THE GRADIENT DESCENT METHOD FROM THE PERSPECTIVE OF FRACTIONAL CALCULUS

PHAM VIET HAI AND JOEL A. ROSENFELD

Abstract. The main results of this manuscript lie at the intersection between optimization theory and fractional calculus. Motivated by gradient methods in optimization theory, we give methods based on $\psi$-fractional derivatives of order $\alpha$ in order to solve unconstrained optimization problems. The convergence of these methods is analyzed in detail. This paper also presents an Adams-Bashforth-Moulton (ABM) method for the estimation of solutions to equations involving $\psi$-fractional derivatives. Numerical examples using the ABM method show that the fractional order $\alpha$ and weight $\psi$ are tunable parameters, which can be helpful for improving the performance of gradient descent methods.

1. Introduction

1.1. Optimization Problems. Given a continuously, differentiable function $f : \mathbb{R}^d \to \mathbb{R}$, we are interested in solving the unconstrained optimization problem:

$$\min_{x \in \mathbb{R}^d} f(x).$$

From classical multivariate calculus, if the objective function $f(\cdot)$ attains its local minimum at $y_*$, then $y_*$ is a stationary point, that is

$$\nabla f(y_*) = 0.$$

Furthermore, if it is also assumes that $f(\cdot)$ is convex, then any stationary point must be a global minimum of $f(\cdot)$. Thus, solving Problem (1.1) may be reduced to the problem of finding stationary points. If the description of the gradient $\nabla f(\cdot)$ is simple, one can seek all stationary points and, among them, all global minima, by solving (1.2). However, if this description is complicated, as it usually is in practice, solving (1.2) may be difficult.

Gradient descent is one of the oldest and most fundamental first-order iterative algorithm for solving Problem (1.1). Low computational complexity makes gradient descent an ideal algorithm for very large-scale problems with medium accuracy. Gradient descent originated from the observation that if the objective function, $f(\cdot)$, is defined and differentiable in a neighborhood of a point $\omega$, then the direction of greatest decrease from $\omega$ is the negative gradient of $f$ at $\omega$. With this observation in mind, the gradient descent algorithm starts with a prediction $x_0$ for a local minimum of $f(\cdot)$, and constructs the sequence $x(0), x(1), x(2), \cdots$ such that

$$x(k + 1) = x(k) - \beta(k)\nabla f(x(k)), \quad x(0) = x_0 \in \mathbb{R}^d,$$

where the step size $\beta : \mathbb{Z}_{>0} \to \mathbb{R}_{>0}$ is allowed to vary at every iteration $k$. The result of the gradient descent algorithm is a monotonic sequence

$$f(x(0)) \geq f(x(1)) \geq f(x(2)) \geq \cdots$$

and it is expected that the sequence $\{x(k)\}_{k=0}^{\infty}$ converges to a local minimum. In practice, the stopping criterion is usually of the form $\|\nabla f(x)\| \leq \varepsilon$, and more details can be found in the monographs [1, 12, 14].

2010 Mathematics Subject Classification. Primary: 05C??, Secondary: 05C??

Key words and phrases. sample paper.
Rearranging (1.3) results in
\[
\frac{x(k+1) - x(k)}{\beta(k)} = -\nabla f(x(k)).
\]
Hence, (1.3) may be viewed as the discretization of the ordinary differential equation (1.4)
\[
y'(t) = -\nabla f(y(t)), \quad y(0) = y_0 \in \mathbb{R}^d,
\]
by the explicit Euler scheme with step size \(\beta(k)\). The system (1.4) is called continuous gradient method. The analogy between difference equations and differential equations has been recognized and exploited very regularly. Many results related to difference equations can carry over quite easily to corresponding results for differential equations and vice versa. For instance, solutions of equations (1.3) and (1.4) converge to the unique stationary point at the exponential rate under the assumption that \(f(\cdot)\) is both strongly convex and smooth (14, 16). Furthermore, if a convergence result is proved for a continuous method, then various finite difference schemes for the solution of this Cauchy problem may be constructed.

The present manuscript employs methods from fractional order calculus to improve the convergence rate of continuous time optimization methods. Improvements in convergence rates over continuous time gradient methods using classical calculus are realized through Mittag-Leffler convergence rates using \(\psi\)-fractional derivatives.

1.2. Fractional calculus. The fractional calculus (FC) may be viewed an old and yet novel topic. It can be said an old topic because the first inquiry into the meaning of a fractional derivative is found in a 1965 letter written to L’Hospital by Leibniz about the meaning for derivative of order \(\frac{1}{2}\) of the power function. However, it may be regarded a novel topic as well due to its important applications in engineering, physics and bio-engineering, and other applied sciences. Several monographs and surveys describe the progress in the area of FC. Prominent are the books [6, 9] as well as the survey [17], which includes a historical review and notes concerning scientists that contributed to the development of FC.

The following are distinct approaches to FC by Riemann-Liouville and Hadamard.

**Definition 1.1.** Given \(0 < \alpha \leq 1\), the Riemann-Liouville integral of a function \(x\) is defined as
\[
I^\alpha_a x(t) \triangleq \frac{1}{\Gamma(\alpha)} \int_a^t \frac{(t-\tau)^{\alpha-1} x(\tau)}{\tau} d\tau,
\]
where \(\Gamma(\alpha) \triangleq \int_0^\infty \tau^{\alpha-1} e^{-\tau} d\tau\). The Riemann-Liouville and Caputo fractional derivatives are defined as, respectively
\[
RLD^\alpha_a \triangleq \frac{d}{dt} \circ I^{1-\alpha}_a, \quad \text{and} \quad CD^\alpha_a \triangleq I^{1-\alpha}_a \circ \frac{d}{dt}.
\]

**Definition 1.2.** Given \(0 < \alpha \leq 1\), the Hadamard integral of a function \(x\) is defined as
\[
H^\alpha_1 x(t) \triangleq \frac{1}{\Gamma(\alpha)} \int_1^t \left( \frac{t}{\tau} \right)^{\alpha-1} x(\tau) \frac{d\tau}{\tau},
\]
The Hadamard fractional derivative is defined as
\[
H^\alpha_1 \triangleq \frac{d}{dt} \circ H^{1-\alpha}_1.
\]

In [10], Liang et al. proposed a fractional differential equation, which generalizes equation (1.4). Specifically, these authors replaced the usual derivative \(y'(\cdot)\) in (1.3) with Caputo fractional derivatives \(CD^\alpha_a\), where \(0 < \alpha < 1\). The fractional order is a adjustable parameter, which can be helpful for improving the performance. Numerical examples in [10] reveal that this fractional differential equation may possess faster or slower convergence rate than
equation (1.4), depending on specific problems. We emphasize that the paper [11] deals only with Caputo fractional derivatives and the cases of Riemann-Liouville or Hadamard are not considered.

1.3. Aim and Content. In recent years, several concepts of fractional derivatives have been proposed, studied and applied to practical problems. In [2], the definition of the \( \psi \)-fractional derivatives unifies a large class of fractional derivatives. Leveraging the \( \psi \)-derivative, the present manuscript provides fractional differential equations that resolves a class of unconstrained optimization problems. Significantly, the convergence rate of these differential equations is analyzed in Section 4 and Section 5. Specifically, convergence rates outperforming integer order gradient descent methods are obtained in Theorem 5.2 through the selection of particular \( \psi \). In the case of non-strongly convex \( f(\cdot) \), a convergence rate of \( O(\psi(t)^{-\lambda}) \) is obtained. Subsequently, under the assumption on the strong convexity, Mittag-Leffler convergence, a general type of exponential convergence, is demonstrated, and conditions when an exponential convergence occurs are also established.

The rest of the paper is organized as follows. Section 2 is devoted to recalling basic knowledge on convex analysis and \( \psi \)-fractional derivatives. What makes these \( \psi \)-fractional derivatives interesting is the fact that they are really generalizations of the well-known concepts, including the ones in Definition 1.1 and 1.2. Section 3 contains several technical observations which will be later on referred to, including chain rule, Jensen-type inequality. Motivated by gradient methods in optimization theory, in Section 4 we give fractional differential equations of Riemann-Liouville type in order to solve unconstrained optimization problems. In parallel, we do study fractional differential equations of Caputo type in Section 5. Section 6 gives a generalization of the ABM method of [6] for \( \psi \)-Caputo derivatives, and this numerical method is utilized to implement the optimization procedures of this manuscript. The paper concludes with suggestions of future research in Section 7.

1.4. Notations. Throughout the paper, we denote by \( \mathbb{Z}, \mathbb{R}, \mathbb{C} \) by the set of integers, real numbers, complex numbers, respectively. For a set \( A \subseteq \mathbb{R} \), \( A_{\geq \delta} \) stands for the set \( \{ x \in A : x \geq \delta \} \). Let \( \mathbb{R}^d \) be the set of real \( d \)-dimensional vectors endowed with the Euclidean inner product \( \langle \cdot, \cdot \rangle \) and the standard Euclidean norm \( \| \cdot \| \). The gradient of the function \( f(\cdot) \) at \( x \) is denoted as \( \nabla f(x) \). For two symmetric matrices \( A, B \in \mathbb{R}^{d \times d} \), writing \( A \preceq B \) means that \( \langle Ax, x \rangle \leq \langle Bx, x \rangle \) for all \( x \in \mathbb{R}^d \).

2. Preliminaries

2.1. Convex analysis. This section provides an exposition on convex functions that will be leveraged in the sequel.

**Definition 2.1.** A function \( f : \mathbb{R}^d \to \mathbb{R} \) is of class \( C^m \) for \( m \in \mathbb{Z}_{\geq 0} \) (written \( f(\cdot) \in C^m \)) if the partial derivatives \[
\frac{\partial^{m_1 + \cdots + m_d} f}{\partial x_1^{m_1} \cdots \partial x_d^{m_d}}
\] all exist and are continuous for every selection \( m_1, \ldots, m_d \) with \( m_1 + \cdots + m_d \leq m \).

Throughout the manuscript several assumptions concerning the objective function will be used, including Lipschitz conditions on the gradient, which may be achieved by bounding the Hessian of a twice differentiable function, convexity of the objective function, and a uniform lower bound on the Hessian. These assumptions are summarized in Assumption 2.1. Assumption 2.2 is also included to ensure that the optimization problem is well posed.

**Assumption 2.1.** The function \( f : \mathbb{R}^d \to \mathbb{R} \) is

1. of class \( C^1 \) and \( L_f \)-smooth, i.e.
   \[ \| \nabla f(x) - \nabla f(y) \| \leq L_f \| x - y \|, \; \forall x, y \in \mathbb{R}^d. \]
(2) of class $C^1$ and convex, i.e.,
$$f(x) \geq f(y) + \langle \nabla f(y), x - y \rangle, \quad \forall x, y \in \mathbb{R}^d,$$

(3) of class $C^2$ and there exists a constant $m_f > 0$ such that
$$m_f I_d \leq \nabla^2 f(x), \quad \forall x \in \mathbb{R}^d.$$

**Assumption 2.2.** The set $S(f) \triangleq \{z \in \mathbb{R}^d : \nabla f(z) = 0\} \neq \emptyset$.

**Definition 2.2.** A function $f \in C^1$ is strongly convex with parameter $m_f > 0$ if
$$f(x) \geq f(y) + \langle \nabla f(y), x - y \rangle + \frac{m_f}{2} \|x - y\|^2, \quad \forall x, y \in \mathbb{R}^d.$$

Convexity has several useful consequences.

**Lemma 2.3** ([2] Ineq. (2.1.7)-(2.1.9)). If the function $f(\cdot)$ is convex and $M_f$-smooth, then
$$f(x) + \langle \nabla f(x), y - x \rangle + \frac{1}{2M_f} \|\nabla f(x) - \nabla f(y)\|^2 \leq f(y), \quad \forall x, y \in \mathbb{R}^d,$$

$$\langle \nabla f(x) - \nabla f(y), x - y \rangle \geq \frac{1}{M_f} \|\nabla f(x) - \nabla f(y)\|^2, \quad \forall x, y \in \mathbb{R}^d,$$

and
$$\langle \nabla f(x) - \nabla f(y), x - y \rangle \leq L \|x - y\|^2, \quad \forall x, y \in \mathbb{R}^d.$$

**2.2. Fractional calculus.** In this section, some basic definitions and techniques related to $\psi$-fractional calculus are presented.

**Assumption 2.3.** The function $\psi : \mathbb{R}_{\geq \ell} \to \mathbb{R}_{\geq 0}$ is

1. strictly increasing, of class $C^1$ with $\psi' > 0$.
2. $\sup\{\psi(t) : t \geq \ell\} = \infty$.
3. (of class $C^2$ with $\psi'' \geq 0$.

To simplify notation, we denote
$$J_\psi \triangleq \frac{1}{\psi'(t)} \cdot \frac{d}{dt}.$$

**Definition 2.4** ([2]). Let $0 < \alpha \leq 1$. The $\psi$-Riemann-Liouville integral of a function $x$ is defined by
$$\left( I^\alpha_{a,\psi} x \right)(t) \triangleq \frac{1}{\Gamma(\alpha)} \int_a^t (\psi(t) - \psi(\tau))^{\alpha-1} \psi'(\tau) x(\tau) \, d\tau.$$

The $\psi$-Riemann-Liouville and $\psi$-Caputo fractional derivatives are defined as, respectively
$$R^{\alpha}_{a,\psi} \triangleq J_\psi \circ I^1_{a,\psi} \quad \text{and} \quad C^{\alpha}_{a,\psi} \triangleq I^1_{a,\psi} \circ J_\psi.$$

**Definition 2.5** ([2]). The $\psi$-Riemann-Liouville and $\psi$-Caputo fractional derivatives (or simply: $\psi$-R-L and $\psi$-C derivatives, respectively) of a $d$-dimensional vector function $x(t) = (x_1(t), \ldots, x_d(t))^T$ are defined component-wise as
$$R^{\alpha}_{a,\psi} x(t) \triangleq \left( R^{\alpha}_{a,\psi} x_1(t), \ldots, R^{\alpha}_{a,\psi} x_d(t) \right)^T,$$
$$C^{\alpha}_{a,\psi} x(t) \triangleq \left( C^{\alpha}_{a,\psi} x_1(t), \ldots, C^{\alpha}_{a,\psi} x_d(t) \right)^T,$$
respectively.

**Remark 2.6.** Several well known fractional derivatives manifest as particular cases of $\psi$-fractional derivatives. For example, given appropriate choices of the kernel, $\psi$, we obtain the Caputo fractional derivative (when $\psi(t) = t$) and Hadamard fractional derivative (when $\psi(t) = \ln t$).
In [3], Almeida et al. investigated fractional differential equations with the \( \psi \)-Caputo derivative:

\[
C^\alpha_{a,\psi} x(t) = g(t, x(t)), \quad t \geq a, \quad x(a) = x_0,
\]

where \( \alpha \in (0, 1] \), and \( g : \mathbb{R}_{\geq 0} \times \mathbb{R}^d \to \mathbb{R}^d \) is a continuous vector-valued function. The results regarding the existence and uniqueness of solutions of (2.5) were established by using fixed point theorems, which agrees with the approach for establishing existence and uniqueness in the classical setting.

**Definition 2.7.** A continuous function \( x : \mathbb{R}_{\geq 0} \to \mathbb{R}^d \) is called a solution of equation (2.5) if it satisfies this equation for every \( t \geq a \). In this case, \( x_0 \) is called the initial value of the solution \( x(\cdot) \).

A closely related fractional differential equation to (2.5) arises from the Riemann-Liouville setting,

\[
RLD^\alpha_{a,\psi} y(t) = g(t, y(t)), \quad t \geq a, \quad I^{1-\alpha}_{a,\psi} y(a+) = y_0,
\]

where a function is called a solution to (2.6) if it satisfies the equation for all \( t > a \) as was the case for (2.5).

**Proposition 2.8.** Let \( 0 < \alpha \leq 1 \). Then the following assertions hold.

1. A continuous function \( x(\cdot) \) is a solution of equation (2.5) if and only if it satisfies

\[
x(t) = x(a) + \frac{1}{\Gamma(\alpha)} \int_a^t (\psi(t) - \psi(s))^{\alpha-1} \psi'(s) g(s, x(s)) \, ds, \quad t \geq a.
\]

2. A continuous function \( y(\cdot) \) is a solution of equation (2.6) if and only if it satisfies

\[
y(t) = y_0 \frac{\psi(t) - \psi(a))^{\alpha-1}}{\Gamma(\alpha)} + \frac{1}{\Gamma(\alpha)} \int_a^t (\psi(t) - \psi(s))^{\alpha-1} \psi'(s) g(s, y(s)) \, ds, \quad t \geq a.
\]

**Proof.** The first item was proved in [3] Theorem 2]. Meanwhile, the second one can be obtained by using

\[
I^\alpha_{\ell,\psi} \circ RL D^\alpha_{\ell,\psi} y(t) = y(t) - \frac{1}{\Gamma(\alpha)} I^{1-\alpha}_{\ell,\psi} y(a+)[\psi(t) - \psi(t)]^{\alpha-1}.
\]

\( \square \)

Similar to the exponential function frequently used in the solutions of integer-order systems, a function frequently used in fractional calculus is the Mittag-Leffler function defined as

\[
\mathcal{E}_{\alpha,\beta}(z) = \sum_{j \geq 0} \frac{z^j}{\Gamma(j\alpha + \beta)}, \quad z \in \mathbb{C},
\]

where \( \alpha, \beta > 0 \). In order to simplify notation, we write \( \mathcal{E}_\alpha(\cdot) \) in stead of \( \mathcal{E}_{\alpha,1} \). The following propositions gather some well-known inequalities on fractional calculus, which are required in later proofs.

**Proposition 2.9** ([12]). The function \( \mathcal{E}_{\alpha,\beta}(-t) \) is completely monotone, that is

\[
(-1)^m \frac{d^m}{dt^m} \mathcal{E}_{\alpha,\beta}(-t) \geq 0, \quad \forall t \geq 0, \forall m \in \mathbb{Z}_{\geq 0},
\]

if and only if \( 0 < \alpha \leq 1, \beta \geq \alpha \).

**Proposition 2.10** ([3] Theorem 4.5]). Let \( \alpha > 0 \) and \( \gamma \in \mathbb{C} \). The limit

\[
\lim_{t \to \infty} \mathcal{E}_\alpha(-\gamma t^\alpha) = 0
\]

holds if \( |\arg(\gamma)| < \alpha \pi/2 \).
3. Auxiliary results

This section contains several technical observations which will be later on referred to.

3.1. Chain rule. In this subsection, we show that the fractional-order derivative of a composite function is different from its integer-order derivative.

Proposition 3.1. Let \( \alpha \in (0, 1] \) and \( \ell \in \mathbb{R}_{\geq 0} \). For \( t \geq \ell \), let us define the function \( \zeta_t \) by setting

\[
\zeta_t(s) \triangleq g(f(s)) - g(f(t)) - \left\langle \nabla g(f(t)), f(s) - f(t) \right\rangle.
\]

Then the following identity holds for every \( t \geq \ell \)

\[
\Gamma(1 - \alpha) \left( CD_{\ell,\psi}^{\alpha} g(f(t)) - \left\langle \nabla g(f(t)), CD_{\ell,\psi}^{\alpha} f(t) \right\rangle \right) = -\frac{\psi'(\ell)\zeta_\ell(\ell)}{(\psi(t) - \psi(\ell))^\alpha} - \int_\ell^t \zeta_t(s) \left[ \frac{\psi''(s)}{(\psi(t) - \psi(s))^{\alpha}} + \frac{\alpha\psi'(s)^2}{(\psi(t) - \psi(s))^{\alpha+1}} \right] ds.
\]

Consequently, if the function \( g(\cdot) \) is convex and if the function \( \psi(\cdot) \) satisfies Assumptions 2.3(1,3), then we obtain the following chain rule.

\[
CD_{\ell,\psi}^{\alpha} g(f(t)) \leq \left\langle \nabla g(f(t)), CD_{\ell,\psi}^{\alpha} f(t) \right\rangle.
\]

If the function \( g(\cdot) \) is concave (i.e. \(-g(\cdot) \) is convex) and if the function \( \psi(\cdot) \) satisfies Assumptions 2.3(1,3), then

\[
CD_{\ell,\psi}^{\alpha} g(f(t)) \geq \left\langle \nabla g(f(t)), CD_{\ell,\psi}^{\alpha} f(t) \right\rangle.
\]

Proof. Taking into account the definition of \( CD_{\ell,\psi}^{\alpha} \), we can write

\[
\left. \Gamma(1 - \alpha) \left( CD_{\ell,\psi}^{\alpha} g(f(t)) - \left\langle \nabla g(f(t)), CD_{\ell,\psi}^{\alpha} f(t) \right\rangle \right) \right|_{t}^{\ell} = \int_\ell^t (\psi(t) - \psi(s))^{-\alpha} \psi'(s) \left\langle \nabla g(f(s)) - \nabla g(f(t)), f'(s) \right\rangle ds
\]

\[
= \int_\ell^t (\psi(t) - \psi(s))^{-\alpha} \psi'(s) \zeta'_t(s) ds
\]

\[
= \int_\ell^t (\psi(t) - \psi(s))^{-\alpha} \psi'(s) d\zeta_t(s),
\]

which implies, by integration parts, that

\[
\left. \Gamma(1 - \alpha) \left( CD_{\ell,\psi}^{\alpha} g(f(t)) - \left\langle \nabla g(f(t)), CD_{\ell,\psi}^{\alpha} f(t) \right\rangle \right) \right|_{t}^{\ell} = \lim_{s \to t} (\psi(t) - \psi(s))^{-\alpha} \psi'(s) \zeta_\ell(s) - \left\langle \psi(t) - \psi(\ell), -\alpha \psi'(s) \zeta'_\ell(t) \right\rangle
\]

\[
- \int_\ell^t \zeta'_t(s) [(\psi(t) - \psi(s))^{-\alpha} \psi''(s) + \alpha \psi'(s)^2 (\psi(t) - \psi(s))^{-\alpha-1}] ds.
\]

Thus, we obtain the desired result because

\[
\lim_{s \to t} (\psi(t) - \psi(s))^{-\alpha} \zeta_t(s) = \lim_{s \to t} \frac{\zeta_t(s)}{(\psi(t) - \psi(s))^{\alpha}} = \lim_{s \to t} -\alpha \psi'(s) (\psi(t) - \psi(s))^{\alpha-1} \quad (\text{by L'Hospital's Rule})
\]

\[= 0 \quad (\text{as } 0 < \alpha \leq 1).
\]

\[\square\]
Remark 3.2. Proposition 3.1 generalises existing inequalities in \[1\] and contains both existing works as special cases. Specially, if \( \psi(t) = t \), Proposition 3.1 is reduced to the original inequality in \[3\] Theorem 1.

**Proposition 3.3.** Let \( \alpha \in (0, 1] \) and \( b \in \mathbb{R}_{\geq 0} \). For \( t \geq b \), let \( \zeta_t \) be the function given in \[4,1\]. Then the following identity holds for every \( t \geq b \)

\[
\frac{\Gamma(1 - \alpha)}{b} \left\{ \int_{b}^{t} \psi(s) - \psi(t) \right\} - \alpha \int_{b}^{t} \psi(s)^{-\alpha-1} \psi'(s) \zeta_t(s) \, ds.
\]

**Proof.** By the Newton-Leibniz formula, one has

\[
f(t) = f(b) + \int_{b}^{t} f'(s) \, ds = f(b) + I_{b, t}^{1} (f'/\psi') \, (t),
\]

and so by \[3\] Corollary 1

\[
RL_{b, t}^{\alpha} f(t) = RL_{b, t}^{\alpha} [f(b)](t) + RL_{b, t}^{\alpha} \circ I_{b, t}^{1} (f'/\psi') \, (t)
\]

\[
= \frac{f(b)}{\Gamma(1 - \alpha)} \left[ \psi(t) - \psi(b) \right]^{-\alpha} + I_{b, t}^{1 - \alpha} (f'/\psi') \, (t)
\]

\[
= \frac{1}{\Gamma(1 - \alpha)} \left[ f(b) \left[ \psi(t) - \psi(b) \right]^{-\alpha} + \int_{b}^{t} \psi(s)^{-\alpha} f'(s) \, ds \right].
\]

Hence, we can write

\[
\frac{\Gamma(1 - \alpha)}{b} \left\{ \int_{b}^{t} \psi(s) - \psi(t) \right\} - \alpha \int_{b}^{t} \psi(s)^{-\alpha-1} \psi'(s) \zeta_t(s) \, ds
\]

which implies, by integration parts, that

\[
\frac{\Gamma(1 - \alpha)}{b} \left\{ \int_{b}^{t} \psi(s) - \psi(t) \right\} - \alpha \int_{b}^{t} \psi(s)^{-\alpha-1} \psi'(s) \zeta_t(s) \, ds
\]

\[
= \psi(b) \left[ \psi(t) - \psi(b) \right]^{-\alpha} \left[ \psi(b) - \psi(t) \right] \right\} \psi(t) - \psi(b) \right]^{-\alpha-1} \psi'(s) \zeta_t(s) \, ds
\]

\[
= \psi(b) \left[ \psi(t) - \psi(b) \right]^{-\alpha} \left[ \psi(b) - \psi(t) \right] \right\} \psi(t) - \psi(b) \right]^{-\alpha-1} \psi'(s) \zeta_t(s) \, ds
\]

\[
= \psi(b) \left[ \psi(t) - \psi(b) \right]^{-\alpha} \left[ \psi(b) - \psi(t) \right] \right\} \psi(t) - \psi(b) \right]^{-\alpha-1} \psi'(s) \zeta_t(s) \, ds.
\]

\[
\square
\]
Based on Proposition 3.3, we can directly obtain the following corollary.

**Corollary 3.4.** Let \( \alpha \in (0, 1] \), \( b \in \mathbb{R}_{>0} \) and \( \psi \) be a function satisfying Assumption 2.3(1). If the function \( g \) satisfies Assumption 2.7(2), then

\[
RLD^\alpha_{b, \psi} g(f(t)) \leq \langle \nabla g(f(t)), RLD^\alpha_{b, \psi} f(t) \rangle + \frac{g(0)}{\Gamma(1 - \alpha)}(\psi(t) - \psi(b))^{-\alpha}.
\]

**Proof.** It follows from Assumption 2.1(2), that

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

Hence, we can use Proposition 3.3 to get the desired result. \( \square \)

**Remark 3.5.** Proposition 3.3 generalises existing inequalities in [11] and contains both existing works as its special cases. Specifically, if \( g(x) = \frac{||x||^2}{2} \) and \( \psi(t) = t \), Proposition 3.3 is reduced to the original inequality in [11] Theorem 1.

### 3.2. Jensen-type inequality

Denote \([c, \ell]_d \triangleq [c, \ell] \times \cdots \times [c, \ell] \) \((d \text{ times})\). The following result can be viewed as the Jensen-type inequality for fractional integral.

**Proposition 3.6.** Let \( \varphi : [c, \ell]_d \rightarrow \mathbb{R} \) be a convex function and \( h : [a, b] \rightarrow [c, \ell]_d \). Then

\[
\varphi\left( \frac{\Gamma(\alpha + 1)}{(\psi(t) - \psi(a))^\alpha} \cdot (I^\alpha_{a, \psi} h)(t) \right) \leq \frac{\Gamma(\alpha + 1)}{(\psi(t) - \psi(a))^\alpha} \cdot (I^\alpha_{a, \psi} \varphi \circ h)(t).
\]

**Proof.** Let \( h = (h_1, \cdots, h_d) \). By the assumption,

\[
c \leq h_j(s) \leq \ell, \quad \forall s \in [a, b], \forall j \in \{1, \cdots, d\}.
\]

Multiplying the above inequality by \((\psi(t) - \psi(s))^{\alpha - 1} \psi'(s)/\Gamma(\alpha)\) and integrating the resulting inequality with respect to \( s \) over \([a, t]\), we obtain

\[
\frac{c}{\Gamma(\alpha)} \int_a^t (\psi(t) - \psi(s))^{\alpha - 1} \psi'(s) \, ds \leq (I^\alpha_{a, \psi} h_j)(t) \leq \frac{\ell}{\Gamma(\alpha)} \int_a^t (\psi(t) - \psi(s))^{\alpha - 1} \psi'(s) \, ds.
\]

Subsequently,

\[
c \leq \frac{\Gamma(\alpha + 1)}{(\psi(t) - \psi(a))^\alpha} \cdot (I^\alpha_{a, \psi} h_j)(t) \leq \ell, \quad \forall a \leq t \leq b, \forall j \in \{1, \cdots, d\}.
\]

From the convexity of \( \varphi \), we obtain

\[
\varphi(h(t)) \geq \varphi\left( \frac{\Gamma(\alpha + 1)}{(\psi(t) - \psi(a))^\alpha} \cdot (I^\alpha_{a, \psi} h)(\tau) \right) + \left( h(t) - \frac{\Gamma(\alpha + 1)}{(\psi(t) - \psi(a))^\alpha} \cdot (I^\alpha_{a, \psi} h)(\tau) \right) \varphi'\left( \frac{\Gamma(\alpha + 1)}{(\psi(t) - \psi(a))^\alpha} \cdot (I^\alpha_{a, \psi} h)(\tau) \right).
\]

Multiplying the above inequality by \((\psi(\tau) - \psi(t))^{\alpha - 1} \psi'(t)/\Gamma(\alpha)\) and integrating the resulting inequality with respect to \( t \) over \([a, \tau]\), we obtain

\[
(I^\alpha_{a, \psi} \varphi \circ h)(\tau) \geq \frac{(\psi(\tau) - \psi(a))^\alpha}{\Gamma(\alpha + 1)} \cdot \varphi\left( \frac{\Gamma(\alpha + 1)}{(\psi(\tau) - \psi(a))^\alpha} \cdot (I^\alpha_{a, \psi} h)(\tau) \right) + \left( (I^\alpha_{a, \psi} h)(\tau) - (I^\alpha_{a, \psi} h)(\tau) \right) \varphi'\left( \frac{\Gamma(\alpha + 1)}{(\psi(\tau) - \psi(a))^\alpha} \cdot (I^\alpha_{a, \psi} h)(\tau) \right)
\]

\[
= \frac{(\psi(\tau) - \psi(a))^\alpha}{\Gamma(\alpha + 1)} \cdot \varphi\left( \frac{\Gamma(\alpha + 1)}{(\psi(\tau) - \psi(a))^\alpha} \cdot (I^\alpha_{a, \psi} h)(\tau) \right).
\]

\( \square \)
3.3. Mittag-Leffler function. The following result will be leveraged in the proofs of Theorems 4.3 and 5.3.

Lemma 3.7. Let $\alpha \in (0, 1]$ and $\ell \geq 0$. Suppose that the function $\psi : \mathbb{R}_{\geq \ell} \to \mathbb{R}_{\geq 0}$ satisfies Assumptions 2.3(1-2). Then the limit

$$\lim_{t \to \infty} \int_{\ell}^{t} (\psi(t) - \psi(s))^{\alpha-1} \psi'(s) E_{\alpha}(-L(\psi(s) - \psi(\ell))^\alpha) ds = \frac{\Gamma(\alpha)}{L}$$

holds.

Proof. It was indicated in $[2$, Lemma 2], that the function $E_{\alpha}(-L(\psi(\cdot) - \psi(\ell))^\alpha)$ is the solution of the initial value problem

$$\frac{d}{dt} E_{\alpha}(-L(\psi(\cdot) - \psi(\ell))^\alpha) = -L E_{\alpha}(-L(\psi(\cdot) - \psi(\ell))^\alpha) \quad \text{for} \quad t \geq \ell,$$

and hence we make use of Proposition 2.8 in order to get the desired result. \qed

4. Gradient methods with Riemann-Liouville type fractional derivatives

In this subsection, we study

$$RLD_{\ell, \psi}^\alpha z(t) = -\beta \nabla f(z(t)), \quad (I_{\ell, \psi}^{1-\alpha} z)(t^+) = z_0 \in \mathbb{R}^d, \quad \forall t \geq \ell.$$ \hspace{1cm} (4.1)

For $y_\ast \in S(f)$, let us define the function $\varphi : \mathbb{R}_{\geq \ell} \to \mathbb{R}_{\geq 0}$ by setting

$$\varphi(t) = \frac{1}{2} \| z(t) - y_\ast \|^2, \quad t \geq \ell.$$ \hspace{1cm} (4.2)

4.1. The case of convex $f(\cdot)$. This subsection is devoted to investigating equation (4.1) under the assumption on the non-strongly convexity of $f(\cdot)$.

Theorem 4.1. Let $\alpha \in (0, 1]$. Suppose that the function $f : \mathbb{R}^d \to \mathbb{R}$ satisfies Assumption 2.3(2) and $\psi : \mathbb{R}_{\geq \ell} \to \mathbb{R}_{\geq 0}$ satisfies Assumption 2.3(1). Consider the fractional-order differential equation (4.1), where the step size $\beta$ is a constant. Then we have the following conclusions.

(1) There exists a constant $C$ such that

$$\| z(t) - y_\ast \|^2 \leq C[\psi(t) - \psi(\ell)]^{\alpha-1}, \quad \forall t \geq \ell.$$ \hspace{1cm} (4.3)

(2) If $\alpha = 1$, then $z(\cdot)$ is bounded.

Proof. By Proposition 5.4 and Assumption 2.3(2), we have

$$RLD_{\ell, \psi}^\alpha \varphi(t) \leq -\beta (z(t) - y_\ast, \nabla f(z(t))) \leq \beta [f_\ast - f(z(t))] \leq 0, \quad \forall t \in \mathbb{R}_{\geq \ell}$$

and so by Assumption 2.3(1),

$$0 \geq \beta I_{\ell, \psi}^\alpha (f_\ast - f \circ z)(t) \geq I_{\ell, \psi}^\alpha \circ RLD_{\ell, \psi}^\alpha \varphi(t) = \varphi(t) - \frac{1}{\Gamma(\alpha)} I_{\ell, \psi}^{1-\alpha} \varphi(\ell^+) [\psi(t) - \psi(\ell)]^{\alpha-1},$$

where in the last equality we use $[3$ Theorem 2.6]. \qed
4.2. The case of strongly convex \( f(\cdot) \). It turns out under the assumption of strong convexity, solutions of equation (4.1) admit Mittag-Leffler convergence, which is a general type of exponential convergence to a stationary point.

**Theorem 4.2.** Let \( \alpha \in (0, 1] \). Suppose that the function \( f : \mathbb{R}^d \to \mathbb{R} \) satisfies Assumption \( 2.2(3) \) and \( \psi : \mathbb{R}_{\geq 1} \to \mathbb{R}_{\geq 0} \) satisfies Assumption \( 2.3(1) \). Consider the fractional-order differential equation (4.1), where the step size \( \beta \) is a constant. Then we have the following conclusions.

1. The solution \( z(\cdot) \) converges to \( y_\ast \) with rate:
   \[
   \|z(t) - y_\ast\|^2 \leq \varphi_0(\psi(t) - \psi(\ell))^{\alpha - 1} \mathcal{E}_{\alpha, \alpha}(-\beta m_f(\psi(t) - \psi(\ell))^\alpha), \quad \forall t \geq \ell,
   \]
   where \( \varphi_0 \triangleq \frac{1}{2}(I_{\psi}^{-\alpha}\|z - y_\ast\|^2)(\ell^+). \)

2. If we additionally assume that Assumption \( 2.1(1) \) hold, then
   \[
   f(z(t)) - f(y_\ast) \leq \frac{M_f}{2}\varphi_0(\psi(t) - \psi(\ell))^{\alpha - 1} \mathcal{E}_{\alpha, \alpha}(-\beta m_f(\psi(t) - \psi(\ell))^\alpha), \quad \forall t \geq \ell.
   \]

**Proof.** (1) Note that
   \[
   f(x) \geq f(y) + \langle \nabla f(y), x - y \rangle + \frac{m_f}{2}\|x - y\|^2, \quad \forall x, y \in \mathbb{R}^d.
   \]
In particular with \( y = z(t) \) and \( x = y_\ast \), we get
   \[
   \langle z(t) - y_\ast, \nabla f(z(t)) \rangle \geq f(z(t)) - f(y_\ast) + \frac{m_f}{2}\|z(t) - y_\ast\|^2
   \]
   \[
   \geq \frac{m_f}{2}\|z(t) - y_\ast\|^2, \quad \forall t \geq \ell,
   \]
where the last inequality holds since \( y_\ast \in \text{arg min} f(x) \). By Proposition \( 3.4 \) we have
   \[
   RLD_{t, \psi}^{\alpha}\varphi(t) \leq \langle z(t) - y_\ast, RLD_{t, \psi}^{\alpha}\varphi(t) \rangle
   \]
   \[
   = -\beta \langle z(t) - y_\ast, \nabla f(z(t)) \rangle
   \]
   \[
   \leq -\frac{m_f}{2}\|z(t) - y_\ast\|^2.
   \]
For setting
   \[
   h(t) \triangleq -\frac{\beta m_f}{2}\|z(t) - y_\ast\|^2 - RLD_{t, \psi}^{\alpha}\varphi(t),
   \]
we have \( h(t) \geq 0 \) for all \( t \geq 0 \), and moreover
   \[
   RLD_{t, \psi}^{\alpha}\varphi(t) = -\beta m_f\varphi(t) - h(t).
   \]
By \( \mathbb{I} \) Theorem 5.1, we can write
   \[
   \varphi(t) = \varphi_0(\psi(t) - \psi(\ell))^{\alpha - 1} \mathcal{E}_{\alpha, \alpha}(-\beta m_f(\psi(t) - \psi(\ell))^\alpha)
   \]
   \[
   - \int_{\ell}^{t}(\psi(t) - \psi(\tau))^{\alpha - 1}\psi'(\tau)\mathcal{E}_{\alpha, \alpha}(-\beta m_f(\psi(t) - \psi(\tau))^\alpha)h(\tau)\,d\tau
   \]
   \[
   \leq \varphi_0(\psi(t) - \psi(\ell))^{\alpha - 1} \mathcal{E}_{\alpha, \alpha}(-\beta m_f(\psi(t) - \psi(\ell))^\alpha),
   \]
where in the last inequality, we use Proposition \( \mathbb{I} \).

(2) By inequalities \( \mathbb{I} \) and \( \mathbb{I} \), we get
   \[
   f(z(t)) - f(y_\ast) \leq \frac{M_f}{2}\|z(t) - y_\ast\|^2 \leq \frac{1}{2}M_f\varphi_0(\psi(t) - \psi(\ell))^{\alpha - 1} \mathcal{E}_{\alpha, \alpha}(-\beta m_f(\psi(t) - \psi(\ell))^\alpha).
   \]
\( \square \)
4.3. Convergence at an exponential rate. Theorem 4.2 reveals that the solution of equation (4.1) can converge to a stationary point at the Mittag-Leffler convergence rate. In particular with $\alpha = 1$, we recover the exponential rate $O(e^{-\beta y(t)})$ for the continuous gradient method (4.3). In this subsection, we study the exponential rate when $\alpha \in (0, 1)$.

**Theorem 4.3.** Let $\alpha \in (0, 1)$. Suppose that the function $f : \mathbb{R}^d \to \mathbb{R}$ satisfies Assumption 2.7(1) and $\psi : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ satisfies Assumptions 2.3(1-2). Consider the fractional-order differential equation (4.1), where the step size $\beta$ is constant. If the solution $z(\cdot)$ of equation (5.1) converges to $y_* \in S(f)$ at the exponential rate $O(e^{-\omega p(t)})$, then $y_* = 0$.

**Proof.** Let $z(\cdot)$ be a solution of equation (5.1) converging to a stationary point $y_*$ with the convergence rate $O(e^{-\omega p(t)})$. Then there exists $t_1 \geq t_2$ such that

$$
\|z(t) - y_*\| \leq e^{-\omega t}, \quad \forall t \geq t_1.
$$

Assume in contrary that $y_* \neq 0$ and so we can set

$$
K \overset{\Delta}{=} \frac{\beta}{\|y_*\|} + 1.
$$

Since $E_\alpha(x) = 2E_{2\alpha}(x^2) - E_\alpha(-x)$, by [3 Corollary 3.8], we can find $t_2 \geq t_1$ with the property that

$$
E_\alpha(-L_f(\psi(t) - \psi(t)) \alpha) > Ke^{-\omega t}, \quad \forall t \geq t_2.
$$

Denote

$$
Q \overset{\Delta}{=} \sup\{\|z(t) - y_*\| : t \in [t, t_2]\}, \quad u(t) \overset{\Delta}{=} z_0(\psi(t) - \psi(t)) \alpha - \Gamma(t) y_*,
$$

By Proposition 2.8, $z(\cdot)$ is of the following form

$$
z(t) = \frac{z_0}{\Gamma(\alpha)}(\psi(t) - \psi(t)) \alpha - \frac{\beta}{\Gamma(\alpha)} \int_0^t (\psi(t) - \psi(s)) \alpha \psi'(s) \nabla f(z(s)) ds,
$$

and so

$$
\Gamma(\alpha) \|u(t)\| \leq \Gamma(\alpha) \|z(t) - y_*\| + \beta \int_0^t (\psi(t) - \psi(s)) \alpha \psi'(s) \|\nabla f(z(s))\| ds
$$

$$
\leq \Gamma(\alpha) \|z(t) - y_*\| + \beta L_f \int_0^t (\psi(t) - \psi(s)) \alpha \psi'(s) \|z(s) - y_*\| ds
$$

$$
= \Gamma(\alpha) \|z(t) - y_*\| + \beta L_f \left\{ \int_{t_2}^t + \int_{t_2}^t \right\} (\psi(t) - \psi(s)) \alpha \psi'(s) \|z(s) - y_*\| ds.
$$

We estimate

$$
\Gamma(\alpha) \|u(t)\| \leq \Gamma(\alpha) \|z(t) - y_*\| + \beta L_f Q \int_0^t (\psi(t) - \psi(s)) \alpha \psi'(s) \|z(s) - y_*\| ds
$$

$$
+ \frac{\beta L_f}{K} \int_{t_2}^t (\psi(t) - \psi(s)) \alpha \psi'(s) E_\alpha(-L_f(\psi(s) - \psi(t)) \alpha) ds
$$

$$
\leq \Gamma(\alpha) \|z(t) - y_*\| + \frac{\beta L_f Q}{\alpha} [(\psi(t) - \psi(t))^\alpha - (\psi(t) - \psi(t_2))^\alpha]
$$

$$
+ \frac{\beta L_f}{K} \int_0^t (\psi(t) - \psi(s)) \alpha \psi'(s) E_\alpha(-L_f(\psi(s) - \psi(t)) \alpha) ds.
$$
Note that by the Mean Value theorem (applied for the function \((\psi(t) - \psi(\ell))^{\alpha}\)), there is \(\delta \in (t, t_2)\) with
\[
(\psi(t) - \psi(\ell))^{\alpha} - (\psi(t) - \psi(t_2))^{\alpha} = -\alpha(\psi(t) - \psi(\ell))^{\alpha-1}\psi'(\delta),
\]
which implies, as \(\alpha \in (0, 1)\) and \(\psi\) satisfies Assumption 2.3(2), that
\[
\lim_{t \to \infty} [(\psi(t) - \psi(\ell))^{\alpha} - (\psi(t) - \psi(t_2))^{\alpha}] = 0.
\]
This limit and Lemma 3.7 show
\[
\lim_{t \to \infty} \|u(t)\| \leq \lim_{t \to \infty} \left\{ \Gamma(\alpha) \|z(t) - y_*\| + \frac{\beta L_f Q}{\alpha} [(\psi(t) - \psi(\ell))^{\alpha} - (\psi(t) - \psi(t_2))^{\alpha}] \right\}
\]
\[
+ \frac{\beta L_f}{K} \lim_{t \to \infty} \int_{\ell}^{t} (\psi(t) - \psi(s))^{\alpha-1}\psi'(s)\mathcal{E}_\alpha(-L_f(\psi(s) - \psi(\ell))^{\alpha}) \, ds
\]
\[
= \frac{\beta \Gamma(\alpha)}{K}.
\]
Thus,
\[
\frac{\beta \Gamma(\alpha)}{K} \geq \lim_{t \to \infty} \|u(t)\| = \Gamma(\alpha) \|y_*\|
\]
but this is impossible.

4.4. Perturbations. In this section, we study
\[\tag{4.5} RLD_{t,\psi}^\alpha z(t) = -\beta \nabla f(z(t)) + g(t), \quad (I_{t,\psi}^{1-\alpha} z)(\ell) = z_0 \in \mathbb{R}^d, \quad \forall t \geq \ell, \]
where the function \(g(\cdot)\) reflects an external action on the system. Thus, (4.5) can be viewed as a perturbation of (4.4).

**Assumption 4.1.** The function \(g : \mathbb{R}_{\geq \ell} \to \mathbb{R}^d\) satisfies
\[
Q \triangleq \sup_{t \geq \ell} \int_{\ell}^{t} (\psi(t) - \psi(s))^{\alpha-1}\psi'(s) \|g(s)\|^2 \, dt < \infty.
\]

**Theorem 4.4.** Let \(\alpha \in (0, 1]\). Suppose that the function \(f : \mathbb{R}^d \to \mathbb{R}\) satisfies Assumption 2.3(3) (in this case, \(S(f) = \{y_*\}\) and \(\psi : \mathbb{R}_{\geq \ell} \to \mathbb{R}_{\geq 0}\) satisfies Assumptions 2.3(1-2). Consider the fractional-order differential equation (4.4), where the step size \(\beta\) is a constant. If the step size \(\beta\) satisfies
\[\tag{4.6} \beta > \frac{1}{m_f},\]
then \(y_*\) is a limit point of \(z(\cdot)\), i.e. there exists a sequence \(\{s_m\} \in \mathbb{R}_{\geq \ell}\) such that
\[
\lim_{m \to \infty} z(s_m) = y_*.
\]

**Proof.** By Proposition 5.1, we have
\[
RLD_{t,\psi}^\alpha \varphi(t) \leq \langle z(t) - y_* , RLD_{t,\psi}^\alpha z(t) \rangle = -\beta \langle z(t) - y_* , \nabla f(z(t)) \rangle + \langle z(t) - y_* , g(t) \rangle
\]
\[
\leq -\frac{\beta m_f}{2} \|z(t) - y_*\|^2 + \langle z(t) - y_* , g(t) \rangle
\]
\[
\leq -\frac{1}{2} (\beta m_f - 1) \|z(t) - y_*\|^2 + \frac{1}{2} \|g(t)\|^2.
\]
Set
\[
h(t) \triangleq -\frac{1}{2} (\beta m_f - 1) \|z(t) - y_*\|^2 + \frac{1}{2} \|g(t)\|^2 - RLD_{t,\psi}^\alpha \varphi(t).
\]
Thus, \(h(t) \geq 0\) for \(t \geq \ell\), and
\[
RLD_{t,\psi}^\alpha \varphi(t) = -(\beta m_f - 1) \varphi(t) + \frac{1}{2} \|g(t)\|^2 - h(t).
\]
Hence,
\[
\varphi(t) - \frac{1}{\Gamma(\alpha)}[\psi(t) - \psi(\ell)]^{\alpha-1} I^{1-\alpha}_{\ell,\psi} \varphi(\ell^+) = I^\alpha_{\ell,\psi} \circ R^\alpha_{\tau} D^\alpha_{\ell,\psi} \varphi(t)
\]
\[
= -(\beta m_f - 1) I^\alpha_{\ell,\psi} \varphi(t) + \frac{1}{2} I^\alpha_{\ell,\psi} \|g\|^2(t) - I^\alpha_{\ell,\psi} h(t)
\]
\[
\leq -(\beta m_f - 1) I^\alpha_{\ell,\psi} \varphi(t) + \frac{1}{2} I^\alpha_{\ell,\psi} \|g\|^2(t),
\]
which gives
\[
(\beta m_f - 1) I^\alpha_{\ell,\psi} \varphi(t) \leq \frac{1}{2} I^\alpha_{\ell,\psi} \|g\|^2(t) + \frac{1}{\Gamma(\alpha)}[\psi(t) - \psi(\ell)]^{\alpha-1} I^{1-\alpha}_{\ell,\psi} \varphi(\ell^+) - \varphi(t)
\]
\[
\leq \frac{1}{2} I^\alpha_{\ell,\psi} \|g\|^2(t) + \frac{1}{\Gamma(\alpha)}[\psi(t) - \psi(\ell)]^{\alpha-1} I^{1-\alpha}_{\ell,\psi} \varphi(\ell^+).
\]
By Assumption 4.1 we can write
\[
\int_\ell^t (\psi(t) - \psi(s))^{\alpha-1} \psi'(s) \|z(s) - y_s\|^2 ds
\]
\[
\leq \frac{Q + \frac{2}{\Gamma(\alpha)}[\psi(t) - \psi(\ell)]^{\alpha-1} I^{1-\alpha}_{\ell,\psi} \varphi(\ell^+)}{\beta m_f - 1},
\]
which implies, by Assumption 2.3(2), that
\[
\inf\{\|z(s) - y_s\| : s \geq \ell\} = 0.
\]

5. Gradient methods with Caputo type

In this section, we study
\[
CD^\alpha_{\ell,\psi} z(t) = -\beta \nabla f(z(t)), \quad z(\ell) = z_0 \in \mathbb{R}^d, \quad \forall t \geq \ell.
\]
For \( y_*, \in S(f) \), let us define the function \( \lambda : \mathbb{R}_{\geq \ell} \rightarrow \mathbb{R}_{\geq 0} \) by setting
\[
\lambda(t) \triangleq \frac{1}{2} \|z(t) - y_*\|^2, \quad t \geq \ell.
\]

5.1. The case of convex \( f(\cdot) \). This subsection is devoted to studying (5.1) under the assumption on that \( f(\cdot) \) is convex but perhaps not strongly convex. Theorem 5.1 indicates that there exists a limit point of \( z(\cdot) \) which belongs to \( S(f) \).

\textbf{Theorem 5.1.} Let \( \alpha \in (0, 1] \). Suppose that the function \( f : \mathbb{R}^d \rightarrow \mathbb{R} \) satisfies Assumption 2.7(2) and \( \psi : \mathbb{R}_{\geq \ell} \rightarrow \mathbb{R}_{\geq 0} \) satisfies Assumptions 2.3(1-2). Consider the fractional-order differential equation (4.1), where the step size \( \beta \) is a constant. Then \( z(\cdot) \) is bounded, and there is a sequence \( \{z(s_m)\} \) converging to a point in \( S(f) \). Furthermore, \( f(\tilde{z}(t)) \) converges to \( f_* \) as \( O(\psi(t)^{-\alpha}) \), where
\[
\tilde{z}(t) \triangleq \frac{\Gamma(\alpha + 1)}{(\psi(t) - \psi(\ell))^\alpha} \cdot (I^\alpha_{\ell,\psi} z)(t).
\]
\textbf{Proof.} By Proposition 5.1 we have
\[
CD^\alpha_{\ell,\psi} \lambda(t) \leq -\beta (z(t) - y_*, \nabla f(z(t))) \leq \beta [f_* - f(z(t))] \leq 0, \quad \forall t \in \mathbb{R}_{\geq \ell}
\]
and so
\[
0 \geq \beta I^\alpha_{\ell,\psi} (f_* - f \circ z)(t) \geq I^\alpha_{\ell,\psi} \circ CD^\alpha_{\ell,\psi} \lambda(t) = \lambda(t) - \lambda(\ell),
\]
where in the last equality we use [8, Theorem 2.6 and equality 16]. As a result, \( z(\cdot) \) is bounded and furthermore

\[
\sup \{ I_{\ell,\psi}^\alpha (f \circ z - f_\ast)(t) : t \geq \ell \} < \infty.
\]

Taking into account the form of \( I_{\ell,\psi}^\alpha \) in Definition 2.4 and Assumption 2.3(2), we must have

\[
f_\ast = \inf \{ f(z(s)) : s \geq \ell \},
\]

which implies, by the definition of an infimum, that there exists a sequence \( \{s_n\} \subseteq \mathbb{R}_{\geq \ell} \) such that \( \lim_{n \to \infty} f(z(s_n)) = f_\ast \). Since \( z(\cdot) \) is bounded, we can extract a convergent subsequence \( \{z(s_{n_k})\} \) of \( \{z(s_n)\} \). Suppose that \( \omega \triangleq \lim_{k \to \infty} z(s_{n_k}) \) and then

\[
f_\ast = \lim_{n \to \infty} f(z(s_n)) = \lim_{k \to \infty} f(z(s_{n_k})) = f(\omega);
\]

namely, \( \omega \) is a stationary point.

Since \( f(\cdot) \) is convex, it follows from Proposition 3.6 that

\[
\Gamma(\alpha + 1) \cdot (I_{\ell,\psi}^\alpha (f \circ z)(t) \geq f(\widehat{z}(t)).
\]

Thus,

\[
0 \leq f(\widehat{z}(t)) - f_\ast \leq \frac{\Gamma(\alpha + 1)}{(\psi(t) - \psi(\ell))^\alpha} \cdot I_{\ell,\psi}^\alpha (f \circ z - f_\ast)(t) \leq \frac{\Gamma(\alpha + 1)\lambda(\ell)}{\beta(\psi(t) - \psi(\ell))^\alpha},
\]

which implies at least an \( O(\psi(t)^{-\alpha}) \) convergence rate. \( \Box \)

5.2. The case of strongly convex \( f(\cdot) \). Under the assumption on the strong convexity, the following result establishes Mittag-Leffler convergence to the optimal point, which is a general type of exponential convergence.

**Theorem 5.2.** Let \( \alpha \in (0, 1) \). Suppose that the function \( f : \mathbb{R}^d \to \mathbb{R} \) satisfies Assumption 2.4 and \( \psi : \mathbb{R}_{\geq \ell} \to \mathbb{R}_{\geq 0} \) satisfies Assumption 2.3(1). Consider the fractional-order differential equation (5.1), where the step size \( \beta \) is a constant. Then we have the following conclusions.

1. The solution \( z(\cdot) \) converges to \( y_\ast \) with rate:

\[
\|z(t) - y_\ast\|^2 \leq \|z_0 - y_\ast\|^2 e_{\alpha}(\psi(t) - \psi(\ell))^\alpha, \quad \forall t \geq \ell.
\]

2. If we additionally assume that Assumption 2.1(1) holds, then

\[
f(z(t)) - f(y_\ast) \leq \frac{1}{2} M_f \|z_0 - y_\ast\|^2 e_{\alpha}(\psi(t) - \psi(\ell))^\alpha, \quad \forall t \geq \ell.
\]

**Proof.** (1) By Proposition 3.3 we have

\[
CD_{\ell,\psi}^\alpha \lambda(t) \leq \langle z(t) - y_\ast, CD_{\ell,\psi}^\alpha z(t) \rangle = -\beta \langle z(t) - y_\ast, \nabla f(z(t)) \rangle \leq -\frac{\beta m_f}{2} \|z(t) - y_\ast\|^2, \quad \text{(by Assumption 2.1(3))}.
\]

For setting

\[
h(t) \triangleq -\frac{\beta m_f}{2} \|z(t) - y_\ast\|^2 - CD_{\ell,\psi}^\alpha \lambda(t),
\]

we have \( h(t) \geq 0 \) for all \( t \geq \ell \), and moreover

\[
CD_{\ell,\psi}^\alpha \lambda(t) = -\beta m_f \lambda(t) - h(t).
\]
By [8, Theorem 5.2], we can write

\[
\lambda(t) = \lambda(\ell) \mathcal{E}_\alpha(-\beta m_f(\psi(t) - \psi(\ell))^\alpha)
- \int_\ell^t (\psi(t) - \psi(\tau))^{\alpha-1} \psi'(\tau) \mathcal{E}_{\alpha,\alpha}(-\beta m_f(\psi(t) - \psi(\tau))^\alpha) h(\tau) \, d\tau
\leq \lambda(\ell) \mathcal{E}_\alpha(-\beta m_f(\psi(t) - \psi(\ell))^\alpha),
\]

where in the last inequality, we use Proposition 2.9.

(2) The proof is similar to the second item of Theorem 5.2. □

5.3. Convergence at an exponential rate. Theorem 5.2 reveals that the solution of equation (5.1) can converge to a stationary point at the Mittag-Leffler convergence rate. In particular with \( \alpha = 1 \), this convergence speed reduces to the exponential rate \( O(e^{-\beta m_f \psi(t)}) \).

The following result indicates that if \( \alpha \in (0,1) \), then there is no nontrivial solution of equation (5.1) converging to a stationary point with such exponential rate.

\textbf{Theorem 5.3.} Let \( \alpha \in (0,1) \). Suppose that the function \( f : \mathbb{R}^d \to \mathbb{R} \) satisfies Assumption 2.1(1) and \( \psi : \mathbb{R} \geq \ell \to \mathbb{R} \geq 0 \) satisfies Assumptions 2.3(1-2). Consider the fractional-order differential equation (5.1), where the step size \( \beta \) is constant. Then every solution of equation (5.1) does not converge to any stationary point at the exponential rate \( O(e^{-\omega \psi(t)}) \).

\textbf{Proof.} Assume in contrary that \( z(\cdot) \) is a solution of equation (5.1) which converges to a stationary point \( y^* \) with the convergence rate \( O(e^{-\omega \psi(t)}) \). Then there exists \( t_1 \geq \ell \) such that

\[
\|z(t) - y^*\| \leq e^{-\omega \psi(t)}, \quad \forall t \geq t_1.
\]

Denote

\[
K \triangleq \beta \|z(\ell) - y^*\| + 1.
\]

Since \( \mathcal{E}_\alpha(x) = 2\mathcal{E}_{2\alpha}(x^2) - \mathcal{E}_\alpha(-x) \), by [8, Corollary 3.8], we can find \( t_2 \geq t_1 \) with the property that

\[
\mathcal{E}_\alpha(-L_f(\psi(t) - \psi(\ell))^\alpha) > K e^{-\omega \psi(t)}, \quad \forall t \geq t_2.
\]

Denote

\[
Q \triangleq \sup\{\|z(t) - y^*\| : t \in [\ell, t_2]\}.
\]

By Proposition 2.8, \( z(\cdot) \) is of the following form

\[
z(t) = z_0 - \frac{\beta}{\Gamma(\alpha)} \int_\ell^t (\psi(t) - \psi(s))^{\alpha-1} \psi'(s) \nabla f(z(s)) \, ds,
\]

and so

\[
\Gamma(\alpha) \|z_0 - y^*\| \leq \Gamma(\alpha) \|z(t) - y^*\| + \beta L_f \left\{ \int_\ell^{t_2} + \int_{t_2}^t \right\} (\psi(t) - \psi(s))^{\alpha-1} \psi'(s) \|z(s) - y^*\| \, ds.
\]
We estimate
\[
\Gamma(\alpha) \|z_0 - y_*\| \leq \Gamma(\alpha) \|z(t) - y_*\| + \beta L_f Q \int_\ell^{t_2} (\psi(t) - \psi(s))^{\alpha-1} \psi'(s) \, ds \\
+ \frac{\beta L_f}{K} \int_\ell^{t_2} (\psi(t) - \psi(s))^{\alpha-1} \psi'(s) \mathcal{E}_\alpha(-L_f(\psi(s) - \psi(\ell))) \, ds \\
\leq \Gamma(\alpha) \|z(t) - y_*\| + \frac{\beta L_f Q}{\alpha} [(\psi(t) - \psi(\ell))^{\alpha} - (\psi(t) - \psi(t_2))^{\alpha}] \\
+ \frac{\beta L_f}{K} \int_\ell^{t} (\psi(t) - \psi(s))^{\alpha-1} \psi'(s) \mathcal{E}_\alpha(-L_f(\psi(s) - \psi(\ell))) \, ds.
\]
Since
\[
\lim_{t \to \infty} [(\psi(t) - \psi(\ell))^{\alpha} - (\psi(t) - \psi(t_2))^{\alpha}] = 0,
\]
by Lemma 3.7 we have
\[
\Gamma(\alpha) \|z_0 - y_*\| \leq \lim_{t \to \infty} \left\{ \Gamma(\alpha) \|z(t) - y_*\| + \frac{\beta L_f Q}{\alpha} [(\psi(t) - \psi(\ell))^{\alpha} - (\psi(t) - \psi(t_2))^{\alpha}] \right\} \\
+ \frac{\beta L_f}{K} \lim_{t \to \infty} \int_\ell^{t} (\psi(t) - \psi(s))^{\alpha-1} \psi'(s) \mathcal{E}_\alpha(-L_f(\psi(s) - \psi(\ell))) \, ds \\
= \frac{\beta \Gamma(\alpha)}{K}.
\]
But this is impossible.

5.4. Perturbations. In this section, we study
\[
(5.4) \quad C^D_{\ell,\psi} z(t) = -\beta \nabla f(z(t)) + g(t), \quad z(\ell) = z_0 \in \mathbb{R}^d, \quad \forall t \geq \ell,
\]
where the function \(g(\cdot)\) reflects an external action on the system. Equation (4.3) can be viewed as a perturbation of (4.1).

**Theorem 5.4.** Let \(\alpha \in (0, 1]\). Suppose that the function \(f : \mathbb{R}^d \to \mathbb{R}\) satisfies Assumption [2.7/3] (in this case, \(S(f) = \{f\} \)) and \(\psi : \mathbb{R}_{\geq \ell} \to \mathbb{R}_{\geq 0}\) satisfies Assumption [2.8/1-2]. Consider the fractional-order differential equation (5.4), where the step size \(\beta\) is a constant. If the step size \(\beta\) satisfies (1.16), then \(y_*\) is a limit point of \(z(\cdot)\), i.e. there exists a sequence \(\{s_m\} \subseteq \mathbb{R}_{\geq \ell}\) such that
\[
\lim_{m \to \infty} z(s_m) = y_*,
\]
Proof. By Proposition 5.1 we have
\[
C^D_{\ell,\psi} \lambda(t) \leq \langle z(t) - y_*, C^D_{\ell,\psi} z(t) \rangle = -\beta \langle z(t) - y_*, \nabla f(z(t)) \rangle + \langle z(t) - y_*, g(t) \rangle \\
\leq -\frac{\beta m_f}{2} \|z(t) - y_*\|^2 + \langle z(t) - y_*, g(t) \rangle \\
\leq -\frac{1}{2}(\beta m_f - 1) \|z(t) - y_*\|^2 + \frac{1}{2} \|g(t)\|^2.
\]
For setting
\[
h(t) \triangleq \frac{1}{2}(\beta m_f - 1) \|z(t) - y_*\|^2 + \frac{1}{2} \|g(t)\|^2 - C^D_{\ell,\psi} \lambda(t),
\]
we have \(h(t) \geq 0\) for \(t \geq \ell\) and
\[
C^D_{\ell,\psi} \lambda(t) = - (\beta m_f - 1) \lambda(t) + \frac{1}{2} \|g(t)\|^2 - h(t).
\]
Hence,
\[
\lambda(t) = \lambda(\ell) - \frac{1}{2} I_{t,\psi}^\alpha \|g\|^2(t) - I_{t,\psi}^\alpha h(t)
\]
which gives
\[
(\beta m_f - 1) I_{t,\psi}^\alpha \lambda(t) \leq \frac{1}{2} I_{t,\psi}^\alpha \|g\|^2(t) + \lambda(\ell) - \lambda(t)
\]
By Assumption 4.1, we can write
\[
\int_\ell^t (\psi(t) - \psi(s))^{\alpha-1} \psi'(s) \|z(s) - y_*\|^2 ds \leq \frac{Q + 2\lambda(\ell)}{\beta m_f - 1}, \quad \forall t \geq \ell,
\]
which implies, by Assumption 2.3 that
\[
\inf \{ \|z(s) - y_*\| : s \geq \ell \} = 0.
\]

6. Design numerical method

In the previous sections, we showed \(\psi\)-fractional derivative based methods are also a tool solving Problem (1.1). A problem arising is how to simulate these methods in practice. This section studies that problem in the more general context of fractional differential equations by extending the Adams-Bashforth-Moulton (ABM) method to \(\psi\)-fractional derivatives. After proving the convergence of the ABM method, we offer numerical examples and indicate that the fractional order \(\alpha\) and weight \(\psi\) can be adjusted in order to improve the performance.

6.1. ABM method for \(\psi\)-fractional derivatives. Given a function \(g : [0, T] \times \mathbb{R}^d \to \mathbb{R}^d\) that is Lipschitz continuous with Lipschitz constant \(L\), a solution of equation (2.5) is given as
\[
x(t) = x(a) + \frac{1}{\Gamma(\alpha)} \int_a^t (\psi(t) - \psi(s))^{\alpha-1} \psi'(s) g(s, x(s)) ds.
\]

The Adams-Bashforth-Moulton method for this initial value problem is a predictor-corrector method motivated by the work of Diethelm in [6]. Just as in [6], the numerical method uses a piecewise constant predictor and a piecewise linear corrector. The advantage of utilizing this approach to the design of a numerical method for the fractional differential equations at hand is that the only substantial adjustments to the ABM method of [6] is an alteration of the coefficients employed in the numerical method. Moreover, in contrast to work such as [13], is that the evaluation of the ABM method at each time-step is \(O(k)\) in computation time, where \(k\) is the \(k\)-th step. Whereas, in [13], each timestep requires matrix inversion for interpolation, which typically requires \(O(k^3)\) computation time. While the ABM method has slower convergence, the advantage gained through computation time makes it more practical for problems with longer time horizons.

Let \(0 = t_0 < t_1 < \ldots < t_{k+1}\). As in [6], let
\[
\phi_{j,k+1}(z) \triangleq \begin{cases} \frac{z-t_j}{t_{j+1}-t_j} & \text{if } t_j < z \leq t_{j+1} \\ \frac{z-t_{j-1}}{t_{j+1}-t_j} & \text{if } t_{j-1} < z \leq t_j \\ 0 & \text{otherwise.} \end{cases}
\]
Given a twice continuously differentiable function $f : [0, t_{k+1}] \to \mathbb{R}$, a piecewise linear expression of $f$ is given as
\[
\hat{f} \triangleq \sum_{i=0}^{k+1} f(t_i) \phi_i, k+1, \text{ and}
\]
\[
\sup_t |f(t) - \hat{f}(t)| \leq \max_{i=1, \ldots, k+1} |t_i - t_{i-1}|^2 \cdot \sup_t |f''(t)|.
\]
For regularly spaced $t_i$, with step size $h > 0$, it follows that
\[
(6.2) \quad \sup_t |f(t) - \hat{f}(t)| \leq h^2 \sup_t |f''(t)|.
\]
If a piecewise constant approximation of $f$ is given as $\hat{f} \triangleq \sum_{i=0}^{k} f(t_i) \chi_{[t_i, t_{i+1}]}$, where $\chi_A$ is the indicator function for the set $A$, then it can be also be shown that
\[
(6.3) \quad \sup_t |f(t) - \hat{f}(t)| \leq h \sup_t |f'(t)|.
\]
Given a collection of points $\{(t_i, x_i)\}_{i=0}^{k}$, the predictor for $x_{k+1}$ is then given as
\[
(6.4) \quad x_{k+1}^P \triangleq x_0 + \frac{1}{\Gamma(\alpha)} \sum_{i=0}^{k} b_i, k+1 g(t_i, x_i),
\]
where
\[
b_i, k+1 \triangleq \int_{t_i}^{t_{i+1}} (\psi(t_{k+1}) - \psi(\tau))^{\alpha-1} \psi'(\tau) d\tau.
\]
In this setting $t \mapsto g(t, x(t))$ is being approximated by $\hat{g} = \sum_{i=0}^{k} g(t_i, x_i) \chi_{[t_i, t_{i+1}]}$ which utilizes the approximated points $x_i \approx x(t_i)$. Consequently,
\[
\|x(t_{k+1}) - x_{k+1}^P\| = \frac{1}{\Gamma(\alpha)} \left\| \int_0^{t_{k+1}} (\psi(t_{k+1}) - \psi(\tau))^{\alpha-1} \psi'(\tau) [g(\tau, x(\tau))
\right.
\]
\[
- \sum_{i=0}^{k} g(t_i, x_i) \chi_{[t_i, t_{i+1}]}(\tau)] d\tau \right\|
\]
\[
= \frac{1}{\Gamma(\alpha)} \sum_{i=0}^{k} \left\| \int_{t_i}^{t_{i+1}} (\psi(t_{k+1}) - \psi(\tau))^{\alpha-1} \psi'(\tau) [g(\tau, x(\tau))
\right.
\]
\[
- \sum_{i=0}^{k} g(t_i, x_i) \chi_{[t_i, t_{i+1}]}(\tau)] d\tau \right\|.
\]
Examining each summand and noting that
\[
g(\tau, x(\tau)) - g(t_i, x_i) \chi_{[t_i, t_{i+1}]}(\tau) = g(\tau, x(\tau)) - g(t_i, x(t_i)) + g(t_i, x(t_i)) - g(t_i, x_i) \chi_{[t_i, t_{i+1}]}(\tau),
\]
we have
\[
\left\| \int_{t_i}^{t_{i+1}} (\psi(t_{k+1}) - \psi(\tau))^{\alpha-1} \psi'(\tau) [g(\tau, x(\tau)) - g(t_i, x(t_i)) \chi_{[t_i, t_{i+1}]}(\tau)] d\tau \right\|
\]
\[
\leq \left\| \int_{t_i}^{t_{i+1}} (\psi(t_{k+1}) - \psi(\tau))^{\alpha-1} \psi'(\tau) [g(\tau, x(\tau)) - g(t_i, x(t_i)) d\tau \right\| + b_i, k+1 L \|x(t_i) - x_i\|.
\]
Assuming \( t \mapsto g(t, x(t)) \) is continuously differentiable and employing the mean value theorem, the following is obtained:

\[
\left\| \sum_{i=0}^{k} \int_{t_i}^{t_{i+1}} (\psi(t_{k+1}) - \psi(\tau))^{\alpha-1} \psi'(\tau) (g(\tau, x(\tau)) - g(t_i, x(t_i))) d\tau \right\| \\
\leq \sup_{t \in [t_0, t_{k+1}]} \left\| \frac{d}{dt} g(t, x(t)) \right\| \sum_{i=0}^{k} \left( \int_{t_i}^{t_{i+1}} (\psi(t_{k+1}) - \psi(\tau))^{\alpha-1} \psi'(\tau) (\tau - ih) d\tau \right) \\
\leq \sup_{t \in [t_0, t_{k+1}]} \left\| \frac{d}{dt} g(t, x(t)) \right\| h \left( \frac{(\psi(t_{k+1}) - \psi(t_0))^{\alpha}}{\alpha} \right)
\]

Combining the above results yields

\[
\| x(t_{k+1}) - x^P_{k+1} \| \leq \frac{1}{\Gamma(\alpha)} \left( \sup_{t \in [t_0, t_{k+1}]} \left\| \frac{d}{dt} g(t, x(t)) \right\| \frac{h (\psi(t_{k+1}) - \psi(t_0))^{\alpha}}{\alpha} \right) \\
+ L \sup_{i=1, \ldots, k} \| x(t_i) - x_i \| \sum_{i=0}^{k} b_{i,k+1} \\
= \frac{1}{\Gamma(\alpha)} \left( \sup_{t \in [t_0, t_{k+1}]} \left\| \frac{d}{dt} g(t, x(t)) \right\| \frac{h (\psi(t_{k+1}) - \psi(t_0))^{\alpha}}{\alpha} \right) \\
+ L \sup_{i=1, \ldots, k} \| x(t_i) - x_i \| \frac{(\psi(t_{k+1}) - \psi(t_0))^{\alpha}}{\alpha}.
\]

This inequality completes the analysis for the predictor step. The corrector is given as

\[
(6.5) \quad x_{k+1} \triangleq x_0 + \frac{1}{\Gamma(\alpha)} \left( \sum_{i=0}^{k} a_{i,k+1} g(t_i, x_i) + a_{k+1,k+1} g(t_{k+1}, x^P_{k+1}) \right),
\]

where

\[
a_{i,k+1} \triangleq \int_{0}^{t_{k+1}} (\psi(t_{k+1}) - \psi(\tau))^{\alpha-1} \psi'(\alpha) \phi_i, k+1(\tau) d\tau.
\]

Thus, \( a_{i,k+1} \) is

\[
a_{i,k+1} = \begin{cases} 
\frac{(\psi(t_{k+1}) - \psi(t_0))^{\alpha}}{(\psi(t_{k+1}) - \psi(t_i))^{\alpha}} - \frac{1}{t_i-t_0} \int_{t_0}^{t_i} (\psi(t_{k+1}) - \psi(\tau))^{\alpha} d\tau; & i = 0 \\
\frac{1}{t_i-t_{i-1}} \int_{t_{i-1}}^{t_i} (\psi(t_{k+1}) - \psi(\tau))^{\alpha} d\tau - \frac{1}{t_{i+1}-t_i} \int_{t_i}^{t_{i+1}} (\psi(t_{k+1}) - \psi(\tau))^{\alpha} d\tau; & 1 \leq i \leq k \\
\frac{1}{t_{k+1}-t_k} \int_{t_k}^{t_{k+1}} (\psi(t_{k+1}) - \psi(\tau))^{\alpha} d\tau; & i = k + 1
\end{cases}
\]

**Remark 6.1.** For practical implementation a simple left or right quadrature rule may be employed to give approximate values for \( a_{i,k+1} \):

\[
\tilde{a}_{i,k+1} \triangleq \begin{cases} 
\frac{(\psi(t_{k+1}) - \psi(t_0))^{\alpha}}{(\psi(t_{k+1}) - \psi(t_i))^{\alpha}} - \frac{(\psi(t_{k+1}) - \psi(t_0))^{\alpha}}{(\psi(t_{k+1}) - \psi(t_{i+1}))^{\alpha}}; & i = 0 \\
\frac{(\psi(t_{k+1}) - \psi(t_i))^{\alpha}}{(\psi(t_{k+1}) - \psi(t_i))^{\alpha}} - \frac{(\psi(t_{k+1}) - \psi(t_{i+1}))^{\alpha}}{(\psi(t_{k+1}) - \psi(t_{i+1}))^{\alpha}}; & 1 \leq i \leq k \\
\frac{(\psi(t_{k+1}) - \psi(t_i))^{\alpha}}{\alpha}; & i = k + 1
\end{cases}
\]
Leveraging the Lipschitz property of (6.7),

20 PHAM VIET HAI AND JOEL A. ROSENFELD

Now consider,

\[ \int_{t_0}^{t_{k+1}} (\psi(t_{k+1}) - \psi(t))^{(n-1)} \psi'(t) g(\tau, x(\tau)) d\tau - \sum_{i=0}^{k+1} a_{i,k+1} g(t_i, x(t_i)) \]

\[ = \frac{1}{\Gamma(\alpha)} \left| \int_{t_0}^{t_{k+1}} (\psi(t_{k+1}) - \psi(t))^{(n-1)} \psi'(t) g(\tau, x(\tau)) d\tau - \sum_{i=0}^{k+1} a_{i,k+1} g(t_i, x(t_i)) \right| \]

Leveraging the Lipschitz property of \( g \), it follows that

\[ \sum_{i=0}^{k+1} a_{i,k+1} g(t_i, x(t_i)) - \sum_{i=0}^{k} a_{i,k+1} g(t_i, x_i) - a_{k+1,k+1} g(t_{k+1}, x_{k+1}) \]

\[ \leq L \sup_{i=1,...,k} \| x(t_i) - x_i \| \sum_{i=0}^{k} a_{i,k+1} + La_{k+1,k+1}\| x(t_{k+1}) - x_{k+1} \| \]

\[ = L \sup_{i=1,...,k} \| x(t_i) - x_i \| \left( \frac{(\psi(t_{k+1}) - \psi(t_0))^\alpha}{\alpha} - \frac{(\psi(t_{k+1}) - \psi(t_k))^\alpha}{\alpha} \right) \]

\[ + L \left( \frac{(\psi(t_{k+1}) - \psi(t_k))^\alpha}{\alpha} \right) \| x(t_{k+1}) - x_{k+1} \| \]

Now consider,

\[ \left| \int_{t_0}^{t_{k+1}} (\psi(t_{k+1}) - \psi(t))^{(n-1)} \psi'(t) g(\tau, x(\tau)) d\tau - \sum_{i=0}^{k+1} a_{i,k+1} g(t_i, x(t_i)) \right| \]

\[ = \left| \int_{t_0}^{t_{k+1}} (\psi(t_{k+1}) - \psi(t))^{(n-1)} \psi'(t) \left( g(\tau, x(\tau)) - \sum_{i=0}^{k+1} g(t_i, x(t_i)) \phi_{i,k+1}(\tau) \right) d\tau \right| \]

\[ \leq h^2 \sup_t \left\| \frac{d^2}{dt^2} g(t, x(t)) \right\| \left| \int_{t_0}^{t_{k+1}} (\psi(t_{k+1}) - \psi(t))^{(n-1)} \psi'(t) d\tau \right| \] (use (6.2))

\[ = h^2 \sup_t \left\| \frac{d^2}{dt^2} g(t, x(t)) \right\| \frac{(\psi(t_{k+1}) - \psi(t_0))^\alpha}{\alpha}. \]

The following inequalities have been established:

\[ \| x(t_{k+1}) - x_{k+1} \| \leq h^2 \sup_t \left\| \frac{d^2}{dt^2} g(t, x(t)) \right\| \frac{(\psi(t_{k+1}) - \psi(t_0))^\alpha}{\alpha} + \]

\[ + L \left( \frac{(\psi(t_{k+1}) - \psi(t_k))^\alpha}{\alpha} \right) \| x(t_{k+1}) - x_{k+1} \| \]

(6.6)

\[ \| x(t_{k+1}) - x_{k+1} \| \leq \frac{1}{\Gamma(\alpha)} \left( \sup_{t\in[t_0,t_{k+1}]} \left\| \frac{d}{dt} g(t, x(t)) \right\| h \frac{(\psi(t_{k+1}) - \psi(t_0))^\alpha}{\alpha} \right) \]

(6.7)

and

\[ + L \sup_{i=1,...,k} \| x(t_i) - x_i \| \frac{(\psi(t_{k+1}) - \psi(t_0))^\alpha}{\alpha} \].
Theorem 6.2. Let $\alpha \in (0,1]$. Suppose that the function $\psi : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ satisfies Assumption 2.3(1). Let $x(t)$ be a solution to (2.3), and suppose that $t \mapsto g(t,x(t))$ is twice continuously differentiable over $[t_0,T]$, and that $g$ is Lipshitz continuous. Then predictor-corrector scheme above yields the convergence rate
\[ \|x(t_k) - x_k\| = O(h^{1+\alpha}) \]
for a suitably chosen $T$.

Proof. The proof will proceed by induction on $k$. For $k = 0$, $x_0 = x(t_0)$, so the result holds automatically. Now suppose that the result holds for $i = 1, \ldots, k$. Then, by (6.7),
\[
\|x(t_{k+1}) - x_{k+1}\| \leq \frac{1}{\Gamma(\alpha)} \left( \sup_{t \in [t_k,t_{k+1}]} \left\| \frac{d}{dt} g(t,x(t)) \right\| h^\alpha \frac{\psi(t_{k+1}) - \psi(t_0)}{\alpha} \right) \\
+ L \sup_{i=1,\ldots,k} \|x(t_i) - x_i\| \|x(t_{k+1}) - x(t_i)\| \frac{\psi(t_{k+1}) - \psi(t_0)}{\alpha}.
\]
To further simplify computations, note that
\[
\frac{\psi(t_{k+1}) - \psi(t_0)}{\alpha} \leq \psi'(T)^\alpha (t_{k+1} - t_0)^\alpha \leq \psi'(T)^\alpha (T - t_0)^\alpha,
\]
and
\[
\frac{\psi(t_{k+1}) - \psi(t_k)}{\alpha} \leq \psi'(T)^\alpha \frac{h^\alpha}{\alpha}.
\]
The applying the above to (6.10) yields
\[
\|x(t_{k+1}) - x_{k+1}\| \leq h^2 \sup_t \left\| \frac{d^2}{dt^2} g(t,x(t)) \right\| \psi'(T)^\alpha \frac{(T - t_0)^\alpha}{\alpha} \\
+ L\psi'(T)^\alpha \frac{h^{1+\alpha}}{\alpha \Gamma(\alpha)} \left( \psi'(T)^\alpha \frac{(T - t_0)^\alpha}{\alpha} + \psi'(T)^\alpha \frac{h^\alpha}{\alpha} \right) \\
+ L\psi'(T)^\alpha \frac{h^{1+2\alpha}}{\alpha \Gamma(\alpha)} \left( \frac{d}{dt} g(t,x(t)) \right)^\alpha \psi'(T)^\alpha \frac{(T - t_0)^\alpha}{\alpha} \\
+ L^2 \frac{h^{1+2\alpha}}{\alpha \Gamma(\alpha)} C \psi'(T)^\alpha \frac{(T - t_0)^\alpha}{\alpha} \\
= h^1 + \alpha \sup_t \left\| \frac{d^2}{dt^2} g(t,x(t)) \right\| \psi'(T)^\alpha \frac{(T - t_0)^\alpha}{\alpha} \\
+ L\psi'(T)^\alpha \frac{1}{\alpha \Gamma(\alpha)} \sup_{t \in [t_k,t_{k+1}]} \left\| \frac{d}{dt} g(t,x(t)) \right\| \psi'(T)^\alpha \frac{(T - t_0)^\alpha}{\alpha} \\
+ L^2 \frac{h^\alpha}{\alpha \Gamma(\alpha)} C \psi'(T)^\alpha \frac{(T - t_0)^\alpha}{\alpha}.
\]

With a selection of $T$ suitably close to $t_0$, the term in parentheses can be kept less than $C$. Hence, the convergence rate is established. \qed

Remark 6.3. Theorem 6.2 requires smoothness of both $g$ and $x$ for the convergence of the numerical method. Results for fractional order differential equations using the Caputo derivative show that $x$ inherits some smoothness on $[0,T]$ from $g$ (cf. [6]), and these results are expected to carry over to the present case. Moreover, through the selection of $\psi$ having higher order zeros at $t = 0$, numerical experiments indicate that $x$ may be smooth over $[0,T]$ as well.
Figure 1. This figure shows the results of the gradient descent methods of this manuscript as applied to the optimization of the Booth function. This figure presents the norm difference between the known optimal point and the two dimensional state vector. Here it can be seen that the selection of $\psi = t^4$ leads to a method that matches integer order gradient descent.

6.2. Numerical examples. Presented in Figure 1, Figure 2, and Figure 3 are the numerical experiments utilizing the functions given in [10]. The gradient descent methods of this manuscript were utilized with $\psi = t, t*\ln(t+1), t^2, t^4$, and for comparison the integer order gradient descent using RK4 has also been included. It can be seen that setting $\psi = t^2$ and $\psi = t^4$ leads to better performance than integer order gradient descent.

The plots given show the norm convergence of the two dimensional state to the optimal point, so the ideal is to demonstrate convergence of this quantity to zero. The spacing $h$ is indicated in the title of each figure as is the fractional order of the derivative. For each fractional numerical method, the ABM method was used with 5 corrector steps, where the corrector was applied iteratively to improve convergence.

On the Booth function, $f(x) = (x_1 + 2x_2 - 7)^2 + (2x_1 + x_2 - 5)^2$, it can be seen in Figure 1 that with the initial condition $(10, 5)$, $\alpha = 0.8$ and $\psi = t^4$, matching convergence rates are achieved between the integer order gradient descent method and the fractional gradient descent methods of this manuscript. Of the methods employed, the Caputo approached the optimal point most slowly. For the negative radial exponential function, $f(x) = -\exp\left(\frac{1}{2}\|x\|\right)$, with initial condition $(1, 5)$, there is a much more dramatic difference between the methods, where $\psi = t^2$ converges very quickly as seen in Figure 2. The noise after the trajectory achieves the optimal point is due to the nondifferentiability of the function at that point. Finally, for the Zakharov function, $f(x) = \|x\|^2 + (\sum_{i=1}^n 0.5ix_i)^2 + (\sum_{i=1}^n 0.5ix_i)^4$ (with $n = 2$ in this case), with initial condition $(10, 5)$, integer order gradient descent performs poorly compared to $\psi = t^2$ and $\psi = t^4$, where the fastest convergence is achieved by $\psi = t^4$ in Figure 3.

7. Conclusions

In the paper, we design $\psi$-fractional derivatives based methods solving unconstrained optimization problems. The convergence analysis of these methods is carried out for both strongly convex and non-strongly convex cases. The key element of our analysis is the identification of a Lyapunov-type function, which allows to establish convergence of generated trajectories in the Riemann-Liouville as well as in the Caputo case. Chain rules and Jensen-type inequality play essential roles in the analysis of these Lyapunov functions. Numerical examples using the ABM method reveal that the fractional order $\alpha$ and weight $\psi$ are tunable parameters, which can be helpful for improving the convergence speed. Future research may include extensions to constrained optimization problems, discrete-time methods and improvement of the convergence speed for specific problems. Moreover, the numerical methods given in Section 6 generalize that of [6], and they have the same limitations where
This figure shows the gradient descent methods of this manuscript as applied to the optimization of the negative radial exponential function. This figure presents the norm difference between the known optimal point and the two dimensional state vector. Here it can be seen that the selection of $\psi = t^4$ gives a dramatic improvement over the other methods shown in this figure.

This figure shows the gradient descent methods of this manuscript as applied to the optimization of the Zakharov function. This figure presents the norm difference between the known optimal point and the two dimensional state vector. Here it can be seen that the selection of $\psi = t^2$ and $\psi = t^4$ leads to a method that outperforms integer order gradient descent.

the convergence rate is valid only for a finite time horizon. Future developments for the numerical methods will be to provide a method and proof of convergence that is valid for arbitrarily large time horizons.

References

[1] Norelys Aguila-Camacho, Manuel A Duarte-Mermoud, and Javier A Gallegos. Lyapunov functions for fractional order systems. Communications in Nonlinear Science and Numerical Simulation, 19(9):2951–2957, 2014.
[2] Ricardo Almeida. A Caputo fractional derivative of a function with respect to another function. Communications in Nonlinear Science and Numerical Simulation, 44:460–481, 2017.
[3] Ricardo Almeida, Agnieszka B Malinowska, and M Teresa T Monteiro. Fractional differential equations with a Caputo derivative with respect to a kernel function and their applications. Mathematical Methods in the Applied Sciences, 41(1):336–352, 2018.
[4] S. Boyd and L. Vandenberghe. Convex Optimization. Cambridge University Press, Cambridge, 2004.
[5] Weisheng Chen, Hao Dai, Yanfei Song, and Zhengqiang Zhang. Convex Lyapunov functions for stability analysis of fractional order systems. IET Control Theory & Applications, 11(7):1070–1074, 2017.
[6] Kai Diethelm. The Analysis of Fractional Differential Equations: An Application-Oriented Exposition Using Differential Operators of Caputo Type. Springer-Verlag, Berlin, 2010.
[7] R. Gorenflo, A. A. Kilbas, F. Mainardi, and S. V. Rogosin. *Mittag-Leffler functions, Related Topics and Applications*. Springer Monogr. Math., Springer, Heidelberg, Springer, 2014.

[8] Fahad Jarad and Thabet Abdeljawad. Generalized fractional derivatives and Laplace transform. *Discrete & Continuous Dynamical Systems-S*, pages 1775–1786, 2019.

[9] Anatoly A Kilbas, Hari M Srivastava, and Juan J Trujillo. *Theory and Applications of Fractional Differential Equations*, volume 204. North-Holland mathematics studies, 2006.

[10] Shu Liang, Leyi Wang, and George Yin. Fractional differential equation approach for convex optimization with convergence rate analysis. *Optimization Letters*, pages 1–11, 2019.

[11] Song Liu, Xiang Wu, Xian-Feng Zhou, and Wei Jiang. Asymptotical stability of Riemann–Liouville fractional nonlinear systems. *Nonlinear Dynamics*, 86(1):65–71, 2016.

[12] Y. Nesterov. *Introductory Lectures on Convex Optimization. A basic course*. Applied Optimization, 87. Kluwer Academic Publishers, Boston, MA, 2004.

[13] Joel A Rosenfeld and Warren E Dixon. Approximating the caputo fractional derivative through the mittag-leffler reproducing kernel hilbert space and the kernelized adams–bashforth–moulton method. *SIAM Journal on Numerical Analysis*, 55(3):1201–1217, 2017.

[14] Andrzej P Ruszczynski. *Nonlinear optimization*, volume 13. Princeton university press, 2006.

[15] W. R. Schneider. Completely monotone generalized Mittag-Leffler functions. *Exposition. Math.*, 14:3–24, 1996.

[16] Damien Scieur, Vincent Roulet, Francis Bach, and Alexandre d’Aspremont. Integration methods and optimization algorithms. In *Advances in Neural Information Processing Systems*, pages 1109–1118, 2017.

[17] Duarte Valério, Juan J Trujillo, Margarita Rivero, JA Tenreiro Machado, and Dumitru Baleanu. Fractional calculus: A survey of useful formulas. *The European Physical Journal Special Topics*, 222(8):1827–1846, 2013.

(P. V. Hai) 
E-mail address: phamviethai86@gmail.com

(Joel A. Rosenfeld) Department of Mathematics and Statistics, University of South Florida, Tampa, Florida. 
E-mail address: rosenfeldj@usf.edu