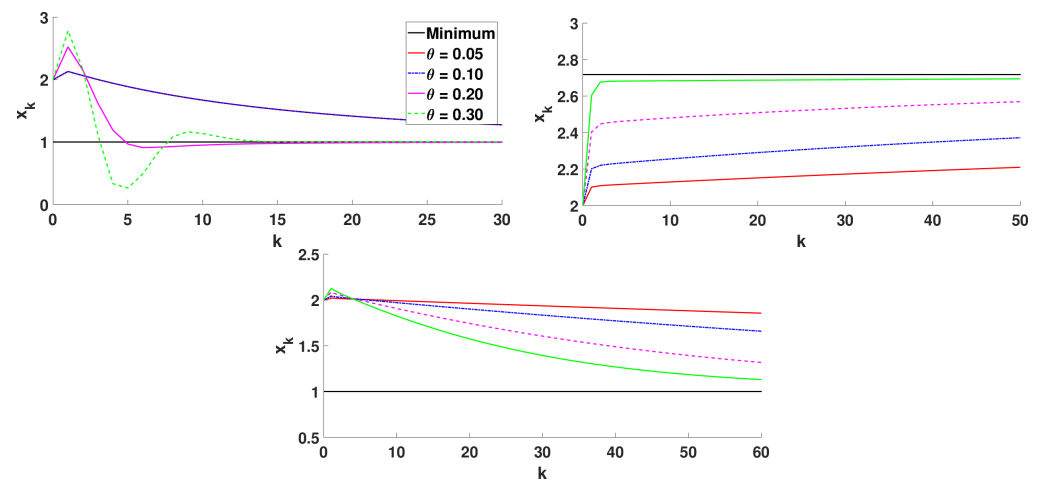


Figure 2. Algorithm 1 for different step sizes, θ .

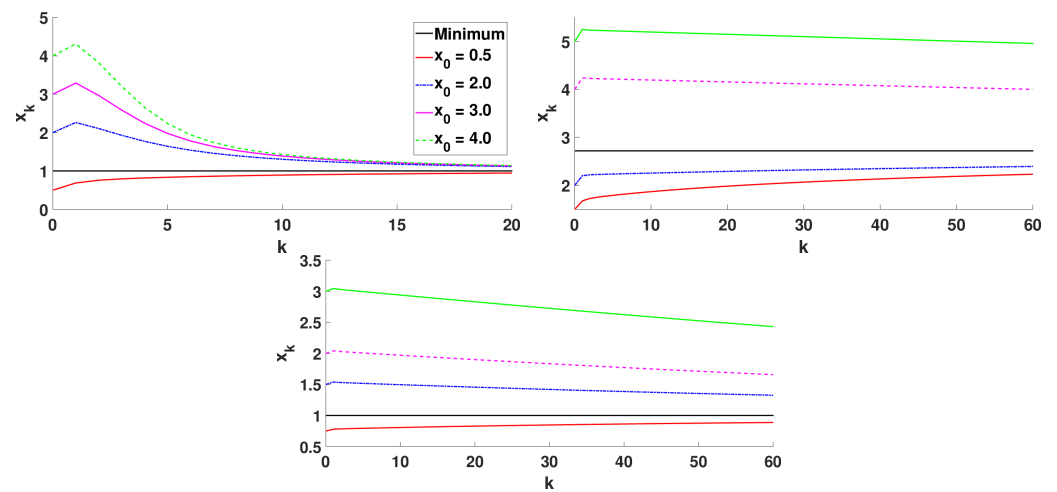


Figure 3. Algorithm 1 for different initial approximations, x_0 .

Now, we show the numerical simulations of Algorithm 1 with a variable order of differentiation $\alpha(x)$. In Figure 4, we consider $\theta = 0.1$, $\beta = 0.1$, $x_0 = 2$, $\epsilon = 10^{-9}$, and the variable order Functions (54)–(58). In Figure 5, we exhibit the behaviour of the algorithm for $\alpha(x)$ given by (58) and different values of β .

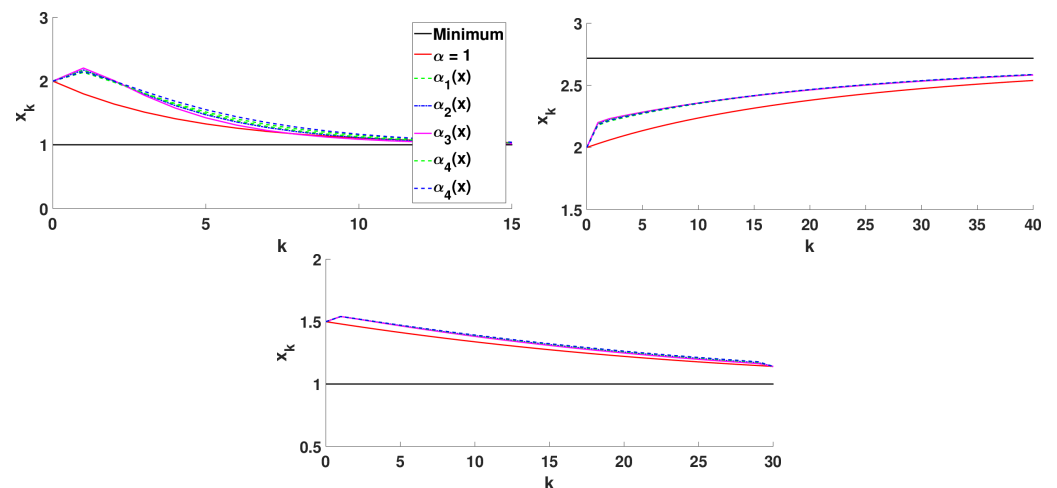


Figure 4. Algorithm 1 for different variable orders of differentiation.

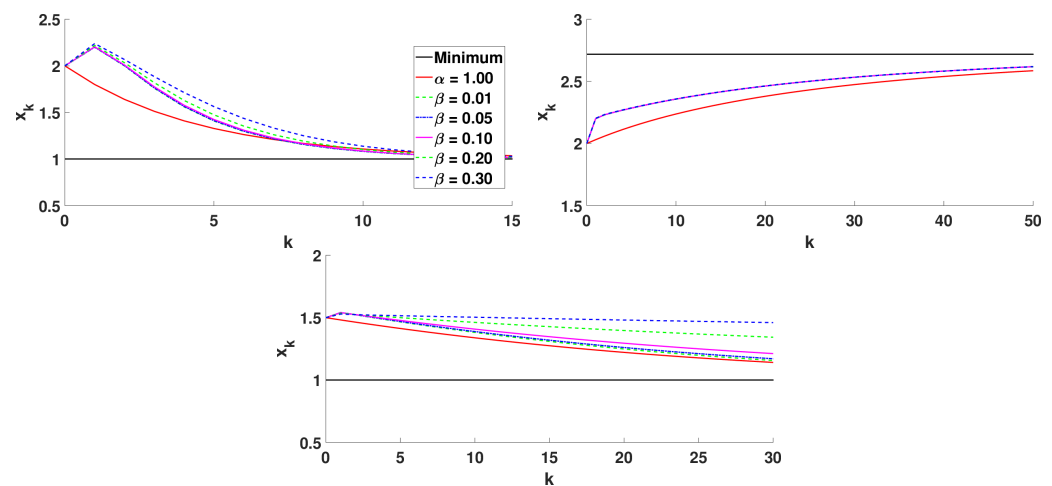


Figure 5. Algorithm 1 with $\alpha(x) = 1 - \tanh(\beta f(x))$ and different values of β .

From these plots, we conclude that in the one-dimensional case, the consideration of variable order differentiation can speed up the convergence, but it is, in general, slower than the classical gradient descent method with integer derivative. A further improvement of the algorithm could be made by considering a variable step size optimized in each iteration. This idea is implemented in the next section, where we consider optimization problems in \mathbb{R}^2 .

5.2. The Two-Dimensional Case

5.2.1. Untrained Approach

Motivated by the ideas presented in [14,15], we extend the results presented in Section 5.1 to the two-dimensional case. We consider a function ψ in the conditions of Definition 2 and the vector-valued function $\Psi : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ given by $\Psi(X) = (\psi(x), \psi(y))$ with $X = (x, y)$. Moreover, let $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ be a function that admits a local minimum at the point X^* in $\mathbb{D}_\rho(X^*) = \{X \in \mathbb{R}^2 : \|X - X^*\| < \rho\}$ for some $\rho > 0$. We want to find a local minimum point of the function $f(\Psi(X)) = f(\psi(x), \psi(y))$, with $X = (x, y)$, through the iterative method

$$X_{k+1} = X_k - \theta \left({}^H\nabla_{X_{k-1}^+}^{\alpha, \mu; \psi} f \right) (\Psi(X_k)), \quad \theta > 0. \quad (60)$$

We assume that f admits a Taylor series centered at the point $(a_1, a_2) \in \mathbb{D}_\rho(X^*)$, given by

$$f(x, y) = \sum_{p=0}^{+\infty} \sum_{q=0}^{+\infty} \left(\frac{\partial^{p+q} f}{\partial x^p \partial y^q} \right) (a_1, a_2) \frac{1}{p! q!} (x - a_1)^p (y - a_2)^q, \quad (61)$$

with a domain of convergence $D \subseteq \mathbb{R}^2$, such that $X^* \in D$. Then, the ψ -Hilfer fractional gradient in (60) is given by

$$\left({}^H\nabla_{X_{k-1}^+}^{\alpha, \mu; \psi} f \right) (\Psi(X_k)) = \left(\left({}^H\partial_{x_{k-1}^+}^{\alpha, \mu; \psi} f \right) (\Psi(X_k)), \left({}^H\partial_{y_{k-1}^+}^{\alpha, \mu; \psi} f \right) (\Psi(X_k)) \right), \quad (62)$$

where $\left({}^H\partial_{x_{k-1}^+}^{\alpha, \mu; \psi} f \right) (\Psi(X_k))$ and $\left({}^H\partial_{y_{k-1}^+}^{\alpha, \mu; \psi} f \right) (\Psi(X_k))$ denote the partial ψ -Hilfer derivatives of f , with respect to x and y , of order $\alpha \in [0, 1]$, type $\mu \in [0, 1]$, and with the lower limit of integration replaced by x_{k-1} and y_{k-1} , respectively. Taking into account (61) and (5), we have the following expressions for the components of (62)

$$\begin{aligned} & \left({}^H_{x_{k-1}^+} \partial_x^{\alpha, \mu; \psi} f \right) (\Psi(X_k)) \\ &= \sum_{p=p_0}^{+\infty} \sum_{q=0}^{+\infty} \left(\frac{\partial^{p+q} f}{\partial_x^p \partial_y^q} \right) (\Psi(X_{k-1})) \frac{(\psi(y_k) - \psi(y_{k-1}))^q}{q!} \frac{(\psi(x_k) - \psi(x_{k-1}))^{p-\alpha}}{\Gamma(p+1-\alpha)}, \end{aligned} \quad (63)$$

$$\begin{aligned} & \left({}^H_{y_{k-1}^+} \partial_x^{\alpha, \mu; \psi} f \right) (\Psi(X_k)) \\ &= \sum_{p=0}^{+\infty} \sum_{q=q_0}^{+\infty} \left(\frac{\partial^{p+q} f}{\partial_x^p \partial_y^q} \right) (\Psi(X_{k-1})) \frac{(\psi(x_k) - \psi(x_{k-1}))^p}{p!} \frac{(\psi(y_k) - \psi(y_{k-1}))^{q-\alpha}}{\Gamma(q+1-\alpha)}, \end{aligned} \quad (64)$$

where

$$p_0 = q_0 = \begin{cases} 0, & \text{if } \mu \in [0, 1] \\ 1, & \text{if } \mu = 1 \end{cases}. \quad (65)$$

The iterative method proposed in (60) takes into account the short memory characteristics of the fractional derivatives and, as in the one-dimensional case, we can see from (63) and (64) that the method does not depend on the type of derivative μ . Furthermore, due to the freedom of choice of the μ parameter ($\mu = 1$ or $\mu \in [0, 1]$) and the ψ function, we can deal with several fractional derivatives (see Section 5 in [19]). We obtain the following convergence results:

Theorem 8. *If the algorithm (60) is convergent where the fractional gradient is given by (62)–(64), then it converges to the minimum point of $f(\Psi(\cdot))$.*

Proof. Let X^* be the minimum point of $f(\Psi(\cdot))$. We prove that the sequence $(X_k)_{k \in \mathbb{N}}$ converges to X^* by contradiction. Assume that X_k converges to a different $X \neq X^*$ and $f'_x(\Psi(X)) \neq 0 \neq f'_y(\Psi(X))$. As the algorithm is convergent, we have that $\lim_{k \rightarrow +\infty} \|X_k - X\| = 0$. Moreover, for any small positive ϵ , there exists a sufficiently large number $N \in \mathbb{N}$, such that

$$(\psi(x_{k-1}) - \psi(x))^2 < \frac{\epsilon^2}{2} < (\psi(x^*) - \psi(x))^2 \quad (66)$$

and

$$(\psi(y_{k-1}) - \psi(y))^2 < \frac{\epsilon^2}{2} < (\psi(y^*) - \psi(y))^2 \quad (67)$$

for any $k > N$. Thus,

$$\delta_1 = \inf_{p, q > N} \left| \frac{\partial^{p_0} f}{\partial x^{p_0}} (\Psi(X_{k-1})) \right| > 0 \quad \text{and} \quad \delta_2 = \inf_{p, q > N} \left| \frac{\partial^{q_0} f}{\partial y^{q_0}} (\Psi(X_{k-1})) \right| > 0 \quad (68)$$

must hold. From (62)–(64) we have

$$\begin{aligned}
& \|X_{k+1} - X_k\|^2 \\
&= \theta^2 \left\| {}^H\nabla_{x_{k-1}^+}^{\alpha, \mu; \psi} f(\Psi(X_{k-1})) \right\|^2 \\
&= \theta^2 \left(\sum_{p=0}^{+\infty} \sum_{q=0}^{+\infty} \left(\frac{\partial^{p+p_0+q} f}{\partial x^{p+p_0} \partial y^q} \right) (\Psi(X_{k-1})) \frac{(\psi(y_k) - \psi(y_{k-1}))^q}{q!} \frac{(\psi(x_k) - \psi(x_{k-1}))^{p+p_0-\alpha}}{\Gamma(p+p_0+1-\alpha)} \right)^2 \\
&\quad + \theta^2 \left(\sum_{p=0}^{+\infty} \sum_{q=0}^{+\infty} \left(\frac{\partial^{p+q+q_0} f}{\partial x^p \partial y^{q+q_0}} \right) (\Psi(X_{k-1})) \frac{(\psi(x_k) - \psi(x_{k-1}))^p}{p!} \frac{(\psi(y_k) - \psi(y_{k-1}))^{q+q_0-\alpha}}{\Gamma(q+q_0+1-\alpha)} \right)^2 \\
&\geq \theta^2 \left(\frac{\partial^{p_0} f}{\partial x^{p_0}} (\Psi(X_{k-1})) \frac{(\psi(x_k) - \psi(x_{k-1}))^{p_0-\alpha}}{\Gamma(p_0+1-\alpha)} \right)^2 \\
&\quad - \theta^2 \left(\sum_{p=1}^{+\infty} \sum_{q=1}^{+\infty} \left(\frac{\partial^{p+p_0+q} f}{\partial x^p \partial y^q} \right) (\Psi(X_{k-1})) \frac{(\psi(y_k) - \psi(y_{k-1}))^q}{q!} \frac{(\psi(x_k) - \psi(x_{k-1}))^{p+p_0-\alpha}}{\Gamma(p+p_0+1-\alpha)} \right)^2 \\
&\quad + \theta^2 \left(\frac{\partial^{q_0} f}{\partial y^{q_0}} (\Psi(X_{k-1})) \frac{(\psi(y_k) - \psi(y_{k-1}))^{q_0-\alpha}}{\Gamma(q_0+1-\alpha)} \right)^2 \\
&\quad - \theta^2 \left(\sum_{p=1}^{+\infty} \sum_{q=1}^{+\infty} \left(\frac{\partial^{p+q+q_0} f}{\partial x^p \partial y^q} \right) (\Psi(X_{k-1})) \frac{(\psi(x_k) - \psi(x_{k-1}))^p}{p!} \frac{(\psi(y_k) - \psi(y_{k-1}))^{q+q_0-\alpha}}{\Gamma(q+q_0+1-\alpha)} \right)^2.
\end{aligned}$$

Considering

$$C_1 = \sup_{p,q \geq N} \frac{\left| \frac{\partial^{p_0} f}{\partial x^{p_0}} (\Psi(X_{k-1})) \right|}{\Gamma(p_0+1-\alpha) q!} \quad \text{and} \quad C_2 = \sup_{p,q \geq N} \frac{\left| \frac{\partial^{q_0} f}{\partial y^{q_0}} (\Psi(X_{k-1})) \right|}{\Gamma(q_0+1-\alpha) p!},$$

from the previous expression, we obtain

$$\begin{aligned}
\|X_{k+1} - X_k\|^2 &\geq \theta^2 \left(\frac{\partial^{p_0} f}{\partial x^{p_0}} (\Psi(X_{k-1})) \frac{(\psi(x_k) - \psi(x_{k-1}))^{p_0-\alpha}}{\Gamma(p_0+1-\alpha)} \right)^2 \\
&\quad - \theta^2 C_1^2 \left(\sum_{p=1}^{+\infty} \sum_{q=1}^{+\infty} (\psi(y_k) - \psi(y_{k-1}))^q (\psi(x_k) - \psi(x_{k-1}))^{p+p_0-\alpha} \right)^2 \\
&\quad + \theta^2 \left(\frac{\partial^{q_0} f}{\partial y^{q_0}} (\Psi(X_{k-1})) \frac{(\psi(y_k) - \psi(y_{k-1}))^{q_0-\alpha}}{\Gamma(q_0+1-\alpha)} \right)^2 \\
&\quad - \theta^2 C_2^2 \left(\sum_{p=1}^{+\infty} \sum_{q=1}^{+\infty} (\psi(x_k) - \psi(x_{k-1}))^p (\psi(y_k) - \psi(y_{k-1}))^{q+q_0-\alpha} \right)^2.
\end{aligned}$$

The double series that appears in the previous expression are of a geometric type with positive radius less than 1. Hence, by the sum of a geometric series, we have

$$\begin{aligned}
\|X_{k+1} - X_k\|^2 &\geq \theta^2 \left(\frac{\partial^{p_0} f}{\partial x^{p_0}}(\Psi(X_{k-1})) \frac{(\psi(x_k) - \psi(x_{k-1}))^{p_0 - \alpha}}{\Gamma(p_0 + 1 - \alpha)} \right)^2 \\
&\quad - \theta^2 C_1^2 \left(\frac{\psi(y_k) - \psi(y_{k-1})}{1 - (\psi(y_k) - \psi(y_{k-1}))} \right)^2 \left(\frac{(\psi(x_k) - \psi(x_{k-1}))^{1+p_0-\alpha}}{1 - (\psi(x_k) - \psi(x_{k-1}))} \right)^2 \\
&\quad + \theta^2 \left(\frac{\partial^{q_0} f}{\partial y^{q_0}}(\Psi(X_{k-1})) \frac{(\psi(y_k) - \psi(y_{k-1}))^{q_0 - \alpha}}{\Gamma(q_0 + 1 - \alpha)} \right)^2 \\
&\quad - \theta^2 C_2^2 \left(\frac{\psi(x_k) - \psi(x_{k-1})}{1 - (\psi(x_k) - \psi(x_{k-1}))} \right)^2 \left(\frac{(\psi(y_k) - \psi(y_{k-1}))^{1+q_0-\alpha}}{1 - (\psi(y_k) - \psi(y_{k-1}))} \right)^2,
\end{aligned}$$

which is equivalent to

$$\begin{aligned}
&\|X_{k+1} - X_k\|^2 \\
&\geq \theta^2 \left(\left(\frac{\partial^{p_0} f}{\partial x^{p_0}}(\Psi(X_{k-1})) \right)^2 - C_1^2 \left(\frac{\psi(y_k) - \psi(y_{k-1})}{1 - (\psi(y_k) - \psi(y_{k-1}))} \right)^2 \left(\frac{\psi(x_k) - \psi(x_{k-1})}{1 - (\psi(x_k) - \psi(x_{k-1}))} \right)^2 \right) \\
&\quad \times \left((\psi(x_k) - \psi(x_{k-1}))^{p_0 - \alpha} \right)^2 \\
&\quad + \theta^2 \left(\left(\frac{\partial^{q_0} f}{\partial y^{q_0}}(\Psi(X_{k-1})) \right)^2 - C_2^2 \left(\frac{\psi(x_k) - \psi(x_{k-1})}{1 - (\psi(x_k) - \psi(x_{k-1}))} \right)^2 \left(\frac{\psi(y_k) - \psi(y_{k-1})}{1 - (\psi(y_k) - \psi(y_{k-1}))} \right)^2 \right) \\
&\quad \times \left((\psi(y_k) - \psi(y_{k-1}))^{q_0 - \alpha} \right)^2 \\
&\geq d_1 \left((\psi(x_k) - \psi(x_{k-1}))^{p_0 - \alpha} \right)^2 + d_2 \left((\psi(y_k) - \psi(y_{k-1}))^{q_0 - \alpha} \right)^2, \tag{69}
\end{aligned}$$

where

$$d_1 = d_1(\epsilon) = \theta^2 \left(\left(\frac{\delta_1}{\Gamma(p_0 + 1 - \alpha)} \right)^2 - C_1^2 \left(\frac{\epsilon}{1 - \epsilon} \right)^4 \right)$$

and

$$d_2 = d_2(\epsilon) = \theta^2 \left(\left(\frac{\delta_2}{\Gamma(q_0 + 1 - \alpha)} \right)^2 - C_2^2 \left(\frac{\epsilon}{1 - \epsilon} \right)^4 \right).$$

One can always find ϵ sufficiently small, such that

$$\begin{aligned}
&\left(\frac{\delta_1}{\Gamma(p_0 + 1 - \alpha)} \right)^2 - C_1^2 \left(\frac{\epsilon}{1 - \epsilon} \right)^4 > \frac{\epsilon^{2\alpha}}{\theta^2} \\
&\Leftrightarrow \left(\frac{\delta_1}{\Gamma(p_0 + 1 - \alpha)} \right)^2 > \frac{\epsilon^{2\alpha}}{\theta^2} + C_1^2 \left(\frac{\epsilon}{1 - \epsilon} \right)^4 \tag{70}
\end{aligned}$$

and

$$\begin{aligned} & \left(\frac{\delta_2}{\Gamma(q_0 + 1 - \alpha)} \right)^2 - C_1^2 \left(\frac{\epsilon}{1 - \epsilon} \right)^4 > \frac{\epsilon^{2\alpha}}{\theta^2} \\ & \Leftrightarrow \left(\frac{\delta_2}{\Gamma(q_0 + 1 - \alpha)} \right)^2 > \frac{\epsilon^{2\alpha}}{\theta^2} + C_2^2 \left(\frac{\epsilon}{1 - \epsilon} \right)^4, \end{aligned} \quad (71)$$

because the functions

$$g_1(\epsilon) = \frac{\epsilon^{2\alpha}}{\theta^2} + C_1^2 \left(\frac{\epsilon}{1 - \epsilon} \right)^4 \quad \text{and} \quad g_2(\epsilon) = \frac{\epsilon^{2\alpha}}{\theta^2} + C_2^2 \left(\frac{\epsilon}{1 - \epsilon} \right)^4$$

are positive increasing for $\alpha \in [0, 1]$, $\theta \in [0, 1]$, and $\epsilon \in [0, 1]$. Hence, from (69) and taking into account (70) and (71), we obtain

$$\|X_{k+1} - X_k\|^2 > \epsilon^{2\alpha} \epsilon^{2p_0 - 2\alpha} + \epsilon^{2\alpha} \epsilon^{2q_0 - 2\alpha} = \begin{cases} 2\epsilon^2, & \text{if } \mu = 1 \\ 2, & \text{if } \mu \in [0, 1] \end{cases}. \quad (72)$$

On the other hand, from assumption (71), we have

$$\begin{aligned} & \|X_{k+1} - X_k\|^2 \\ & \leq \|\Psi(X_{k+1}) - \Psi(X_k)\|^2 \\ & \leq (\psi(x_{k+1}) - \psi(x_k))^2 + (\psi(y_{k+1}) - \psi(y_k))^2 \\ & \leq (\psi(x_{k+1}) - \psi(x))^2 + (\psi(x) - \psi(x_k))^2 + (\psi(y_{k+1}) - \psi(y))^2 + (\psi(y) - \psi(y_k))^2 \\ & < \frac{\epsilon^2}{2} + \frac{\epsilon^2}{2} + \frac{\epsilon^2}{2} + \frac{\epsilon^2}{2} = 2\epsilon^2, \end{aligned}$$

which contradicts (72). This completes the proof. \square

Despite the convergence of (60), it is important to point out that sometimes the function $f(\Psi(\cdot))$ is not smooth enough. In these cases, the algorithm involving the double series (63) and (64) cannot be implemented. Moreover, in the same way as was done for the one-dimensional case, assuming that f is at least of the class C^1 , we only consider the following terms of (63) and (64)

$$\begin{aligned} & \frac{f'_x(\Psi(X_{k-1}))}{\Gamma(2 - \alpha)} (\psi(x_k) - \psi(x_{k-1}))^{1-\alpha} \text{ (term } p = 1 \text{ and } q = 0 \text{ in (63))}, \\ & \frac{f'_y(\Psi(X_{k-1}))}{\Gamma(2 - \alpha)} (\psi(y_k) - \psi(y_{k-1}))^{1-\alpha}, \text{ (term } p = 0 \text{ and } q = 1 \text{ in (64))}. \end{aligned}$$

Thus, the higher order terms are eliminated and we have the following update of (62):

$$\begin{aligned} & \left({}^H\nabla_{X_{k-1}^+}^{\alpha, \mu; \psi} f \right) (\Psi(X_k)) \\ & = \left(\frac{f'_x(\Psi(X_{k-1}))}{\Gamma(2 - \alpha)} (\psi(x_k) - \psi(x_{k-1}))^{1-\alpha}, \frac{f'_y(\Psi(X_{k-1}))}{\Gamma(2 - \alpha)} (\psi(y_k) - \psi(y_{k-1}))^{1-\alpha} \right). \end{aligned}$$

To avoid the appearance of complex numbers, we also consider

$$\begin{aligned} & \left({}^H\nabla_{X_{k-1}^+}^{\alpha, \mu; \psi} f \right) (\Psi(X_k)) \\ &= \left(\frac{f'_x(\Psi(X_{k-1}))}{\Gamma(2-\alpha)} |\psi(x_k) - \psi(x_{k-1})|^{1-\alpha}, \frac{f'_y(\Psi(X_{k-1}))}{\Gamma(2-\alpha)} |\psi(y_k) - \psi(y_{k-1})|^{1-\alpha} \right). \end{aligned} \quad (73)$$

The implementation of (73) is presented in Algorithm 2. In a similar way as it was done in Theorem 8, we can state the following results.

Theorem 9. *If the algorithm (60) is convergent where the fractional gradient is given by (73), then it converges to the minimum point of $f(\Psi(\cdot))$.*

Algorithm 2: 2D ψ -FGM with higher order truncation

Inputs:

Functions: $\psi(x), f'_x(\Psi(X)), f'_y(\Psi(X))$

Fixed parameters: $\alpha, a, \theta, \epsilon$

Initial guess: $X_0 = [x_0, y_0]$

Output: k -iteration: $X_k = [x_k, y_k]$

Initialization

$k = 2, X_1 = [a, a]$

while $\left\| {}^H\nabla_{X_k^+}^{\alpha, \mu; \psi} f(\Psi(X_{k+1})) \right\| \geq \epsilon$ **do**

$$\begin{aligned} x_k &= x_{k-1} - \theta \frac{f'_x(\Psi(X_{k-2}))}{\Gamma(2-\alpha)} |\psi(x_{k-1}) - \psi(x_{k-2})|^{1-\alpha} \\ y_k &= y_{k-1} - \theta \frac{f'_y(\Psi(X_{k-2}))}{\Gamma(2-\alpha)} |\psi(y_{k-1}) - \psi(y_{k-2})|^{1-\alpha} \\ X_k &= [x_k, y_k] \\ k &= k + 1 \end{aligned}$$

end

5.2.2. The Trained Approach

In this section, we refine Algorithm 2 in two ways that train our algorithm in each iteration to find the most accurate X_k . First, we consider a variable step size $\theta_k > 0$ that is updated in each iteration by minimizing the following function

$$\phi(\theta_k) = f \left(\Psi(X_k) - \theta_k \left({}^H\nabla_{X_{k-1}^+}^{\alpha, \mu; \psi} f \right) (\Psi(X_k)) \right).$$

In the second refinement, we adjust the order of integration α with X_k . More precisely, if f is a non-negative function with a unique minimum point X^* , we can consider any of the functions (54)–(58). In our approach, we only consider the following variable fractional order

$$\alpha(X) = 1 - \frac{2}{\pi} \arctan(\beta J(X)), \quad (74)$$

where the loss function is $J(X) = \|\nabla f(\Psi(X))\|$. From (74), we infer that when $J(X) \approx 0$, one has $\alpha(X) \approx 1$, and when $J(X) \gg 0$, one has $\alpha(X) \approx 0$. As a consequence of the previous refinements, we have the following iterative method:

$$X_{k+1} = X_k - \theta_k \left({}^H\nabla_{X_{k-1}^+}^{\alpha(X_k), \mu; \psi} f \right) (\Psi(X_k)), \quad (75)$$

where the fractional gradient is given by

$$\begin{aligned} & \left({}^H\nabla_{X_{k-1}^+}^{\alpha(X_k), \mu; \psi} f \right) (\Psi(X_k)) \\ &= \left(\frac{f'_x(\Psi(X_{k-1}))}{\Gamma(2-\alpha(X_k))} |\psi(x_k) - \psi(x_{k-1})|^{1-\alpha(X_k)}, \frac{f'_y(\Psi(X_{k-1}))}{\Gamma(2-\alpha(X_k))} |\psi(y_k) - \psi(y_{k-1})|^{1-\alpha(X_k)} \right). \end{aligned} \quad (76)$$

The implementation of (76) is presented in Algorithm 3. Likewise, with a variable fractional order $\alpha(X)$, the following theorem follows.

Theorem 10. *If the algorithm (75) is convergent where the fractional gradient is given by (76), then it converges to the minimum point of $f(\Psi(\cdot))$.*

The proof of this result follows the same reasoning of the proof of Theorem 8, and therefore, is omitted.

Algorithm 3: 2D ψ -FGM with variable fractional order and optimized step size

Inputs:

Functions: $\psi(x), f'_x(\Psi(X)), f'_y(\Psi(X))$

Fixed parameters: $\alpha(X), a, \epsilon, \beta$

Initial guess: $X_0 = [x_0, y_0]$

Output: k -iteration: $X_k = [x_k, y_k]$

Initialization

$k = 2, X_1 = [a, a]$

while $\left\| {}^H\nabla_{X_k^+}^{\alpha(X_k), \mu; \psi} f(\Psi(X_{k+1})) \right\| \geq \epsilon$ **do**

$$\begin{aligned} & J(X_{k-2}) = \left\| \nabla f(\Psi(X_{k-2})) \right\|, \alpha(X_{k-2}) = \text{subs}(\alpha(X), J(X), J(X_{k-2})) \\ & \theta_k = \text{Solve} \left[\text{Diff} \left[f \left(\Psi(X_k) - \theta \left({}^H\nabla_{X_{k-1}^+}^{\alpha(X_k), \mu; \psi} f \right) (\Psi(X_k)) \right) \right] = 0, \theta \right] \\ & x_k = x_{k-1} - \theta_k \frac{f'_x(\Psi(X_{k-2}))}{\Gamma(2-\alpha)} |\psi(x_{k-1}) - \psi(x_{k-2})|^{1-\alpha(X_{k-2})} \\ & y_k = y_{k-1} - \theta_k \frac{f'_y(\Psi(X_{k-2}))}{\Gamma(2-\alpha)} |\psi(y_{k-1}) - \psi(y_{k-2})|^{1-\alpha(X_{k-2})} \\ & X_k = [x_k, y_k] \\ & k = k + 1 \end{aligned}$$

end

5.2.3. Numerical Simulations

In this section, we implement Algorithms 2 and 3 for finding the local minimum point of the function $f(\Psi(\cdot))$ for particular choices of f and ψ . For the function f , we consider the following cases:

- $f_1(x, y) = 4x^2 - 4xy + 2y^2$ with minimum point at $(0, 0)$,
- Matyas function: $f_2(x, y) = 0.26(x^2 + y^2) - 0.48xy$ with minimum point at $(0, 0)$,
- Wayburn and Seader No. 1 function: $f_3(x, y) = (x^6 + y^4 - 17)^2 + (2x + y - 4)^2$ with minimum point at $(1, 2)$.

The function f_1 is a classic convex quadratic function in \mathbb{R}^2 and can be considered an academic example for implementing our algorithms. The choice of functions f_2 and f_3 is due to the fact that they are benchmark functions used to test and evaluate several characteristics of optimization algorithms, such as convergence rate, precision, robustness, and general performance. More precisely, the Matyas function has a plate shape and the Wayburn and Seader No. 1 function has a valley shape, which implies slow convergence to the minimum point of the corresponding function. For the functions ψ in the vector function Ψ , we consider the choices

- $\psi_1(x) = x$, with $x \in I = [0, +\infty[$,
- $\psi_2(x) = x^2$, with $x \in I = [0, +\infty[$,

- $\psi_3(x) = \ln x$, with $x \in I = [1, +\infty[$.

For the numerical simulations, we consider some combinations of the functions f_i , $i = 1, 2$, and ψ_j , $j = 1, 2, 3$, and compare the results in the following scenarios:

- Algorithm 3 with $\alpha(X) = 1$ that corresponds to the classical 2D gradient descent method,
- Algorithm 2 with $\alpha(X) = 0.8$ and step size $\theta = 0.1$,
- Algorithm 3 with $\alpha(X) = 1 - \frac{2}{\pi} \arctan(\beta J(X))$, with $\beta = 0.1$.

Figure 6 shows the target functions $f_1(\Psi_1(X)) = 4x^2 - 4xy + 2y^2$ and $f_1(\Psi_2(X)) = 4x^4 - 4x^2y^2 + 2y^4$, both with a local minimum point at $X^* = (0, 0)$. Figures 7 and 8 show the X_k iterates in the corresponding contour plots of the functions. The plots on the right show the amplification close to the minimum point. The results of numerical simulations are summarized in Table 1. The stopping criterion used was $\epsilon = 10^{-9}$.

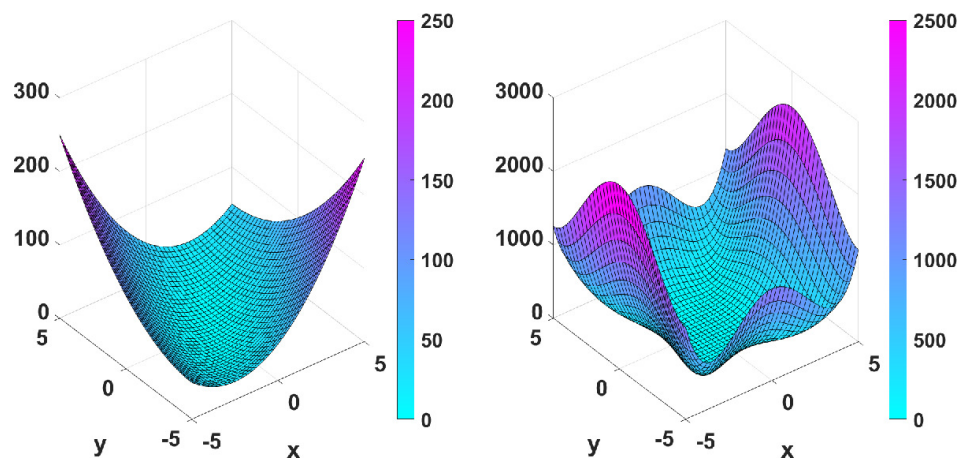


Figure 6. Graphical representations of $f_1(\Psi_1(X))$ and $f_1(\Psi_2(X))$.

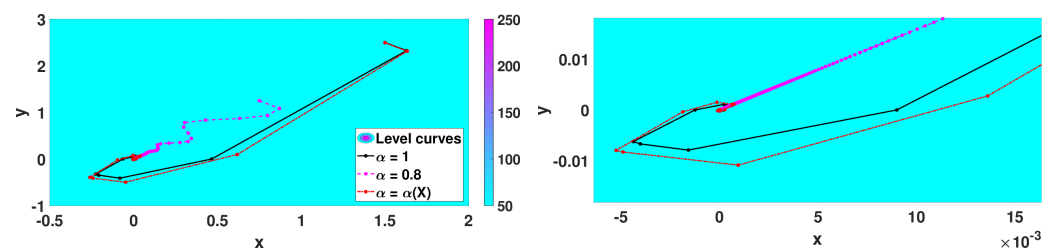


Figure 7. Iterates for $f_1(\Psi_1(X))$.

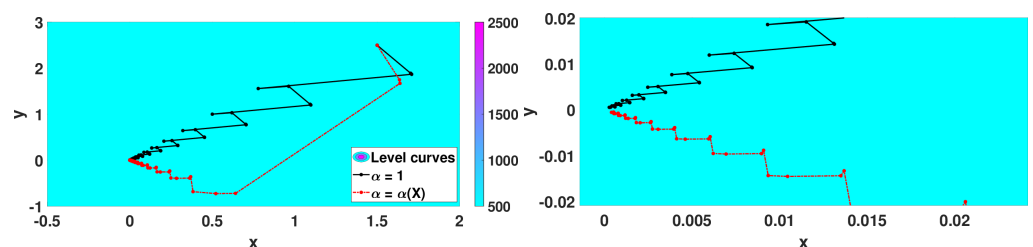


Figure 8. Iterates for $f_1(\Psi_2(X))$.

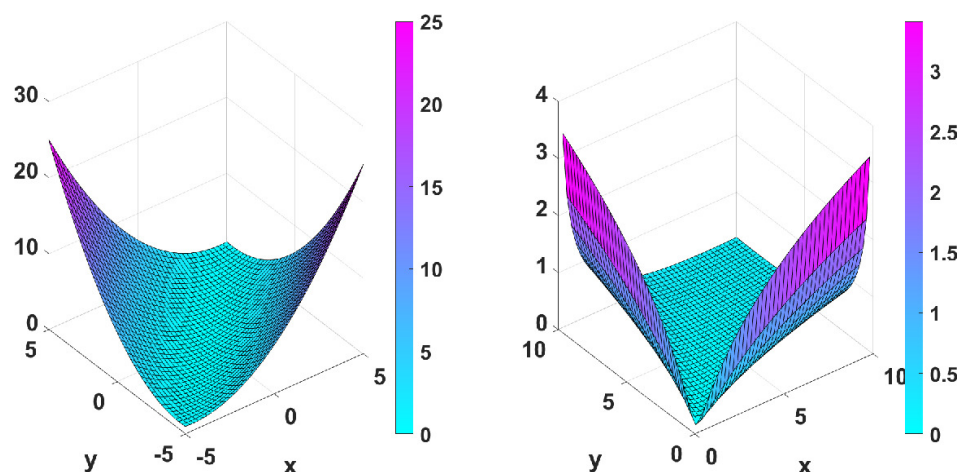
Table 1. Information about the k -iteration associated with Figures 7 and 8.

	α	θ	k	X_0	X_k	$\ X_k - X^*\ $
$f_1(\Psi_1(X))$						
Classical Gradient	1.0	optimized	49	[1.50, 2.50]	$[7.8188 \times 10^{-11}, 1.3031 \times 10^{-10}]$	1.5197×10^{10}
Algorithm 2	0.8	0.1	2480	[0.75, 1.25]	$[3.3860 \times 10^{-8}, 5.3902 \times 10^{-8}]$	6.3655×10^{-10}
Algorithm 3	variable	optimized	50	[1.50, 2.50]	$[9.3483 \times 10^{-11}, 1.4213 \times 10^{-10}]$	1.7012×10^{-10}
$f_1(\Psi_2(X))$						
Classical Gradient	1.0	optimized	77	[1.50, 2.50]	$[3.3370 \times 10^{-4}, 5.5617 \times 10^{-4}]$	6.4860×10^{-4}
Algorithm 2	0.8	0.1	divergence	[0.75, 1.25]	—	—
Algorithm 3	variable	optimized	73	[1.50, 2.50]	$[3.9399 \times 10^{-4}, -5.4845 \times 10^{-4}]$	6.7530×10^{-4}

In Table 1, we present the information concerning the X_k iterates of the implemented algorithms. When we consider Ψ_1 , we achieve the global minimum point in the three cases; however, it is clear that Algorithm 2 leads to the worst results in terms of speediness, whereas the classical case and Algorithm 3 have similar results. If we consider Ψ_2 and we restrict our analysis to the two fastest algorithms, we conclude that, in this case, Algorithm 3 provides a more accurate approximation in fewer iterations. We point out that the objective function $f_1(\Psi_2(\cdot))$ is a function with less convexity near the minimum point when compared with the objective function $f_1(\Psi_1(\cdot))$, which leads to an optimization problem that is more challenging under the numerical point of view.

In the next set of figures and tables, we consider the Matyas function to test our algorithms. More precisely, we consider the functions $f_2(\Psi_1(X)) = 0.26(x^2 + y^2) - 0.48xy$ with local minimum at the point $X^* = (0, 0)$, and $f_2(\Psi_3(X)) = 0.26(\ln^2(x) + \ln^2(y)) - 0.48 \ln(x) \ln(y)$ with local minimum at the point $X^* = (1, 1)$.

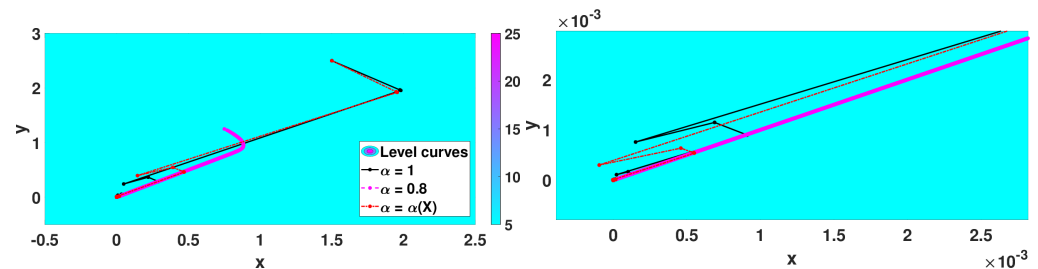
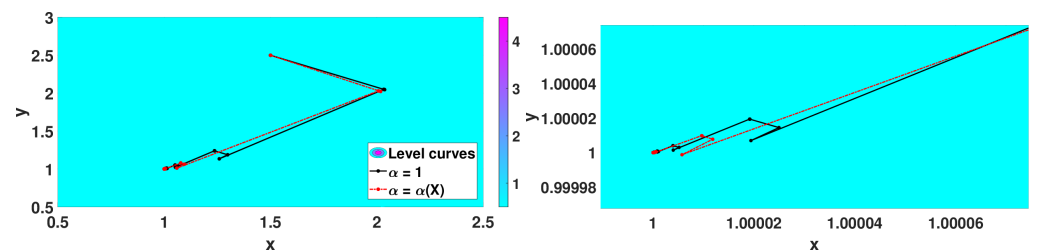
Figure 9 shows both functions $f_2(\Psi_1(\cdot))$ and $f_2(\Psi_3(\cdot))$ with a plate-like shape. Considering the same stopping criteria, we have the following results.

**Figure 9.** Graphical representations of $f_2(\Psi_1(X))$ and $f_2(\Psi_3(X))$.

From the analysis of Table 2, we see that the three methods converge in the case of $f_2(\Psi_1)$, but Algorithm 2 is the worst in terms of iterations. The classic case and Algorithm 3 have similar results in terms of precision; however, Algorithm 3 presents better performance in terms of the number of iterations. In the case of $f_2(\Psi_3)$, the Algorithm 2 diverges and the other two are convergent. Algorithm 3 required half of the iterations compared with the classical gradient method.

Table 2. Information about the k -iteration associated with Figures 10 and 11.

	α	θ	k	X_0	X_k	$\ X_k - X^*\ $
$f_2(\Psi_1(X))$						
Classical Gradient	1.0	optimized	43	[1.50, 2.50]	$[8.9458 \times 10^{-9}, 8.8320 \times 10^{-9}]$	1.2571×10^{-8}
Algorithm 2	0.8	1.0	94777	[0.75, 1.25]	$[1.7395 \times 10^{-6}, 1.7395 \times 10^{-6}]$	2.4600×10^{-6}
Algorithm 3	variable	optimized	27	[1.50, 2.50]	$[1.4056 \times 10^{-8}, 1.3811 \times 10^{-8}]$	1.9727×10^{-8}
$f_2(\Psi_3(X))$						
Classical Gradient	1.0	optimized	51	[1.50, 2.50]	[1, 1]	0.5×10^{-16}
Algorithm 2	0.8	0.1	divergence	[0.75, 1.25]	—	—
Algorithm 3	variable	optimized	29	[1.50, 2.50]	[1, 1]	0.5×10^{-6}

**Figure 10.** Iterates for $f_2(\Psi_1(X))$.**Figure 11.** Iterates for $f_2(\Psi_3(X))$.

In the final set of figures and tables, the function f_3 is composed with Ψ_1 and Ψ_2 . Taking into account the results obtained previously for the Matyas function, where it is clear that Algorithm 2 leads to worse results in terms of rapidness and accuracy, we only implemented the classical gradient method and Algorithm 3. The following figure shows the graphical representation of $f_3(\Psi_1(X)) = (x^6 + y^4 - 17)^2 + (2x + y - 4)^2$ with local minimum at the point $X^* = (1, 2)$ and $f_3(\Psi_2(X)) = (x^{12} + y^8 - 17)^2 + (2x^2 + y^2 - 4)^2$ with local maximum at $X^* = (0, 0)$ (or a local minimum of the function $-f_3(\Psi_2(X))$).

The plots in Figure 12 show that both functions are valley-shaped. Figures 13 and 14 and Table 3 show the obtained numerical results.

In this last case, we see that the classic gradient method and Algorithm 3 provide very good approximations. Algorithm 3 performs better in terms of the number of iterations. For instance, in the case of $f_3(\Psi_1)$, the number of iterations decreased around 97% in comparison with the classical gradient method.

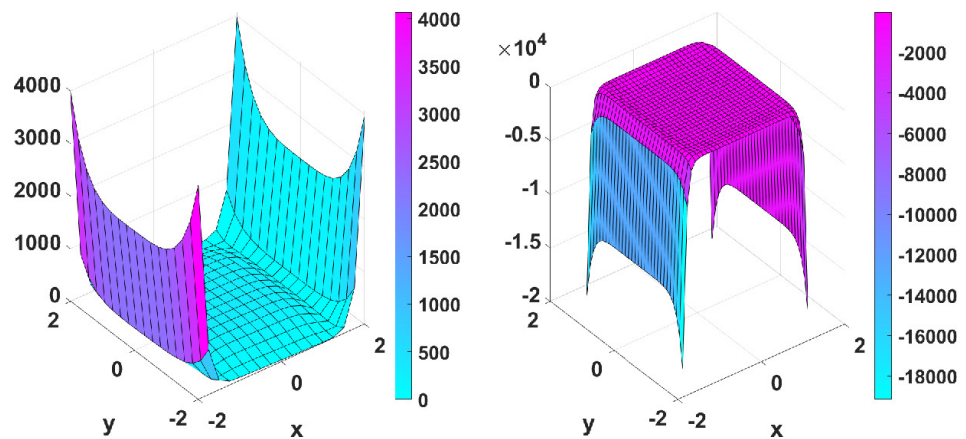


Figure 12. Graphical representations of $f_3(\Psi_1(X))$ and $-f_3(\Psi_2(X))$

Table 3. Information about the k -iteration associated with Figures 13 and 14.

	α	θ	k	X_0	X_k	$\ X_k - X^*\ $
$f_3(\Psi_1(X))$						
Classical Gradient	1.0	optimized	2923	[1.50, 2.50]	[1, 2]	0.5×10^{-16}
Algorithm 3	variable	optimized	71	[1.50, 2.50]	[1, 2]	1.378364×10^{-10}
$f_3(\Psi_2(X))$						
Classical Gradient	1.0	optimized	51	[0.50, 0.25]	$[1.6446 \times 10^{-12}, -1.3283 \times 10^{-11}]$	1.3385×10^{-11}
Algorithm 3	variable	optimized	39	[0.50, 0.25]	$[-1.0755 \times 10^{-12}, 2.1842 \times 10^{-11}]$	2.1868×10^{-11}

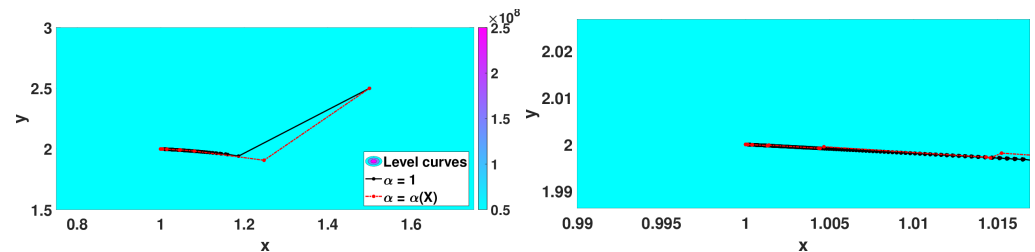


Figure 13. Iterates for $f_3(\Psi_1(X))$.

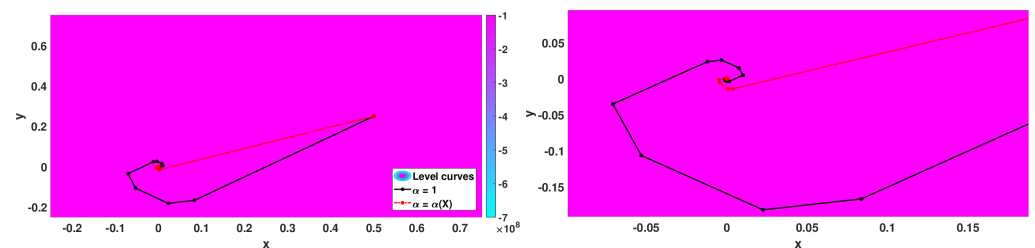


Figure 14. Iterates for $f_3(\Psi_2(X))$.

6. Conclusions

In this work, we study the classical gradient method from the perspective of the ψ -Hilfer fractional derivative. In the first part of the article, we consider the continuous gradient method and perform the convergence analysis for strongly and non-strongly convex cases. The identification of functions of the Lyapunov type, together with the auxiliary results demonstrated, allowed us to establish the convergence of the generating trajectories in the case of ψ -Hilfer.

In the second part of the paper, we first show that the ψ -Hilfer FGM with the ψ -Hilfer gradient given as a power series can converge to a point different from the extreme point. To

work out this problem, we propose an algorithm with a variable lower bound of integration, reducing the influence of long-time memory terms. By truncating the higher order terms, we obtain the ψ -FGM, which allows easy implementation in practice. Furthermore, we optimized the step size in each iteration and considered a variable order of differentiation to increase the precision and speed of the method. These two tunable parameters improved the performance of the method in terms of speed of convergence.

Our numerical simulations showed that the proposed FGM achieved the approximation with equal or better precision, but in much fewer iterations compared with the classical gradient method with optimized step size. We emphasize that in our 2D numerical simulations, the Matyas function and the Wayburn and Seader No. 1 functions are well-known benchmark functions used to test optimization methods. These functions have the shapes of plates and valleys, respectively, representing an extra challenge in numerical simulations.

In future works, it would be interesting to further develop this theory to see its application in the field of convolutional neural networks.

Author Contributions: Conceptualization, N.V., M.M.R. and M.F.; investigation, N.V., M.M.R. and M.F.; writing—original draft, N.V., M.M.R. and M.F.; writing—review and editing, N.V., M.M.R. and M.F. All authors have read and agreed to the published version of the manuscript.

Funding: The work of the authors was supported by Portuguese funds through the CIDMA (Center for Research and Development in Mathematics and Applications) and FCT (Fundação para a Ciência e a Tecnologia), within projects UIDB/04106/2020 and UIDP/04106/2020. N. Vieira was also supported by FCT via the 2018 FCT program of Stimulus of Scientific Employment—Individual Support (Ref: CEECIND/01131/2018).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Lin, J.Y.; Liao, C.W. New IIR filter-based adaptive algorithm in active noise control applications: Commutation error-introduced LMS algorithm and associated convergence assessment by a deterministic approach. *Automatica* **2008**, *44*, 2916–2922. [\[CrossRef\]](#)
2. Pu, Y.F.; Zhou, J.L.; Zhang, Y.; Zhang, N.; Huang, G.; Siarry, P. Fractional extreme value adaptive training method: Fractional steepest descent approach. *IEEE Trans. Neural Netw. Learn. Syst.* **2015**, *26*, 653–662. [\[CrossRef\]](#) [\[PubMed\]](#)
3. Ge, Z.W.; Ding, F.; Xu, L.; Alsaedi, A.; Hayat, T. Gradient-based iterative identification method for multivariate equation-error autoregressive moving average systems using the decomposition technique. *J. Frankl. Inst.* **2019**, *356*, 1658–1676. [\[CrossRef\]](#)
4. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444. [\[CrossRef\]](#) [\[PubMed\]](#)
5. Wong, C.C.; Chen, C.C. A hybrid clustering and gradient descent approach for fuzzy modeling. *IEEE Trans. Syst. Man Cybern. Part B Cybern.* **1999**, *29*, 686–693. [\[CrossRef\]](#) [\[PubMed\]](#)
6. Ren, Z.G.; Xu, C.; Zhou, Z.C.; Wu, Z.Z.; Chen, T.H. Boundary stabilization of a class of reaction-advection-diffusion systems via a gradient-based optimization approach. *J. Frankl. Inst.* **2019**, *356*, 173–195. [\[CrossRef\]](#)
7. Tan, Y.; He, Z.; Tian, B. A novel generalization of modified LMS algorithm to fractional order. *IEEE Signal Process. Lett.* **2015**, *22*, 1244–1248. [\[CrossRef\]](#)
8. Cheng, S.S.; Wei, Y.H.; Chen, Y.Q.; Li, Y.; Wang, Y. An innovative fractional order LMS based on variable initial value and gradient order. *Signal Process.* **2017**, *133*, 260–269. [\[CrossRef\]](#)
9. Raja, M.A.Z.; Chaudhary, N.I. Two-stage fractional least mean square identification algorithm for parameter estimation of CARMA systems. *Signal Process.* **2015**, *107*, 327–339. [\[CrossRef\]](#)
10. Shah, S.M.; Samar, R.; Khan, N.M.; Raja, M.A.Z. Design of fractional-order variants of complex LMS and NLMS algorithms for adaptive channel equalization. *Nonlinear Dyn.* **2017**, *88*, 839–858. [\[CrossRef\]](#)
11. Viera-Martin, E.; Gómez-Aguilar, J.F.; Solís-Pérez, J.E.; Hernández-Pérez, J.A.; Escobar-Jiménez, R.F. Artificial neural networks: A practical review of applications involving fractional calculus. *Eur. Phys. J. Spec. Top.* **2022**, *231*, 2059–2095. [\[CrossRef\]](#) [\[PubMed\]](#)
12. Sheng, D.; Wei, Y.; Chen, Y.; Wang, Y. Convolutional neural networks with fractional order gradient method. *Neurocomputing* **2020**, *408*, 42–50. [\[CrossRef\]](#)

13. Wang, Y.L.; Jahanshahi, H.; Bekiros, S.; Bezzina, F.; Chu, Y.M.; Aly, A.A. Deep recurrent neural networks with finite-time terminal sliding mode control for a chaotic fractional-order financial system with market confidence. *Chaos Solitons Fractals* **2021**, *146*, 110881. [\[CrossRef\]](#)
14. Chen, Y.Q.; Gao, Q.; Wei, Y.H.; Wang, Y. Study on fractional order gradient methods. *Appl. Math. Comput.* **2017**, *314*, 310–321. [\[CrossRef\]](#)
15. Wei, Y.; Kang, Y.; Yin, W.; Wang, Y.; Generalization of the gradient method with fractional order gradient direction. *J. Frankl. Inst.* **2020**, *357*, 2514–2532 [\[CrossRef\]](#)
16. Samko, S.G.; Kilbas, A.A.; Marichev, O.I. *Fractional Integrals and Derivatives: Theory and Applications*; Gordon and Breach: New York, NY, USA, 1993.
17. Kilbas, A.A.; Srivastava, H.M.; Trujillo, J.J. *Theory and Applications of Fractional Differential Equations*; North-Holland Mathematics Studies; Elsevier: Amsterdam, The Netherlands, 2006; Volume 204.
18. Almeida, R. A Caputo fractional derivative of a function with respect to another function. *Commun. Nonlinear Sci. Numer. Simul.* **2017**, *44*, 460–481. [\[CrossRef\]](#)
19. Sousa, J.V.C.; Oliveira, E.C.; On the ψ -Hilfer derivative. *Commun. Nonlinear Sci. Numer. Simulat.* **2018**, *60*, 72–91. [\[CrossRef\]](#)
20. Gorenflo, R.; Kilbas, A.A.; Mainardi, F.; Rogosin, S.V. *Mittag-Leffler Functions, Related Topics and Applications*; Springer: Berlin/Heidelberg, Germany, 2020.
21. Hai, P.V.; Rosenfeld, J.A. The gradient descent method from the perspective of fractional calculus. *Math. Meth. Appl. Sci.* **2021**, *44*, 5520–5547. [\[CrossRef\]](#)
22. Kucche, K.D.; Mali, A.D.; Sousa, J.V.C. On the nonlinear ψ -Hilfer fractional differential equations. *Comput. Appl. Math.* **2019**, *38*, 73. [\[CrossRef\]](#)

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.