Mathematical Models and Methods in Applied Sciences
Vol. 33, No. 2 (2023) 377–453

DOI: 10.1142/S0218202523500100

Time fractional gradient flows: Theory and numerics

Wenbo Li

Department of Mathematics, University of Tennessee,
Knoxville TN 37996, USA

Current Address: Institute of Computational Mathematics and Scientific/
Engineering Computing of the Chinese Academy of Sciences,
Beijing 100190, P. R. China
liwenbo@lsec.cc.ac.cn

Abner J. Salgado*

Department of Mathematics, University of Tennessee,
Knoxville TN 37996, USA
asalgad1@utk.edu

Received 3 January 2021
Revised 16 August 2022
Accepted 15 December 2022
Published 23 February 2023
Communicated by Qiang Du

We develop the theory of fractional gradient flows: an evolution aimed at the minimization of a convex, lower semicontinuous energy, with memory effects. This memory is characterized by the fact that the negative of the (sub)gradient of the energy equals the so-called Caputo derivative of the state. We introduce the notion of energy solutions, for which we provide existence, uniqueness and certain regularizing effects. We also consider Lipschitz perturbations of this energy. For these problems we provide an *a posteriori* error estimate and show its reliability. This estimate depends only on the problem data, and imposes no constraints between consecutive time-steps. On the basis of this estimate we provide an *a priori* error analysis that makes no assumptions on the smoothness of the solution.

Keywords: Caputo derivative; gradient flows; *a posteriori* error estimate; variable time stepping.

AMS Subject Classification: 34G20, 35R11, 65J08, 65M06, 65M15, 65M50

1. Introduction

In recent times problems involving fractional derivatives have garnered considerable attention, as it is claimed that they better describe certain fundamental relations

*Corresponding author.

between the processes of interest; see, for instance Refs. 44, 19 and 75. In this, and many other references the models considered are linear. However, it is well known that real world phenomena are not linear, not even smooth. It is only natural then to consider nonlinear/nonsmooth models with fractional derivatives.

The purpose of this work is to develop the theory and numerical analysis of so-called *time-fractional* gradient flows: an evolution equation aimed at the minimization of a convex and lower semicontinuous (l.s.c.) energy, but where the evolution has memory effects. This memory is characterized by the fact that the negative of the (sub)gradient of the energy equals the so-called Caputo derivative of the state.

The Caputo derivative, introduced in Ref. 14, is one of the existing models of fractional derivatives. It is defined, for $\alpha \in (0,1)$, by

$$D_c^\alpha w(t) = \frac{1}{\Gamma(1-\alpha)} \int_0^t \frac{\dot{w}(r)}{(t-r)^\alpha} \mathrm{d}r, \tag{1.1}$$

where Γ denotes the Gamma function. This definition, from the onset, seems unnatural. To define a derivative of a fractional order, it seems necessary for the function to be at least differentiable. Below we briefly describe several attempts at circumventing this issue. We focus, in particular, on the results developed in a series of papers by Li and Liu, see Refs. 36, 39, 37 and 38, where they developed a distributional theory for this derivative; see also Ref. 23. The authors of these works also constructed, in Ref. 37, so-called deconvolution schemes that aim at discretizing this derivative. With the help of this definition and the schemes that they develop the authors were able to study several classes of equations, in particular time fractional gradient flows.

Let us be precise in what we mean by this term. Let $T > 0$ be a final time, $\mathcal{H}$ be a separable Hilbert space, $\Phi : \mathcal{H} \to \mathbb{R} \cup \{+\infty\}$ be a convex and l.s.c. functional, which we will call *energy*. Given $u_0 \in \mathcal{H}$, and $f : (0,T] \to \mathcal{H}$ we seek for a function $u : [0,T] \to \mathcal{H}$ that satisfies

$$\begin{cases} D_c^\alpha u(t) + \partial\Phi(u(t)) \ni f(t), & t \in (0,T], \\ u(0) = u_0, \end{cases} \tag{1.2}$$

where by $\partial\Phi$ we denote the subdifferential of Φ. Our objectives in this work can be stated as follows: We will introduce the notion of "energy solutions" of (1.2), and we will refine the results regarding existence, uniqueness, and regularizing effects provided in Ref. 39. This will be done by generalizing, to non-uniform time steps the "deconvolution" schemes of Refs. 37 and 39, and developing a sort of "fractional minimizing movements" scheme. We will also provide an *a priori* error estimate that seems optimal in light of the regularizing effects proved above. We also develop an *a posteriori* error estimate, in the spirit of Ref. 51 and show its reliability.

We comment, in passing, that nonlinear evolution problems with fractional time derivative have been considered in other works. From a modeling point of view, their advantages have been observed in Refs. 19 and 15. Some other types of nonlinear problems have been studied in Refs. 9, 67, 2, 34, 33, 65, 74, 53 and 62, where for a particular type of nonlinear problem, other "energy dissipation inequalities" than those we obtain are derived. Regularity properties for nonlinear problems with fractional time derivatives have been obtained in Refs. 32, 18, 31, 1, 73, 72, 71 and 69. Of particular interest to us are Ref. 39 which we described above and Ref. 3 which also considers time fractional gradient flows. The assumptions on the data, however, are slightly different than ours. As such, some of the results in Ref. 3 are stronger, and some weaker than ours; in particular, we conduct a numerical analysis of this problem. Nevertheless, we refer to this reference for a nice historical account and particular applications to PDEs.

Before proceeding any further, let us detail here what we believe are the major advancements that this contribution aims to put forward:

- **Theory: energy solutions.** We develop the theory of energy solutions for problems of the form (1.2). All that is needed in this context, is for the energy Φ to be convex, l.s.c., and bounded from below. No other assumptions regarding smoothness or structure of the energy are made. Nevertheless, we prove existence, uniqueness, and certain regularizing effects within this solution class.
- **Theory: Lipschitz perturbations.** The theory of energy solutions is extended to the case of a Lipschitz perturbation to our convex energy. In this case we also develop an existence and uniqueness theory.
- **Numerics: unconditionally stable scheme.** Our extension to arbitrary time steps of existing deconvolution schemes is, one of the few discretizations of the Caputo derivative that is unconditionally stable over arbitrary meshes. Most works consider either uniform, or suitable graded temporal meshes. We comment that, while this work was under review, another discretization with similar properties appeared in Ref. 43, where the authors consider the time fractional Allen–Cahn equation. Under some restrictions on the time step and the spatial discretization parameter, the scheme of Ref. 43 is also maximum principle preserving.
- **Numerics: *a posteriori* error estimator.** We construct a reliable *a posteriori* error estimator for our numerical scheme. This is the first of its kind for time-fractional problems.
- **Numerics: error estimates.** We prove optimal error estimates for our numerical scheme. These do not assume any regularity beyond what the notion of energy solutions accommodates. For some special cases, where there is additional regularity, these are improved.

Our presentation will be organized as follows. We will establish notation and the framework we will adopt in Sec. 2. Here, in particular, we will study several properties of a particular space, which we denote by $L^p_\alpha(0,T;\mathcal{H})$, and that will be used to characterize the requirements on the right-hand side f of (1.2). In addition, we also review the various proposed generalizations of the classical definition of Caputo derivatives, with particular attention to that of Refs. 36, 39 and 38; since this is the one we shall adopt. In Sec. 3, we generalize the deconvolution schemes of Refs. 37 and 39 and their properties, to the case of non-uniform time stepping. Many of the simple properties of these schemes are lost in this case, but we retain enough of them for our purposes. Section 4 introduces the notion of energy solutions for (1.2) and shows existence and uniqueness of these. This is accomplished by introducing, on the basis of our generalized deconvolution formulas, a fractional minimizing movements scheme; and showing that the discrete solutions have enough compactness to pass to the limit in the size of the partition. In Sec. 5, we provide an error analysis of the fractional minimizing movements scheme. First, we show how an error estimate follows as a side result from the existence proof. Then, in the spirit of Ref. 51, we provide an *a posteriori* error estimator for our scheme and show its reliability. This estimator is then used to independently show rates of convergence. This section is concluded with some particular instances in which the rate of convergence can be improved. Section 6 is dedicated to the case in which we allow a Lipschitz perturbation of the subdifferential. We extend the existence, uniqueness, *a priori*, and *a posteriori* approximation results of the fractional gradient flow. To show the extent of applicability of our developments, in Sec. 7, we present a series of example problems to which our setting applies. Finally, Sec. 8 presents some simple numerical experiments that illustrate, explore, and expand our theory.

2. Notation and Preliminaries

Let us begin by presenting the main notation and assumptions we shall operate under. We will denote by $T \in (0, \infty)$ our final (positive) time. By $\mathcal{H}$ we will always denote a separable Hilbert space with scalar product $\langle \cdot, \cdot \rangle$ and norm $\| \cdot \|$. As it is by now customary, by C we will denote a nonessential constant whose value may change at each occurrence.

2.1. *Convex energies*

The *energy* will be a convex, l.s.c., functional $\Phi : \mathcal{H} \to \mathbb{R} \cup \{+\infty\}$ with nonempty effective domain of definition, that is,

$$D(\Phi) = \{w \in \mathcal{H} : \Phi(w) < +\infty\} \neq \emptyset.$$

We will always assume that our energy is bounded from below, that is,

$$\Phi_{\inf} = \inf_{u \in \mathcal{H}} \Phi(u) > -\infty.$$

As we are not assuming smoothness in our energy beyond convexity, a useful substitute for its derivative is the subdifferential, that is,

$$\partial \Phi(w) = \{\xi \in \mathcal{H} : \langle \xi, v - w \rangle \leq \Phi(v) - \Phi(w)\ \forall\, v \in \mathcal{H}\}.$$

The effective domain of the subdifferential is $D(\partial\Phi) = \{w \in \mathcal{H} : \partial\Phi(w) \neq \emptyset\}$. Recall that, in our setting, we always have that $\overline{D(\partial\Phi)} = \overline{D(\Phi)}$. We refer the reader to Refs. 16 and 56 for basic facts on convex analysis.

In applications, it is sometimes useful to obtain error estimates on (semi)norms stronger than those of the ambient space, and that are dictated by the structure of the energy. For this reason, we introduce the following *coercivity modulus* of Φ, see Definition 2.3 in Ref. 51.

Definition 2.1. (Coercivity modulus) For every $w_1 \in D(\Phi)$ and $w_2 \in D(\partial\Phi)$, let $\sigma(w_1; w_2) \geq 0$ be

$$\sigma(w_1; w_2) = \Phi(w_2) - \Phi(w_1) - \sup_{\xi \in \partial\Phi(w_1)} \langle \xi, w_2 - w_1 \rangle.$$

Then for every $w_1, w_2 \in D(\partial\Phi)$ we define

$$\begin{aligned}\rho(w_1, w_2) &= \sigma(w_1; w_2) + \sigma(w_2; w_1) \\ &= \inf_{\xi_1 \in \partial\Phi(w_1), \xi_2 \in \partial\Phi(w_2)} \langle \xi_1 - \xi_2, w_1 - w_2 \rangle.\end{aligned}$$

We comment that, by the definition, $\rho(\cdot, \cdot)$ is symmetric, whereas $\sigma(\cdot; \cdot)$ might not be. Furthermore, the separability of $\mathcal{H}$ guarantees that σ and ρ are both Borel measurable; see Remark 2.4 in Ref. 51. One may also refer to Sec. 2.3 of Ref. 51 for discussions and properties of σ and ρ for certain choices of Φ. Definition 2.1 enables us to write

$$\xi \in \partial\Phi(w) \Leftrightarrow \langle \xi, v - w \rangle + \sigma(w; v) \leq \Phi(v) - \Phi(w), \quad \forall\, v \in \mathcal{H}. \tag{2.1}$$

2.2. *Vector-valued time dependent functions*

We will follow standard notation regarding Bochner spaces of vector-valued functions, see Sec. 1.5 in Ref. 54. For any $w \in L^1(0, T; \mathcal{H})$ and $E \subset [0, T]$ that is measurable, we define the average by

$$\fint_E w(t)\mathrm{d}t = \frac{1}{|E|} \int_E w(t)\mathrm{d}t,$$

where $|E|$ denotes the Lebesgue measure of E.

Since eventually we will have to deal with time discretization, we also introduce notation for time-discrete vector-valued functions. Let $\mathcal{P}$ be a partition of the time interval $[0, T]$

$$\mathcal{P} = \{0 = t_0 < t_1 < \cdots < t_{N-1} < t_N = T\}, \tag{2.2}$$

with variable steps $\tau_n = t_n - t_{n-1}$ and $\tau = \max\{\tau_n : n \in \{1, \ldots, N\}\}$. We will always denote by N the size of a partition. For $t \in [0, T]$ we define

$$\lfloor t \rfloor_{\mathcal{P}} = \max\{r \in \mathcal{P} : r < t\}, \quad \lceil t \rceil_{\mathcal{P}} = \min\{r \in \mathcal{P} : t \leq r\},$$

and $n(t)$ to be the index of $\lceil t \rceil_{\mathcal{P}}$, so that $t \in (\lfloor t \rfloor_{\mathcal{P}}, \lceil t \rceil_{\mathcal{P}}] = (t_{n(t)-1}, t_{n(t)}]$. Given a partition $\mathcal{P}$, for $\mathbf{W} = \{W_i\}_{i=1}^N \subset \mathcal{H}^N$ we define its *piecewise constant* interpolant with respect to $\mathcal{P}$ to be the function $\overline{W}_{\mathcal{P}} \in L^\infty(0, T; \mathcal{H})$ defined by

$$\overline{W}_{\mathcal{P}}(t) = W_{n(t)}. \tag{2.3}$$

2.2.1. *The space* $L^p_\alpha(0,T;\mathcal{H})$

To quantify the assumptions we need on the right-hand side f of (1.2) we introduce the following space.

Definition 2.2. (Space $L^p_\alpha(0,T;\mathcal{H})$) Let $p \in [1,\infty)$ and $\alpha \in (0,1)$. We say that the function $w : [0,T] \to \mathcal{H}$ belongs to the space $L^p_\alpha(0,T;\mathcal{H})$ if and only if

$$\|w\|_{L^p_\alpha(0,T;\mathcal{H})} = \sup_{t\in[0,T]} \left(\int_0^t (t-s)^{\alpha-1} \|w(s)\|^p \mathrm{d}s \right)^{1/p} < \infty. \tag{2.4}$$

Let us show some basic embedding results about this space.

Proposition 2.3. (Embedding) *Let* $p \in [1,\infty)$, $\alpha \in (0,1)$, *and* $q > p/\alpha$. *Then we have that*

$$L^q(0,T;\mathcal{H}) \hookrightarrow L^p_\alpha(0,T;\mathcal{H}) \hookrightarrow L^p(0,T;\mathcal{H}).$$

Proof. The second embedding is immediate. For any $t \in (0,T]$

$$\int_0^t \|w(s)\|^p \mathrm{d}s \le \sup_{s\in[0,t]} (t-s)^{1-\alpha} \int_0^t (t-s)^{\alpha-1}\|w(s)\|^p \mathrm{d}s \le T^{1-\alpha} \|w\|^p_{L^p_\alpha(0,T;\mathcal{H})},$$

where we used that $1-\alpha > 0$.

The proof of the first embedding is a simple application of Hölder inequality. Indeed, we have

$$\left(\int_0^t (t-s)^{\alpha-1} \|w(s)\|^p \mathrm{d}s \right)^{1/p} \le \left(\frac{q-p}{q\alpha-p} \right)^{(q-p)/q} t^{\alpha-p/q} \|w\|_{L^q(0,t;\mathcal{H})},$$

and hence

$$\|w\|_{L^p_\alpha(0,T;\mathcal{H})} \le \left(\frac{q-p}{q\alpha-p} \right)^{(q-p)/q} T^{\alpha-p/q} \|w\|_{L^q(0,T;\mathcal{H})}, \tag{2.5}$$

as we intended to show. □

When dealing with discretization we will approximate the right-hand side f of (1.2) by its local averages over a partition $\mathcal{P}$. Thus, we must provide a bound on this operation that is independent of the partition.

Lemma 2.4. (Continuity of averaging) *Let* $p \in [1,\infty)$, $\alpha \in (0,1)$, $f \in L^p_\alpha(0,T;\mathcal{H})$, *and* $\mathcal{P}$ *be a partition of* $[0,T]$ *as in* (2.2). *Define* $\mathbf{F} = \{ ⨍_{t_{n-1}}^{t_n} f(t)\mathrm{d}t\}_{n=1}^N \subset \mathcal{H}^N$ *and let* $\overline{F}_\mathcal{P}$ *be defined as in* (2.3). *Then, there exists a constant* $C > 0$ *only depending on* p *and* α *such that*

$$\|\overline{F}_\mathcal{P}\|_{L^p_\alpha(0,T;\mathcal{H})} \le C \|f\|_{L^p_\alpha(0,T;\mathcal{H})}.$$

Proof. Let $p \in (1,\infty)$. We first, for $n \in \{1,\ldots,N\}$, bound the integral

$$\int_0^{t_n} (t_n - s)^{\alpha-1} \|\overline{F}_\mathcal{P}(s)\|^p \mathrm{d}s.$$

To achieve this, we decompose this integral as

$$\int_0^{t_n}(t_n-s)^{\alpha-1}\|\overline{F}_{\mathcal{P}}(s)\|^p\mathrm{d}s=\sum_{k=1}^{n}\int_{t_{k-1}}^{t_k}(t_n-s)^{\alpha-1}\|\overline{F}_{\mathcal{P}}(s)\|^p\mathrm{d}s$$

$$=\sum_{k=1}^{n}\|F_k\|^p\int_{t_{k-1}}^{t_k}(t_n-s)^{\alpha-1}\mathrm{d}s. \tag{2.6}$$

We use Hölder inequality in the definition of F_k to obtain that

$$\|F_k\|^p=\left\|\fint_{t_{k-1}}^{t_k}f(s)\mathrm{d}s\right\|^p$$

$$\le\fint_{t_{k-1}}^{t_k}(t_n-s)^{\alpha-1}\|f(s)\|^p\mathrm{d}s\left(\fint_{t_{k-1}}^{t_k}(t_n-s)^{\frac{1-\alpha}{p-1}}\mathrm{d}s\right)^{p-1}. \tag{2.7}$$

Since, for every $p\in(1,\infty)$ the function $s\mapsto s^{\alpha-1}$ belongs to the Muckenhoupt class $A_p(\mathbb{R}_+)$, see Example 7.1.7 in Ref. 29, there exists a constant $C_{p,\alpha}$ that only depends on p and α such that

$$\fint_a^b s^{\alpha-1}\mathrm{d}s\left(\fint_a^b s^{\frac{1-\alpha}{p-1}}\mathrm{d}s\right)^{p-1}\le C_{p,\alpha},\quad\forall\, 0\le a<b.$$

Therefore, for any k, we have

$$\fint_{t_{k-1}}^{t_k}(t_n-s)^{\alpha-1}\mathrm{d}s\left[\fint_{t_{k-1}}^{t_k}(t_n-s)^{\frac{1-\alpha}{p-1}}\mathrm{d}s\right]^{p-1}$$

$$=\fint_{t_n-t_{k-1}}^{t_n-t_k}s^{\alpha-1}\mathrm{d}s\left[\fint_{t_n-t_{k-1}}^{t_n-t_k}s^{\frac{1-\alpha}{p-1}}\mathrm{d}s\right]^{p-1}\le C_{p,\alpha}. \tag{2.8}$$

Substituting (2.7) and (2.8) into (2.6) we get

$$\int_0^{t_n}(t_n-s)^{\alpha-1}\|\overline{F}_{\mathcal{P}}(s)\|^p\mathrm{d}s\le\sum_{k=1}^{n}C_{p,\alpha}\int_{t_{k-1}}^{t_k}(t_n-s)^{\alpha-1}\|f(s)\|^p\mathrm{d}s$$

$$=C_{p,\alpha}\int_0^{t_n}(t_n-s)^{\alpha-1}\|f(s)\|^p\mathrm{d}s\le C_{p,\alpha}\|f\|^p_{L^p_\alpha(0,T;\mathcal{H})}.$$

Now consider $t\in[0,T]$. Taking advantage of the estimate we obtained above, we write

$$\int_0^t(t-s)^{\alpha-1}\|\overline{F}_{\mathcal{P}}(s)\|^p\mathrm{d}s=\int_0^{\lfloor t\rfloor_{\mathcal{P}}}(t-s)^{\alpha-1}\|\overline{F}_{\mathcal{P}}(s)\|^p\mathrm{d}s$$

$$+\int_{\lfloor t\rfloor_{\mathcal{P}}}^{t}(t-s)^{\alpha-1}\|\overline{F}_{\mathcal{P}}(s)\|^p\mathrm{d}s$$

$$=\int_0^{\lfloor t\rfloor_{\mathcal{P}}}(t-s)^{\alpha-1}\|\overline{F}_{\mathcal{P}}(s)\|^p\mathrm{d}s$$

$$
\begin{aligned}
&+ \|\overline{F}_{\mathcal{P}}(\lceil t \rceil_{\mathcal{P}})\|^p \int_{\lfloor t \rfloor_{\mathcal{P}}}^{t} (t-s)^{\alpha-1} \mathrm{d}s \\
&\le \int_0^{\lfloor t \rfloor_{\mathcal{P}}} (\lfloor t \rfloor_{\mathcal{P}} - s)^{\alpha-1} \|\overline{F}_{\mathcal{P}}(s)\|^p \mathrm{d}s \\
&\quad + \|\overline{F}_{\mathcal{P}}(\lceil t \rceil_{\mathcal{P}})\|^p \int_{\lfloor t \rfloor_{\mathcal{P}}}^{\lceil t \rceil_{\mathcal{P}}} (\lceil t \rceil_{\mathcal{P}} - s)^{\alpha-1} \mathrm{d}s \\
&\le C_{p,\alpha} \|f\|^p_{L^p_\alpha(0,T;\mathcal{H})} \\
&\quad + \int_0^{\lceil t \rceil_{\mathcal{P}}} (\lceil t \rceil_{\mathcal{P}} - s)^{\alpha-1} \|\overline{F}(s)\|^p \mathrm{d}s \\
&\le 2C_{p,\alpha} \|f\|^p_{L^p_\alpha(0,T;\mathcal{H})}.
\end{aligned} \tag{2.9}
$$

Therefore by taking supremum over $t \in [0,T]$ and $C = (2C_{p,\alpha})^{1/p}$, we finish the proof of this lemma.

For $p = 1$, the proof proceeds almost the same way as before. The only difference worth noting is that, instead of (2.7), we have

$$
\|F_k\| = \left\| ⨍_{t_{k-1}}^{t_k} f(s)\mathrm{d}s \right\| \le ⨍_{t_{k-1}}^{t_k} (t_n - s)^{\alpha-1} \|f(s)\| \mathrm{d}s \sup_{s \in [t_{k-1}, t_k]} \frac{1}{(t_n - s)^{\alpha-1}}.
$$

Next, we observe that, since $\alpha - 1 \in (-1, 0)$, then the function $s \mapsto s^{\alpha-1}$ belongs to the Muckenhoupt class $A_1(\mathbb{R}_+)$. Thus,

$$
\sup_{s \in [a,b]} \frac{1}{s^{\alpha-1}} ⨍_a^b s^{\alpha-1} \mathrm{d}s \le C_\alpha, \quad \forall\, 0 \le a < b.
$$

With this information, the proof proceeds without change. □

It turns out that averaging is not only continuous, but possesses suitable approximation properties in this space. Namely, we have a control on the difference between fractional integrals of $f \in L^p_\alpha(0,T;\mathcal{H})$ and its averages.

Lemma 2.5. (Approximation) *Let $p \in [1,\infty)$, $\alpha \in (0,1)$, $f \in L^p_\alpha(0,T;\mathcal{H})$, and $\mathcal{P}$ be a partition of $[0,T]$ as in (2.2). Let p' be the Hölder conjugate of p, $\mathbf{F} = \{⨍_{t_{n-1}}^{t_n} f(t)\mathrm{d}t\}_{n=1}^N \subset \mathcal{H}^N$, and let $\overline{F}_{\mathcal{P}}$ be defined as in (2.3). Then we have*

$$
\begin{aligned}
\sup_{t \in [0,T]} \left\| \int_0^t (t-s)^{\alpha-1} \left(f(s) - \overline{F}_{\mathcal{P}}(s) \right) \mathrm{d}s \right\| &\le C \tau^{\alpha/p'} \|f - \overline{F}_{\mathcal{P}}\|_{L^p_\alpha(0,T;\mathcal{H})} \\
&\le C' \tau^{\alpha/p'} \|f\|_{L^p_\alpha(0,T;\mathcal{H})},
\end{aligned} \tag{2.10}
$$

where the constants C, C' depend only on p and α. In addition, for any $\beta \in (0,1)$ we also have

$$\sup_{r\in[0,T]} \int_0^r (r-t)^{\alpha-1} \left\| \int_0^t (t-s)^{\beta-1}(f(s) - \overline{F}_{\mathcal{P}}(s))\mathrm{d}s \right\|^p \mathrm{d}t$$
$$\leq C_1 \tau^{p\beta} \|f - \overline{F}_{\mathcal{P}}\|^p_{L^p_\alpha(0,T;\mathcal{H})} \leq C_1' \tau^{p\beta} \|f\|^p_{L^p_\alpha(0,T;\mathcal{H})}, \qquad (2.11)$$

where the constants C_1, C_1' depend on p, α, and β. As usual, when $p = 1$, we have $p' = \infty$ and $1/p'$ is treated as 0.

Proof. We first notice that the second inequalities in both (2.10) and (2.11) follow directly from Theorem 2.4 and the triangle inequality.

To show the first inequality in (2.10), given $\mathcal{P}$ we consider $t \in [0,T]$. Using that $f - \overline{F}_{\mathcal{P}}$ has zero mean on each subinterval of the partition, we can write

$$\begin{aligned}
&\int_0^t (t-s)^{\alpha-1} \left(f(s) - \overline{F}_{\mathcal{P}}(s)\right) \mathrm{d}s \\
&\quad = \int_{\lfloor t \rfloor_{\mathcal{P}}}^t (t-s)^{\alpha-1}(f(s) - \overline{F}_{\mathcal{P}}(s))\mathrm{d}s \\
&\qquad + \sum_{k=1}^{n(t)-1} \int_{t_{k-1}}^{t_k} (t-s)^{\alpha-1}(f(s) - \overline{F}_{\mathcal{P}}(s))\mathrm{d}s \\
&\quad = \int_{\lfloor t \rfloor_{\mathcal{P}}}^t (t-s)^{\alpha-1}(f(s) - \overline{F}_{\mathcal{P}}(s))\mathrm{d}s + \sum_{k=1}^{n(t)-1} \int_{t_{k-1}}^{t_k} ((t-s)^{\alpha-1} \\
&\qquad - (t-t_{k-1})^{\alpha-1})(f(s) - \overline{F}_{\mathcal{P}}(s))\mathrm{d}s = \mathrm{I}_1(t) + \mathrm{I}_2(t). \qquad (2.12)
\end{aligned}$$

For the first term, denoted $I_1(t)$, we have

$$\begin{aligned}
\|\mathrm{I}_1(t)\| &\leq \left(\int_{\lfloor t \rfloor_{\mathcal{P}}}^t (t-s)^{\alpha-1} \left\|f(s) - \overline{F}_{\mathcal{P}}(s)\right\|^p \mathrm{d}s \right)^{1/p} \left(\int_{\lfloor t \rfloor_{\mathcal{P}}}^t (t-s)^{\alpha-1}\mathrm{d}s \right)^{1/p'} \\
&\leq \|f - \overline{F}_{\mathcal{P}}\|_{L^p_\alpha(0,T;\mathcal{H})} \left(\frac{1}{\alpha}(t - \lfloor t \rfloor_{\mathcal{P}})^\alpha \right)^{1/p'} \leq C_1 \tau^{\alpha/p'} \|f - \overline{F}_{\mathcal{P}}\|_{L^p_\alpha(0,T;\mathcal{H})},
\end{aligned}$$

where C_1 only depends on p and α. For the second term, noticing that $t - t_{k-1} + \tau > t - s$ for $s \in (t_{k-1}, t_k)$ we have

$$\begin{aligned}
\|\mathrm{I}_2\| &\leq \int_0^{\lfloor t \rfloor_{\mathcal{P}}} \left((t-s)^{\alpha-1} - (t-s+\tau)^{\alpha-1}\right) \|f(s) - \overline{F}_{\mathcal{P}}(s)\|\mathrm{d}s \\
&\leq \left[\int_0^{\lfloor t \rfloor_{\mathcal{P}}} (t-s)^{\alpha-1} \left\|f(s) - \overline{F}_{\mathcal{P}}(s)\right\|^p \mathrm{d}s \right]^{1/p}
\end{aligned}$$

$$\times \left[\int_0^{\lfloor t \rfloor_{\mathcal{P}}} (t-s)^{\alpha-1} \left[1 - \left[\frac{t-s+\tau}{t-s} \right]^{\alpha-1} \right]^{p'} \mathrm{d}s \right]^{1/p'}$$
$$\leq \|f - \overline{F}_{\mathcal{P}}\|_{L^p_\alpha(0,T;\mathcal{H})} \left(\int_0^{\lfloor t \rfloor_{\mathcal{P}}} (t-s)^{\alpha-1} - (t-s+\tau)^{\alpha-1} \mathrm{d}s \right)^{1/p'} .$$

Since

$$\int_0^{\lfloor t \rfloor_{\mathcal{P}}} (t-s)^{\alpha-1} - (t-s+\tau)^{\alpha-1} \mathrm{d}s = \frac{1}{\alpha}(t^\alpha - (t - \lfloor t \rfloor_{\mathcal{P}})^\alpha - (t+\tau)^\alpha + (t - \lfloor t \rfloor_{\mathcal{P}} + \tau)^\alpha)$$
$$\leq \frac{1}{\alpha}((t - \lfloor t \rfloor_{\mathcal{P}} + \tau)^\alpha - (t - \lfloor t \rfloor_{\mathcal{P}})^\alpha) \leq \frac{\tau^\alpha}{\alpha},$$

we obtain

$$\|\mathrm{I}_2(t)\| \leq C_2 \tau^{\alpha/p'} \|f - \overline{F}_{\mathcal{P}}\|_{L^p_\alpha(0,T;\mathcal{H})},$$

and (2.10) follows after combining the bounds for $\mathrm{I}_1(t)$ and $\mathrm{I}_2(t)$ that we have obtained.

To prove (2.11) we apply the Hölder inequality to (2.12) with α replaced by β to get

$$\left\| \int_0^t (t-s)^{\beta-1} \left(f(s) - \overline{F}_{\mathcal{P}}(s) \right) \mathrm{d}s \right\|^p \leq \mathrm{II}_1(t)^{p-1} \cdot (\mathrm{II}_2(t) + \mathrm{II}_3(t)),$$

where

$$\mathrm{II}_1(t) = \int_{\lfloor t \rfloor_{\mathcal{P}}}^t (t-s)^{\beta-1} \mathrm{d}s + \sum_{k=1}^{n(t)-1} \int_{t_{k-1}}^{t_k} \left[(t-s)^{\beta-1} - (t-t_{k-1})^{\beta-1} \right] \mathrm{d}s,$$
$$\mathrm{II}_2(t) = \int_{\lfloor t \rfloor_{\mathcal{P}}}^t (t-s)^{\beta-1} \left\| f(s) - \overline{F}_{\mathcal{P}}(s) \right\|^p \mathrm{d}s,$$
$$\mathrm{II}_3(t) = \sum_{k=1}^{n(t)-1} \int_{t_{k-1}}^{t_k} \left((t-s)^{\beta-1} - (t-t_{k-1})^{\beta-1} \right) \left\| f(s) - \overline{F}_{\mathcal{P}}(s) \right\|^p \mathrm{d}s.$$

Arguing as in the bound for $\mathrm{I}_2(t)$

$$\mathrm{II}_1(t) = \frac{1}{\beta}(t - \lfloor t \rfloor_{\mathcal{P}})^\beta + \int_0^{\lfloor t \rfloor_{\mathcal{P}}} [(t-s)^{\beta-1} - (t-s+\tau)^{\beta-1}] \mathrm{d}s \leq \frac{2}{\beta}\tau^\beta.$$

Thus, to obtain (2.11) it suffices to show that, for every $r \in [0,T]$,

$$\int_0^r (r-t)^{\alpha-1} (\mathrm{II}_2(t) + \mathrm{II}_3(t)) \mathrm{d}t \leq C_2 \tau^\beta \|f - \overline{F}_{\mathcal{P}}\|^p_{L^p_\alpha(0,T;\mathcal{H})}$$

with some constant C_2 only depending on p, α, and β. To estimate the fractional integral of II_2 by Fubini's theorem we have

$$\int_0^r (r-t)^{\alpha-1}\mathrm{II}_2(t)\mathrm{d}t = \int_0^r \|f(s)-\overline{F}_{\mathcal{P}}(s)\|^p \times \int_s^{\lceil s\rceil_{\mathcal{P}}\wedge r} (r-t)^{\alpha-1}(t-s)^{\beta-1}\mathrm{d}t\mathrm{d}s, \tag{2.13}$$

where we set $a\wedge b=\min\{a,b\}$. We claim that there exists a constant C_3 depending on α and β such that

$$\int_s^{\lceil s\rceil_{\mathcal{P}}\wedge r} (r-t)^{\alpha-1}(t-s)^{\beta-1}\mathrm{d}t \le C_3(r-s)^{\alpha-1}\tau^\beta. \tag{2.14}$$

On the one hand, for $r-s\le 2\tau$, we simply have

$$\begin{aligned}\int_s^{\lceil s\rceil_{\mathcal{P}}\wedge r} (r-t)^{\alpha-1}(t-s)^{\beta-1}\mathrm{d}t &\le \int_s^r (r-t)^{\alpha-1}(t-s)^{\beta-1}\mathrm{d}t\\ &= \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}(r-s)^{\alpha+\beta-1}\\ &\le \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}(r-s)^{\alpha-1}(2\tau)^\beta.\end{aligned}$$

On the other hand, if $r-s>2\tau$, then

$$\begin{aligned}\int_s^{\lceil s\rceil_{\mathcal{P}}\wedge r} (r-t)^{\alpha-1}(t-s)^{\beta-1}\mathrm{d}t &\le \int_s^{s+\tau} (r-t)^{\alpha-1}(t-s)^{\beta-1}\mathrm{d}t\\ &\le \int_s^{s+\tau} \left(\frac{r-s}{2}\right)^{\alpha-1}(t-s)^{\beta-1}\mathrm{d}t\\ &= \frac{2^{1-\alpha}}{\beta}(r-s)^{\alpha-1}\tau^\beta.\end{aligned}$$

Therefore (2.14) is proved, and thus (2.14) implies that

$$\begin{aligned}\int_0^r (r-t)^{\alpha-1}\mathrm{II}_2(t)\mathrm{d}t &\le C_3\tau^\beta\int_0^r (r-s)^{\alpha-1}\left\|f(s)-\overline{F}_{\mathcal{P}}(s)\right\|^p\mathrm{d}s\\ &\le C_3\tau^\beta\|f-\overline{F}_{\mathcal{P}}\|^p_{L^p_\alpha(0,T;\mathcal{H})}.\end{aligned}$$

For $\mathrm{II}_3(t)$, we again apply Fubini's theorem to obtain

$$\int_0^r (r-t)^{\alpha-1}\mathrm{II}_3(t)\mathrm{d}t = \int_0^r \left\|f(s)-\overline{F}_{\mathcal{P}}(s)\right\|^p \times \int_s^r (r-t)^{\alpha-1}\left((t-s)^{\beta-1}-(t-s+\tau)^{\beta-1}\right)\mathrm{d}t\mathrm{d}s.$$

To conclude, we claim that

$$A = \int_s^r (r-t)^{\alpha-1}((t-s)^{\beta-1}-(t-s+\tau)^{\beta-1})\mathrm{d}t \le C_4\tau^\beta(r-s)^{\alpha-1}, \tag{2.15}$$

for a constant C_4 depending on α and β. Indeed, if this is the case, we have

$$\begin{aligned}\int_0^r (r-t)^{\alpha-1}\mathrm{II}_3(t)\mathrm{d}t &\le C_4\tau^\beta \int_0^r (r-s)^{\alpha-1}\left\|f(s)-\overline{F}_{\mathcal{P}}(s)\right\|^p \mathrm{d}s\\ &\le C_4\tau^\beta \|f-\overline{F}_{\mathcal{P}}\|^p_{L^p_\alpha(0,T;\mathcal{H})},\end{aligned}$$

and we combine the estimates for $\mathrm{II}_2(t)$ and $\mathrm{II}_3(t)$ together and conclude the proof of (2.11).

Let us now turn to the proof of (2.15). First, if $r-s\le\tau$ then it suffices to observe that

$$A \le \int_s^r (r-t)^{\alpha-1}(t-s)^{\beta-1}\mathrm{d}t = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}(r-s)^{\alpha+\beta-1} \le \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}\tau^\beta(r-s)^{\alpha-1}.$$

Now, if $r-s>\tau$, we estimate as

$$\begin{aligned}A &= \int_s^r (r-t)^{\alpha-1}(t-s)^{\beta-1}\mathrm{d}t - \int_s^r (r-t)^{\alpha-1}(t-s+\tau)^{\beta-1}\mathrm{d}t\\ &= \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}(r-s)^{\alpha+\beta-1} - \int_{-\tau}^{r-s}(t+\tau)^{\beta-1}(r-t-s)^{\alpha-1}\mathrm{d}t\\ &\quad + \int_0^\tau (r-s-t+\tau)^{\alpha-1}t^{\beta-1}\mathrm{d}t\\ &= \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}((r-s)^{\alpha+\beta-1}-(r-s+\tau)^{\alpha+\beta-1})\\ &\quad + \int_0^\tau (r-s-t+\tau)^{\alpha-1}t^{\beta-1}\mathrm{d}t.\end{aligned}$$

The first term can be bounded using that $r-s>\tau$ as follows

$$\begin{aligned}(r-s)^{\alpha+\beta-1}-(r-s+\tau)^{\alpha+\beta-1} &\le \max\{\alpha+\beta-1,0\}\tau(r-s)^{\alpha+\beta-2}\\ &\le \tau^\beta(r-s)^{\alpha-1}.\end{aligned}$$

On the other hand, since for $t\in(0,\tau)$ we have that $r-s+\tau-t\ge r-s$, the second term can be estimated as

$$\int_0^\tau (r-s-t+\tau)^{\alpha-1}t^{\beta-1}\mathrm{d}t \le (r-s)^{\alpha-1}\int_0^\tau t^{\beta-1}\mathrm{d}t = \frac{1}{\alpha}(r-s)^{\alpha-1}\tau^\beta.$$

This concludes the proof. □

We refer the reader to Sec. 4 of Ref. 38 for further results concerning the space $L^p_\alpha(0,T;\mathcal{H})$.

2.3. *The Caputo derivative*

As we mentioned in the Introduction, the definition of the Caputo derivative, given in (1.1) seems unnatural. Smoothness of higher order is needed to define a fractional derivative. Several attempts at resolving this discrepancy have been proposed in the literature and we here quickly describe a few of them.

First, one of the main reasons that motivate practitioners to use, among the many possible definitions, the Caputo derivative (1.1) is, first, that $D_c^\alpha 1 = 0$ and second that this derivative allows one to pose initial value problems like (1.2). However, it is by now known that even in the linear case solutions of problems involving the Caputo derivative possess a weak singularity in time Refs. 61, 60 and 59. This singular behavior of the solution forces one to wonder: If fractional derivatives describe processes with memory, why is it sufficient to know the state at one particular point (initial condition) to uniquely describe the state at all future times? Is it possible that the singularity is precisely caused by the fact that we are ignoring the past states of the system? This motivates the following: Set $w(t) = w_0$ for $t \leq 0$. Therefore,

$$\begin{aligned} D_c^\alpha w(t) &= \frac{1}{\Gamma(1-\alpha)} \int_{-\infty}^t \frac{\dot{w}(r)}{(t-r)^\alpha} \mathrm{d}r = \frac{1}{\Gamma(1-\alpha)} \int_{-\infty}^t \frac{(w(r)-w(t))^{\cdot}}{(t-r)^\alpha} \mathrm{d}r \\ &= \frac{1}{\Gamma(-\alpha)} \int_{-\infty}^t \frac{w(r)-w(t)}{(t-r)^{\alpha+1}} \mathrm{d}r = D_m^\alpha w(t), \end{aligned} \tag{2.16}$$

where, in the last step, we integrated by parts. The expression $D_m^\alpha w(t)$ is known as the Marchaud derivative of order α of the function w. This is the way that the Caputo derivative has been understood, for instance, in Refs. 6, 5, 7 and 4. We comment, in passing, that owing to Ref. 10 this fractional derivative satisfies an extension problem similar to the (by now) classical Caffarelli–Silvestre extension in Refs. 13 and 58 for the fractional Laplacian.

Another approach, and the one we shall adopt here, is to notice that (1.1) can be converted, for sufficiently smooth functions, into a Volterra type equation

$$w(t) = w(0) + \frac{1}{\Gamma(\alpha)} \int_0^t (t-s)^{\alpha-1} D_c^\alpha w(s) \mathrm{d}s, \quad \forall\, t \in [0,T]. \tag{2.17}$$

This identity is the beginning of the theory developed in Ref. 36 to extend the notion of Caputo derivative. To be more specific, Ref. 36 considers the set of distributions

$$\mathscr{E}^T = \{w \in \mathscr{D}'(\mathbb{R}; \mathcal{H}) : \exists M_w \in (-\infty, T), \operatorname{supp}(w) \subset [-M_w, T)\}.$$

for a fixed time $T > 0$. Then the modified Riemann–Liouville derivative for any distribution $w \in \mathscr{E}^T$ is defined, following classical references, like Sec. 1.5.5 of Ref. 26, as

$$D_{rl}^\alpha w = w * g_{-\alpha} \in \mathscr{E}^T,$$

where $g_{-\alpha}(t) = \frac{1}{\Gamma(1-\alpha)} D(\theta(t) t^{-\alpha})$, with θ being the Heaviside function, is a distribution supported in $[0, \infty)$ and the convolution is understood as the generalized definition between distributions. Here D denotes the distributional derivative. Reference 36 then uses this to define the generalized Caputo derivative of $w \in L^1_{\text{loc}}([0,T);\mathcal{H})$ associated with w_0 by

$$D_c^\alpha w = D_{rl}^\alpha (w - w_0).$$

If there exists $w(0) \in \mathcal{H}$ such that $\lim_{t\downarrow 0} \fint_0^t \|w(s) - w(0)\| \mathrm{d}s = 0$, then we always impose $w_0 = w(0)$ in this definition. It is shown in Theorem 3.7 of Ref. 36 that for such a function w, (2.17) holds for Lebesgue a.e. $t \in (0,T)$ provided that the generalized Caputo derivative $D_c^\alpha w \in L^1_{\text{loc}}([0,T);\mathcal{H})$.

We also comment that Proposition 3.11(ii) of Ref. [36] implies that for every function $w \in L^2(0,T;\mathcal{H})$ with $D_c^\alpha w \in L^2(0,T;\mathcal{H})$ we have

$$\frac{1}{2} D_c^\alpha \|w\|^2(t) \le \langle D_c^\alpha w(t), w(t) \rangle. \tag{2.18}$$

Finally, we recall that the Mittag–Leffler function of order $\alpha \in (0,1)$ is defined via

$$E_\alpha(z) = \sum_{k=0}^\infty \frac{z^k}{\Gamma(\alpha k + 1)}.$$

We refer the reader to Ref. 28 for an extensive treatise on this function. Here we just mention that this function satisfies, for any $\lambda \in \mathbb{R}$, the identity

$$D_c^\alpha E_\alpha(\lambda t^\alpha) = \lambda E_\alpha(\lambda t^\alpha), \quad E_\alpha(0) = 1. \tag{2.19}$$

2.3.1. *An auxiliary estimate*

Having defined the Caputo derivative of a function, we present an auxiliary result. Namely, an estimate on functions that have piecewise constant, over some partition $\mathcal{P}$, Caputo derivative.

Lemma 2.6. (Continuity) *Let $p \in [1,\infty)$; $\mathcal{P}$ be a partition, as in (2.2), of $[0,T]$; and $w \in L^1(0,T;\mathcal{H})$ be such that its generalized Caputo derivative $D_c^\alpha w \in L^p_\alpha(0,T;\mathcal{H})$, and it is piecewise constant over $\mathcal{P}$. Then we have*

$$\sup_{r\in[0,T]} \int_0^r (r-t)^{\alpha-1} \|w(\lceil t \rceil_{\mathcal{P}}) - w(t)\|^p \mathrm{d}t \le C\tau^{p\alpha} \|D_c^\alpha w\|^p_{L^p_\alpha(0,T;\mathcal{H})}, \tag{2.20}$$

where the constant C depends only on α.

Proof. The representation (2.17) allows us to write

$$\begin{aligned} &w(\lceil t \rceil_{\mathcal{P}}) - w(t) \\ &\quad = \frac{1}{\Gamma(\alpha)} \left[\int_0^t D_c^\alpha w(s) \left((\lceil t \rceil_{\mathcal{P}} - s)^{\alpha-1} - (t-s)^{\alpha-1} \right) \mathrm{d}s \right. \\ &\qquad \left. + \int_t^{\lceil t \rceil_{\mathcal{P}}} D_c^\alpha w(s) (\lceil t \rceil_{\mathcal{P}} - s)^{\alpha-1} \mathrm{d}s \right]. \end{aligned}$$

Therefore by Hölder inequality, we have

$$\begin{aligned}
\|w(\lceil t\rceil_{\mathcal{P}})-w(t)\|^p &\leq \frac{1}{\Gamma^p(\alpha)}\left(\int_0^t |(\lceil t\rceil_{\mathcal{P}}-s)^{\alpha-1}-(t-s)^{\alpha-1}|\mathrm{d}s\right.\\
&\qquad\left.+\int_t^{\lceil t\rceil_{\mathcal{P}}}(\lceil t\rceil_{\mathcal{P}}-s)^{\alpha-1}\mathrm{d}s\right)^{p-1}\\
&\qquad\times\left(\int_0^t \|D_c^\alpha w(s)\|^p\left|(\lceil t\rceil_{\mathcal{P}}-s)^{\alpha-1}-(t-s)^{\alpha-1}\right|\mathrm{d}s\right.\\
&\qquad\left.+\int_t^{\lceil t\rceil_{\mathcal{P}}}\|D_c^\alpha w(s)\|^p(\lceil t\rceil_{\mathcal{P}}-s)^{\alpha-1}\mathrm{d}s\right)\\
&\leq C\tau^{\alpha(p-1)}\left(\int_0^t \|D_c^\alpha w(s)\|^p\left|(\lceil t\rceil_{\mathcal{P}}-s)^{\alpha-1}-(t-s)^{\alpha-1}\right|\mathrm{d}s\right.\\
&\qquad\left.+\frac{(\lceil t\rceil_{\mathcal{P}}-t)^\alpha}{\alpha}\|D_c^\alpha w(t)\|^p\right)\\
&= C_1\tau^{\alpha(p-1)}\int_0^t \|D_c^\alpha w(s)\|^p\left|(\lceil t\rceil_{\mathcal{P}}-s)^{\alpha-1}-(t-s)^{\alpha-1}\right|\mathrm{d}s\\
&\qquad+C_2\tau^{p\alpha}\|D_c^\alpha w(t)\|^p=\mathrm{I}_1(t)+\mathrm{I}_2(t),
\end{aligned}$$

where the constants C, C_1, and C_2 depend only on p and α.

For $\mathrm{I}_2(t)$, we simply have

$$\int_0^r (r-t)^{\alpha-1}\mathrm{I}_2(t)\mathrm{d}t \leq C\tau^{p\alpha}\|D_c^\alpha w\|^p_{L^p_\alpha(0,T;\mathcal{H})}.$$

Now to bound the integral for $\mathrm{I}_1(t)$, we use Fubini's theorem to get

$$\begin{aligned}
\int_0^r (r-t)^{\alpha-1}\mathrm{I}_1(t)\mathrm{d}t &= C_1\tau^{(p-1)\alpha}\int_0^r \|D_c^\alpha w(s)\|^p\\
&\qquad\times\int_s^r (r-t)^{\alpha-1}|(\lceil t\rceil_{\mathcal{P}}-s)^{\alpha-1}-(t-s)^{\alpha-1}|\mathrm{d}t\mathrm{d}s.
\end{aligned}$$

We claim that

$$\int_s^r (r-t)^{\alpha-1}|(\lceil t\rceil_{\mathcal{P}}-s)^{\alpha-1}-(t-s)^{\alpha-1}|\mathrm{d}t \leq C_3(r-s)^{\alpha-1}\tau^\alpha, \qquad (2.21)$$

where C_3 only depends α. If this is true, then we have

$$\begin{aligned}
\int_0^r (r-t)^{\alpha-1}\mathrm{I}_1(t)\mathrm{d}t &\leq C\tau^{p\alpha}\int_0^r \|D_c^\alpha w(s)\|^p(r-s)^{\alpha-1}\mathrm{d}s\\
&\leq C\tau^{p\alpha}\|D_c^\alpha w\|^p_{L^p_\alpha(0,T;\mathcal{H})}.
\end{aligned}$$

The proof of (2.21) proceeds as the one for (2.15). For brevity we shall skip the details. ∎

2.4. *Some comparison estimates*

As a final preparatory step we present some auxiliary results that shall be repeatedly used and are related to differential inequalities involving the Caputo derivative, and a Grönwall-like lemma.

First, we present a comparison principle which is similar to Proposition 4.2 of Ref. 24. The proof can be done easily by contradiction, and therefore it is omitted here.

Lemma 2.7. (Comparison) *Let $g_1, g_2 : [0,T] \times \mathbb{R} \to \mathbb{R}$ be both non-decreasing in their second argument and g_2 be measurable. Assume that $v, w \in C([0,T];\mathbb{R})$ satisfy $v(0) < w(0)$, and there is some $\alpha \in (0,1)$, for which*

$$v(t) \leq g_1(t, v(t)) + \frac{1}{\Gamma(\alpha)} \int_0^t (t-s)^{\alpha-1} g_2(s, v(s)) \mathrm{d}s,$$

$$w(t) > g_1(t, w(t)) + \frac{1}{\Gamma(\alpha)} \int_0^t (t-s)^{\alpha-1} g_2(s, w(s)) \mathrm{d}s,$$

for every $t \in [0,T]$. Then we have $v < w$ on $[0,T]$.

We now present a result that can be interpreted as an extension of Lemma 3.7 in Ref. 51 to the fractional case. However, unlike the classical case, here we have the restriction that $\lambda \geq 0$ because we have to argue from a fractional integral inequality. Nevertheless, this is sufficient for our purposes.

Lemma 2.8. (Fractional Grönwall) *Let $a \in C([0,T];\mathbb{R})$ with $D_c^\alpha a^2 \in L^1_{loc}([0,T);\mathbb{R})$, $b, c, d : [0,T] \to [0,+\infty]$ be measurable functions, and $\lambda \geq 0$. If the following differential inequality is satisfied*

$$D_c^\alpha a^2(t) + b(t) \leq 2\lambda a^2(t) + c(t) + 2d(t)a(t), \quad a.e.\ t \in (0,T), \tag{2.22}$$

then we have

$$\left(\sup_{t\in[0,T]} a^2(t) + \frac{1}{\Gamma(\alpha)} \|b\|_{L^1_\alpha(0,T;\mathbb{R})} \right)^{1/2} \leq 2\widetilde{D}(T) E_\alpha(2\lambda T^\alpha) + \sqrt{a^2(0) + \widetilde{C}(T)} \sqrt{E_\alpha(2\lambda T^\alpha)},$$

where

$$\widetilde{C}(t) = \frac{1}{\Gamma(\alpha)} \|c\|_{L^1_\alpha(0,t;\mathbb{R})}, \quad \widetilde{D}(t) = \frac{1}{\Gamma(\alpha)} \|d\|_{L^1_\alpha(0,t;\mathbb{R})}. \tag{2.23}$$

Proof. From (2.22) we obtain that

$$a^2(t) + \frac{1}{\Gamma(\alpha)} \int_0^t (t-s)^{\alpha-1} b(s) \mathrm{d}s \leq a^2(0) + \frac{1}{\Gamma(\alpha)} \int_0^t (t-s)^{\alpha-1} \left[c(s) + 2d(s)a(s) + 2\lambda a^2(s) \right] \mathrm{d}s$$

$$\leq a^2(0) + \widetilde{C}(t) + 2\widetilde{a}(t)\widetilde{D}(t) + \frac{2\lambda}{\Gamma(\alpha)} \times \int_0^t (t-s)^{\alpha-1}\widetilde{a}^2(s)\, \mathrm{d}s, \tag{2.24}$$

where $\widetilde{a}(t) = \max_{0\leq s\leq t} a(s)$ and the functions $\widetilde{C}, \widetilde{D}$ are defined in (2.23). This immediately implies that

$$\widetilde{a}^2(t) \leq a^2(0) + \widetilde{C}(t) + 2\widetilde{a}(t)\widetilde{D}(t) + \frac{2\lambda}{\Gamma(\alpha)}\int_0^t (t-s)^{\alpha-1}\widetilde{a}^2(s)\mathrm{d}s.$$

In order to bound $\widetilde{a}$, we construct a barrier function $e(t) = K\sqrt{E_\alpha(2\lambda t^\alpha)}$ where the constant K is chosen so that

$$e^2(t) > a^2(0) + \widetilde{C}(t) + 2e(t)\widetilde{D}(t) + \frac{2\lambda}{\Gamma(\alpha)}\int_0^t (t-s)^{\alpha-1}e^2(s)\mathrm{d}s, \quad \forall\, t \in (0,T).$$

Indeed, owing to (2.19) we see that

$$\frac{2\lambda}{\Gamma(\alpha)}\int_0^t (t-s)^{\alpha-1}E_\alpha(2\lambda s^\alpha)\, \mathrm{d}s = E_\alpha(2\lambda t^\alpha) - E_\alpha(0) = E_\alpha(2\lambda t^\alpha) - 1$$

and hence

$$\begin{aligned}
&a^2(0) + \widetilde{C}(t) + e(t)\widetilde{D}(t) + \frac{2\lambda}{\Gamma(\alpha)}\int_0^t (t-s)^{\alpha-1}e^2(s)\, \mathrm{d}s \\
&\quad = a^2(0) + \widetilde{C}(t) + 2K\sqrt{E_\alpha(2\lambda t^\alpha)}\widetilde{D}(t) \\
&\qquad + K^2\left(E_\alpha(2\lambda t^\alpha) - 1\right) < K^2 E_\alpha(2\lambda t^\alpha) = e^2(t),
\end{aligned}$$

for every $t \in (0,T)$ provided that

$$K > \widetilde{D}(T)\sqrt{E_\alpha(2\lambda T^\alpha)} + \sqrt{a^2(0) + \widetilde{C}(T) + \widetilde{D}^2(t)E_\alpha(2\lambda T^\alpha)}. \tag{2.25}$$

Applying Theorem 2.7 we obtain that

$$\widetilde{a}(t) \leq e(t) = K\sqrt{E_\alpha(2\lambda t^\alpha)}.$$

Plugging this back into (2.24) and noticing that this holds for any K satisfying (2.25) we obtain that

$$\begin{aligned}
&\sup_{t\in[0,T]} a^2(t) + \frac{1}{\Gamma(\alpha)}\int_0^t (t-s)^{\alpha-1}b(s)\mathrm{d}s \\
&\quad \leq \left(\widetilde{D}(T)\sqrt{E_\alpha(2\lambda T^\alpha)} + \sqrt{a^2(0) + \widetilde{C}(T) + \widetilde{D}^2(t)E_\alpha(2\lambda T^\alpha)}\right)^2 E_\alpha(2\lambda T^\alpha) \\
&\quad \leq \left(2\widetilde{D}(T)E_\alpha(2\lambda T^\alpha) + \sqrt{a^2(0) + \widetilde{C}(T)}\sqrt{E_\alpha(2\lambda T^\alpha)}\right)^2
\end{aligned}$$

which is the desired result. $\square$

3. Deconvolutional Discretization of the Caputo Derivative

To discretize the Caputo fractional derivative, Refs. 37 and 39 consider a so-called deconvolutional scheme on uniform time grids and prove some properties of this discretization. In this section, we generalize this deconvolutional scheme to the variable time step setting, and prove properties that will be useful in deriving *a posteriori* error estimates later, in Sec. 5.2.

3.1. *The discrete Caputo derivative*

Let $\mathcal{P}$ be a partition as in (2.2). To motivate this discretization, let us assume that $w : [0, T] \to \mathcal{H}$ is such that $D_c^\alpha w(t)$ is piecewise constant on the partition $\mathcal{P}$, with

$$D_c^\alpha w(t) = V_{n(t)}.$$

Then formally by (2.17), we have, for $n \in \{1, \ldots, N\}$,

$$\begin{aligned} w(t_n) &= w(0) + \frac{1}{\Gamma(\alpha)} \int_0^{t_n} (t_n - s)^{\alpha-1} D_c^\alpha w(s) \mathrm{d}s \\ &= w(0) + \frac{1}{\Gamma(\alpha+1)} \sum_{i=1}^{n} \left((t_n - t_{i-1})^\alpha - (t_n - t_i)^\alpha\right) V_i, . \end{aligned} \tag{3.1}$$

Let $\mathbf{K}_\mathcal{P} \in \mathbb{R}^{N\times N}$ be the matrix induced by the partition $\mathcal{P}$, which is defined as

$$\mathbf{K}_{\mathcal{P},ni} = \begin{cases} \dfrac{1}{\Gamma(\alpha+1)}((t_n - t_{i-1})^\alpha - (t_n - t_i)^\alpha), & 1 \le i \le n \le N, \\ 0, & 1 \le n < i \le N. \end{cases} \tag{3.2}$$

Then we can rewrite (3.1) in matrix form as

$$\mathbf{W} = \mathbf{W}_0 + \mathbf{K}_\mathcal{P}\mathbf{V},$$

where $\mathbf{V}, \mathbf{W}, \mathbf{W}_0 \in \mathcal{H}^N$ with $\mathbf{V}_n = V_n$, $\mathbf{W}_n = w(t_n)$, and $(\mathbf{W}_0)_n = w(0)$. Notice that $\mathbf{K}_\mathcal{P}$ is lower triangular and all the elements on and below the main diagonal are positive. Therefore $\mathbf{K}_\mathcal{P}$ is invertible and its inverse is also lower triangular. Thus, the previous identity is equivalent to

$$\mathbf{V} = \mathbf{K}_\mathcal{P}^{-1}(\mathbf{W} - \mathbf{W}_0),$$

in other words

$$V_n = \sum_{i=1}^{n} \mathbf{K}_{\mathcal{P},ni}^{-1}(W_i - W_0) = \mathbf{K}_{\mathcal{P},n0}^{-1} W_0 + \sum_{i=1}^{n} \mathbf{K}_{\mathcal{P},ni}^{-1} W_i,$$

where we set $\mathbf{K}_{\mathcal{P},n0}^{-1} = -\sum_{j=1}^{n} \mathbf{K}_{\mathcal{P},nj}^{-1}$. This motivates the following approximation of the Caputo derivative provided $\mathbf{W} \in \mathcal{H}^N$ and $W_0 \in \mathcal{H}$ are given. For $n \in$

$\{1, \ldots, N\}$ we set

$$(D_{\mathcal{P}}^{\alpha}\mathbf{W})_n = \sum_{i=1}^{n} \mathbf{K}_{\mathcal{P},ni}^{-1}(W_i - W_0) = \sum_{i=0}^{n} \mathbf{K}_{\mathcal{P},ni}^{-1} W_i$$
$$= \sum_{i=0}^{n-1} \mathbf{K}_{\mathcal{P},ni}^{-1}(W_i - W_n). \tag{3.3}$$

3.2. *Properties of* $\mathbf{K}_{\mathcal{P}}^{-1}$

We note that, when the partition is uniform, both $\mathbf{K}_{\mathcal{P}}$ and its inverse will be Toeplitz matrices, and hence the product $\mathbf{K}_{\mathcal{P}}\mathbf{V}$ can be interpreted as the convolution of sequences. Consequently, multiplication by $\mathbf{K}_{\mathcal{P}}^{-1}$ is equivalent to taking a sequence deconvolution. This motivates the name of this scheme and enables Ref. 39 to apply techniques for the deconvolution of a completely monotone sequence and prove properties of $\mathbf{K}_{\mathcal{P}}^{-1}$.

We were not successful in extending, to a general partition $\mathcal{P}$, all the properties of $\mathbf{K}_{\mathcal{P}}^{-1}$ presented in Ref. 39 for the case when the partition is uniform. This is mainly because their techniques are based on ideas that rely on completely monotone sequences, which do not easily extend to a general $\mathcal{P}$. Nevertheless we have obtained sufficient, for our purposes, properties. The following result is the counterpart to Proposition 3.2(1) in Ref. 39.

Proposition 3.1. (Properties of $\mathbf{K}_{\mathcal{P}}^{-1}$) *Let $\mathcal{P}$ be a partition as in* (2.2), *and $\mathbf{K}_{\mathcal{P}}$ be defined in* (3.2). *The matrix $\mathbf{K}_{\mathcal{P}}$ is invertible, and its inverse satisfies*:

$$\mathbf{K}_{\mathcal{P},n0}^{-1} = -\sum_{j=1}^{n} \mathbf{K}_{\mathcal{P},nj}^{-1} < 0, \quad n \in \{1, \ldots, N\}, \tag{3.4}$$

$$\mathbf{K}_{\mathcal{P},ii}^{-1} > 0 \quad i \in \{1, \ldots, N\}, \quad \mathbf{K}_{\mathcal{P},ni}^{-1} < 0 \quad 1 \le i < n \le N. \tag{3.5}$$

Proof. We already showed that $\mathbf{K}_{\mathcal{P}}$ is nonsingular. We prove (3.4) and (3.5) separately.

First, to prove that $\mathbf{K}_{\mathcal{P},n0}^{-1} < 0$. For this, it suffices to show that for a vector $\mathbf{W} \in \mathbb{R}^N$ such that $W_i = 1$ for any $i \geq 1$, then the vector $\mathbf{F} = \mathbf{K}_{\mathcal{P}}^{-1}\mathbf{W}$ satisfies

$$F_n > 0 \quad \forall\, n \geq 1.$$

We prove this by induction on n. For $n = 1$, clearly

$$F_1 = \frac{W_1}{\mathbf{K}_{\mathcal{P},1,1}} = \frac{1}{\mathbf{K}_{\mathcal{P},1,1}} > 0.$$

Suppose that $F_j > 0$ for all $1 \leq j \leq k$, now we want to show that $F_{k+1} > 0$ as well. Notice that

$$1 = W_k = \sum_{j=1}^{k} \mathbf{K}_{\mathcal{P},k,j} F_j, \quad 1 = W_{k+1} = \sum_{j=1}^{k+1} \mathbf{K}_{\mathcal{P},k+1,j} F_j,$$

then taking the difference we have

$$\begin{aligned} 0 &= \sum_{j=1}^{k+1} \mathbf{K}_{\mathcal{P},k+1,j} F_j - \sum_{j=1}^{k} \mathbf{K}_{\mathcal{P},k,j} F_j \\ &= \mathbf{K}_{\mathcal{P},k+1,k+1} F_{k+1} + \sum_{j=1}^{k} (\mathbf{K}_{\mathcal{P},k+1,j} - \mathbf{K}_{\mathcal{P},k,j}) F_j. \end{aligned} \tag{3.6}$$

We claim that $\mathbf{K}_{\mathcal{P},k+1,j} - \mathbf{K}_{\mathcal{P},k,j} < 0$ for any j. In fact, this can be seen through the definition of the entries of $\mathbf{K}_{\mathcal{P}}$

$$\begin{aligned} &\mathbf{K}_{\mathcal{P},k+1,j} - \mathbf{K}_{\mathcal{P},k,j} < 0 \\ &\quad \Leftrightarrow (t_{k+1} - t_{j-1})^\alpha - (t_{k+1} - t_j)^\alpha < (t_k - t_{j-1})^\alpha - (t_k - t_j)^\alpha \\ &\quad \Leftrightarrow \int_0^{t_j - t_{j-1}} (t_{k+1} - t_j + s)^{\alpha-1} \mathrm{d}s < \int_0^{t_j - t_{j-1}} (t_k - t_j + s)^{\alpha-1} \mathrm{d}s. \end{aligned}$$

Using $\mathbf{K}_{\mathcal{P},k+1,j} - \mathbf{K}_{\mathcal{P},k,j} < 0$ and $F_j > 0$ for all $j \in \{1, \dots, k\}$ in (3.6), we see that $\mathbf{K}_{\mathcal{P},k+1,k+1} F_{k+1} > 0$ and thus $F_{k+1} > 0$. Therefore by induction we proved that $\mathbf{K}^{-1}_{\mathcal{P},n0} < 0$ for $n \geq 1$.

Next, we prove that $\mathbf{K}^{-1}_{\mathcal{P},ii} > 0$ and $\mathbf{K}^{-1}_{\mathcal{P},ni} < 0$. Consider a vector $\mathbf{W} \in \mathbb{R}^N$ that is such that $W_i = 1$ and $W_j = 0$ for $j \neq i$. It suffices to prove that for, $\mathbf{F} = \mathbf{K}_{\mathcal{P}}^{-1} \mathbf{W}$, we have $F_i > 0$ and if $n > i$

$$F_n < 0. \tag{3.7}$$

Since $\mathbf{K}_{\mathcal{P}}^{-1}$ is lower triangular, we know $F_j = 0$ for $j \in \{1, \dots, i-1\}$. From $\mathbf{K}_{\mathcal{P}} \mathbf{F} = \mathbf{W}$, we see that

$$1 = W_i = (\mathbf{K}_{\mathcal{P}} \mathbf{F})_i = \sum_{j=1}^{i} \mathbf{K}_{\mathcal{P},ij} F_j = \mathbf{K}^{-1}_{\mathcal{P},ii} F_i$$

and thus $F_i = 1/\mathbf{K}_{\mathcal{P},ii} > 0$. Now we prove by induction that (3.7) holds. First, when $n = i+1$, we have

$$0 = W_{i+1} = (\mathbf{K}_{\mathcal{P}} \mathbf{F})_{i+1} = \mathbf{K}_{\mathcal{P},i+1,i} F_i + \mathbf{K}_{\mathcal{P},i+1,i+1} F_{i+1}$$

and hence

$$F_{i+1} = -\frac{\mathbf{K}_{\mathcal{P},i+1,i} F_i}{\mathbf{K}_{\mathcal{P},i+1,i+1}} < 0.$$

This shows that (3.7) is true for $n = i+1$. Now suppose that we have already shown that $F_n < 0$ for n satisfying $n \in \{i+1, \dots, k\}$, we want to prove $F_{k+1} < 0$. To this aim, notice that

$$0 = W_{k+1} = (\mathbf{K}_{\mathcal{P}} \mathbf{F})_{k+1} = \sum_{j=i}^{k} \mathbf{K}_{\mathcal{P},k+1,j} F_j + \mathbf{K}_{\mathcal{P},k+1,k+1} F_{k+1},$$

therefore we only need to show $\sum_{j=i}^{k} \mathbf{K}_{\mathcal{P},k+1,j} F_j > 0$. Recall that

$$0 = W_k = (\mathbf{K}_{\mathcal{P}} \mathbf{F})_k = \sum_{j=i}^{k} \mathbf{K}_{\mathcal{P},k,j} F_j,$$

and thus, since $\mathbf{K}_{\mathcal{P},k,i} > 0$, we can get

$$\begin{aligned}
\sum_{j=i}^{k} \mathbf{K}_{\mathcal{P},k+1,j} F_j &= \sum_{j=i}^{k} \mathbf{K}_{\mathcal{P},k+1,j} F_j - \frac{\mathbf{K}_{\mathcal{P},k+1,i}}{\mathbf{K}_{\mathcal{P},k,i}} \sum_{j=i}^{k} \mathbf{K}_{\mathcal{P},k,j} F_j \\
&= \sum_{j=i+1}^{k} \left(\mathbf{K}_{\mathcal{P},k+1,j} - \frac{\mathbf{K}_{\mathcal{P},k+1,i}}{\mathbf{K}_{\mathcal{P},k,i}} \mathbf{K}_{\mathcal{P},k,j} \right) F_j.
\end{aligned}$$

Since by the induction hypothesis $F_j < 0$ for $j \in \{i+1, \ldots, k\}$, it only remains to show that

$$\mathbf{K}_{\mathcal{P},k+1,j} - \frac{\mathbf{K}_{\mathcal{P},k+1,i}}{\mathbf{K}_{\mathcal{P},k,i}} \mathbf{K}_{\mathcal{P},k,j} < 0 \Leftrightarrow \frac{\mathbf{K}_{\mathcal{P},k+1,i}}{\mathbf{K}_{\mathcal{P},k,i}} > \frac{\mathbf{K}_{\mathcal{P},k+1,j}}{\mathbf{K}_{\mathcal{P},k,j}}.$$

Applying Cauchy's mean value theorem, there exists $\eta \in (t_k - t_i, t_k - t_{i-1})$ such that

$$\begin{aligned}
\frac{\mathbf{K}_{\mathcal{P},k+1,i}}{\mathbf{K}_{\mathcal{P},k,i}} &= \frac{(t_{k+1} - t_{i-1})^\alpha - (t_{k+1} - t_i)^\alpha}{(t_k - t_{i-1})^\alpha - (t_k - t_i)^\alpha} \\
&= \frac{\alpha(\eta + \tau_{k+1})^{\alpha-1}}{\alpha \eta^{\alpha-1}} = \left(\frac{\eta + \tau_{k+1}}{\eta} \right)^{\alpha-1}.
\end{aligned}$$

Similarly there exists $\xi \in (t_k - t_j, t_k - t_{j-1})$ such that

$$\frac{\mathbf{K}_{\mathcal{P},k+1,j}}{\mathbf{K}_{\mathcal{P},k,j}} = \left(\frac{\xi + \tau_{k+1}}{\xi} \right)^{\alpha-1}.$$

Due to $j > i$, we have $\xi < \eta$ and hence

$$\frac{\mathbf{K}_{\mathcal{P},k+1,j}}{\mathbf{K}_{\mathcal{P},k,j}} = \left(\frac{\xi + \tau_{k+1}}{\xi} \right)^{\alpha-1} < \left(\frac{\eta + \tau_{k+1}}{\eta} \right)^{\alpha-1} = \frac{\mathbf{K}_{\mathcal{P},k+1,i}}{\mathbf{K}_{\mathcal{P},k,i}}.$$

Therefore from the arguments above we see that $F_{k+1} < 0$, and by induction $\mathbf{K}_{\mathcal{P},ni}^{-1} < 0$ for $n > i$. □

Remark 3.2. (Generalization) The discretization of the Caputo derivative, described in (3.3), and its properties presented in Proposition 3.1 can be extended to more general kernels. Indeed, for a general convolutional kernel $g \in L^1(0, T; \mathbb{R})$ the entries of the matrix $\mathbf{K}_{\mathcal{P}}$ will be

$$\mathbf{K}_{\mathcal{P},ni} = \int_{t_n - t_{i-1}}^{t_n - t_i} g(t) \mathrm{d}t.$$

The proof of (3.4) follows *verbatim* provided $g'(t) < 0$, as the reader can readily verify. The proof of (3.5) only requires that the function $G(t) = \ln(g(t))$, satisfies $G''(t) > 0$.

For a uniform time grid $\mathcal{P}$, Theorem 2.3 of Ref. 37 proves that, for every i, the sequence $\{-\mathbf{K}^{-1}_{\mathcal{P},n+i,i}\}_{n\geq 1}$ is completely monotone. The following result holds for a general partition $\mathcal{P}$, and is a direct consequence of Theorem 2.3 in Ref. 37 for uniform time stepping.

Proposition 3.3. (Monotonicity) *Let $\mathcal{P}$ be a partition of* $[0,T]$ *as in* (2.2), *and* $\mathbf{K}_{\mathcal{P}}$ *be defined as in* (3.2). *Then, its inverse satisfies*

(1) *For* $n \in \{1, \dots, N-1\}$,

$$-\sum_{j=1}^{n} \mathbf{K}^{-1}_{\mathcal{P},nj} = \mathbf{K}^{-1}_{\mathcal{P},n0} < \mathbf{K}^{-1}_{\mathcal{P},n+1,0} = -\sum_{j=1}^{n+1} \mathbf{K}^{-1}_{\mathcal{P},n+1,j}. \tag{3.8}$$

(2) *For* $1 \leq i < n < N$,

$$\mathbf{K}^{-1}_{\mathcal{P},ni} < \mathbf{K}^{-1}_{\mathcal{P},n+1,i}. \tag{3.9}$$

Proof. To prove (3.8) it suffices to show that for a vector $\mathbf{W} \in \mathbb{R}^N$ such that $W_i = 1$ for any $i \geq 1$, then the vector $\mathbf{F} = \mathbf{K}_{\mathcal{P}}^{-1}\mathbf{W}$ satisfies

$$F_n > F_{n+1} \quad \forall n \geq 1.$$

We prove this by induction on n. For $n = 1$,

$$\begin{aligned}
1 &= W_1 = (\mathbf{K}_{\mathcal{P}}\mathbf{F})_1 = \mathbf{K}_{\mathcal{P},11}F_1, \\
1 &= W_2 = (\mathbf{K}_{\mathcal{P}}\mathbf{F})_2 = \mathbf{K}_{\mathcal{P},21}F_1 \\
&\quad + \mathbf{K}_{\mathcal{P},22}F_2 = (\mathbf{K}_{\mathcal{P},21} + \mathbf{K}_{\mathcal{P},22})F_1 + \mathbf{K}_{\mathcal{P},22}(F_2 - F_1).
\end{aligned}$$

Clearly,

$$F_1 > 0, \quad \mathbf{K}_{\mathcal{P},11} = (t_1 - t_0)^\alpha < (t_2 - t_0)^\alpha = \mathbf{K}_{\mathcal{P},21} + \mathbf{K}_{\mathcal{P},22}.$$

Hence we have

$$\mathbf{K}_{\mathcal{P},22}(F_2 - F_1) = 1 - (\mathbf{K}_{\mathcal{P},21} + \mathbf{K}_{\mathcal{P},22})F_1 < 1 - \mathbf{K}_{\mathcal{P},11}F_1 = 0,$$

which, since $\mathbf{K}_{\mathcal{P},22} > 0$, implies that $F_2 - F_1 < 0$, i.e. $F_1 > F_2$. So the claim holds for $n = 1$.

Suppose $F_{j+1} < F_j$ for all $1 \leq j < k$, now we want to show that $F_{k+1} < F_k$ as well. Notice that

$$\begin{aligned}
1 = W_k &= \sum_{i=1}^{k} \mathbf{K}_{\mathcal{P},ki}F_i = \sum_{i=0}^{k-1} \left(\sum_{j=i+1}^{k} \mathbf{K}_{\mathcal{P},kj} \right) (F_{i+1} - F_i) \\
&= \sum_{i=0}^{k-1} (t_k - t_i)^\alpha (F_{i+1} - F_i),
\end{aligned}$$

$$1 = W_{k+1} = \sum_{i=1}^{k+1} \mathbf{K}_{\mathcal{P},k+1,i}F_i = \sum_{i=0}^{k} (t_{k+1} - t_i)^\alpha (F_{i+1} - F_i),$$

where we set $F_0 = 0$ in the equations above. Therefore to show $F_{k+1} < F_k$, we only need to prove that

$$\begin{aligned}
0 &< \sum_{i=0}^{k-1} (t_{k+1} - t_i)^\alpha (F_{i+1} - F_i) - 1 \\
&= \sum_{i=0}^{k-1} (t_{k+1} - t_i)^\alpha (F_{i+1} - F_i) - \sum_{i=0}^{k-1} (t_k - t_i)^\alpha (F_{i+1} - F_i) \\
&= \sum_{i=0}^{k-1} ((t_{k+1} - t_i)^\alpha - (t_k - t_i)^\alpha)(F_{i+1} - F_i). \qquad (3.10)
\end{aligned}$$

Since we also have

$$\begin{aligned}
1 = W_{k-1} &= \sum_{i=1}^{k-1} \mathbf{K}_{\mathcal{P},k-1,i} F_i = \sum_{i=0}^{k-2} (t_{k-1} - t_i)^\alpha (F_{i+1} - F_i) \\
&= \sum_{i=0}^{k-1} (t_{k-1} - t_i)^\alpha (F_{i+1} - F_i),
\end{aligned}$$

Taking the difference between the equation above and the one for W_k, we obtain that

$$\begin{aligned}
0 = W_k - W_{k-1} &= \sum_{i=0}^{k-1} (t_k - t_i)^\alpha (F_{i+1} - F_i) - \sum_{i=0}^{k-1} (t_{k-1} - t_i)^\alpha (F_{i+1} - F_i) \\
&= \sum_{i=0}^{k-1} ((t_k - t_i)^\alpha - (t_{k-1} - t_i)^\alpha)(F_{i+1} - F_i).
\end{aligned}$$

In light of this identity, we claim that to obtain (3.10) it suffices to show that, for $i \in \{1, \dots, k-1\}$,

$$\frac{t_{k+1}^\alpha - t_k^\alpha}{t_k^\alpha - t_{k-1}^\alpha} = \frac{(t_{k+1} - t_0)^\alpha - (t_k - t_0)^\alpha}{(t_k - t_0)^\alpha - (t_{k-1} - t_0)^\alpha} > \frac{(t_{k+1} - t_i)^\alpha - (t_k - t_i)^\alpha}{(t_k - t_i)^\alpha - (t_{k-1} - t_i)^\alpha}. \qquad (3.11)$$

If this is true, letting $c = \left(t_{k+1}^\alpha - t_k^\alpha\right) / \left(t_k^\alpha - t_{k-1}^\alpha\right)$ we have

$$\begin{aligned}
&\sum_{i=0}^{k-1} ((t_{k+1} - t_i)^\alpha - (t_k - t_i)^\alpha)(F_{i+1} - F_i) \\
&\quad = \sum_{i=0}^{k-1} (((t_{k+1} - t_i)^\alpha - (t_k - t_i)^\alpha) - c((t_k - t_i)^\alpha - (t_{k-1} - t_i)^\alpha))(F_{i+1} - F_i) \\
&\quad = \sum_{i=1}^{k-1} (((t_{k+1} - t_i)^\alpha - (t_k - t_i)^\alpha) - c((t_k - t_i)^\alpha - (t_{k-1} - t_i)^\alpha))(F_{i+1} - F_i) \\
&\quad = \sum_{i=1}^{k-1} d_i \, (F_{i+1} - F_i),
\end{aligned}$$

where $d_i = ((t_{k+1} - t_i)^\alpha - (t_k - t_i)^\alpha) - c((t_k - t_i)^\alpha - (t_{k-1} - t_i)^\alpha) < 0$ due to (3.11). By the inductive hypothesis, $F_{i+1} - F_i < 0$ for $1 \le i \le k-1$, so the equation above implies (3.10), and hence $F_{k+1} < F_k$ is proved.

To finish the proof, we focus on (3.11), fix i and define $c_1 = t_{k-1} - t_i$, $c_2 = t_k - t_i$, $c_3 = t_{k+1} - t_i$ and function

$$h(x) = \frac{(x+c_3)^\alpha - (x+c_2)^\alpha}{(x+c_2)^\alpha - (x+c_1)^\alpha}.$$

Then (3.11) is equivalent to $h(t_i - t_0) > h(0)$, and it remains to show that $h(x)$ is strictly increasing for $x > 0$. We observe that

$$\frac{\mathrm{d}}{\mathrm{d}x}(\ln(h(x))) = \alpha \left[\frac{(x+c_3)^{\alpha-1} - (x+c_2)^{\alpha-1}}{(x+c_3)^\alpha - (x+c_2)^\alpha} - \frac{(x+c_2)^{\alpha-1} - (x+c_1)^{\alpha-1}}{(x+c_2)^\alpha - (x+c_1)^\alpha} \right].$$

Applying Cauchy's mean-value theorem to the two fractions above, we know there exists $\eta \in (x+c_2, x+c_3)$ and $\xi \in (x+c_1, x+c_2)$ such that

$$\frac{\mathrm{d}}{\mathrm{d}x}(\ln(h(x))) = \alpha \left[\frac{(\alpha-1)\eta^{\alpha-2}}{\alpha\eta^{\alpha-1}} - \frac{(\alpha-1)\xi^{\alpha-2}}{\alpha\xi^{\alpha-1}} \right] = (\alpha-1)\left(\eta^{-1} - \xi^{-1}\right) > 0,$$

where the last inequality holds because $\alpha < 1$ and $\xi < x + c_2 < \eta$. This shows the monotonicity of function h and confirms (3.11). This concludes the inductive step and proves (3.8).

The proof of (3.9) is obtained similarly. For convenience we only write the proof for $i = 1$, but the extension to general i is straightforward. Consider a vector $\mathbf{W} \in \mathbb{R}^N$ such that $W_j = 1$ if $j = 1$ and $W_j = 0$ if $j \neq 1$, then it suffices to prove that vector $\mathbf{F} = \mathbf{K}_{\mathcal{P}}^{-1}\mathbf{W}$ satisfies

$$F_n < F_{n+1} \tag{3.12}$$

for $n \in \{2, \dots, N-1\}$. We prove (3.12) by induction on n. For $n = 2$, observe that

$$W_k = \sum_{j=0}^{k} (t_k - t_j)^\alpha (F_{j+1} - F_j) = \sum_{j=0}^{k-1} (t_k - t_j)^\alpha (F_{j+1} - F_j)$$

from the proof of (3.8) with $F_0 = 0$, we have

$$\begin{aligned}
1 &= W_1 = (t_1 - t_0)^\alpha (F_1 - F_0), \\
0 &= W_2 = (t_2 - t_0)^\alpha (F_1 - F_0) + (t_2 - t_1)^\alpha (F_2 - F_1), \\
0 &= W_3 = (t_3 - t_0)^\alpha (F_1 - F_0) + (t_3 - t_1)^\alpha (F_2 - F_1) + (t_3 - t_2)^\alpha (F_3 - F_2).
\end{aligned}$$

From the first and second equation above, we see that $F_1 > 0$ and $F_2 - F_1 < 0$. Combining the second and the third equation we deduce that

$$0 = W_3 - \frac{t_3^\alpha}{t_2^\alpha} W_2 = \left[(t_3 - t_1)^\alpha - (t_2 - t_1)^\alpha \frac{t_3^\alpha}{t_2^\alpha} \right] (F_2 - F_1) + (t_3 - t_2)^\alpha (F_3 - F_2).$$

Since $(t_3 - t_1)^\alpha - (t_2 - t_1)^\alpha (t_3/t_2)^\alpha = (t_3 - t_1)^\alpha - (t_3 - (t_1 t_3/t_2))^\alpha > 0$, we obtain that $F_3 - F_2 > 0$ which is (3.12) for $n = 2$.

It also remains to prove that when (3.12) holds for $n \in \{2,\dots,k-1\}$, then it also holds for $n=k$, i.e. $F_k < F_{k+1}$, provided that $k<N$. To this aim, we first see that

$$0 = W_{k+1} - \frac{t_{k+1}^\alpha}{t_k^\alpha} W_k = \sum_{j=1}^{k} \left((t_{k+1}-t_j)^\alpha - (t_k - t_j)^\alpha \frac{t_{k+1}^\alpha}{t_k^\alpha} \right) (F_{j+1}-F_j). \quad (3.13)$$

Therefore in order to prove $F_k < F_{k+1}$, we only need to show that

$$\sum_{j=1}^{k-1} \left((t_{k+1}-t_j)^\alpha - (t_k - t_j)^\alpha \frac{t_{k+1}^\alpha}{t_k^\alpha} \right) (F_{j+1}-F_j) < 0. \quad (3.14)$$

Similar to (3.13) we also have

$$0 = W_k - \frac{t_k^\alpha}{t_{k-1}^\alpha} W_{k-1} = \sum_{j=1}^{k-1} \left((t_k - t_j)^\alpha - (t_{k-1} - t_j)^\alpha \frac{t_k^\alpha}{t_{k-1}^\alpha} \right) (F_{j+1}-F_j).$$

Thanks to the inductive hypothesis, we know that $F_{j+1}-F_j < 0$ for $j=2$ and $F_{j+1}-F_j > 0$ for $j \in \{3,\dots,k-1\}$, Therefore using a similar argument used in the proof for (3.8), to prove (3.14) we only need to show that, for $j \in \{2,\dots,k-1\}$,

$$\begin{aligned} &\frac{(t_{k+1}-t_1)^\alpha - (t_k - t_1)^\alpha (t_{k+1}/t_k)^\alpha}{(t_k - t_1)^\alpha - (t_{k-1}-t_1)^\alpha (t_k/t_{k-1})^\alpha} \\ &\qquad > \frac{(t_{k+1}-t_j)^\alpha - (t_k - t_j)^\alpha (t_{k+1}/t_k)^\alpha}{(t_k - t_j)^\alpha - (t_{k-1}-t_j)^\alpha (t_k/t_{k-1})^\alpha}, \end{aligned} \quad (3.15)$$

which is similar to (3.11). We rewrite the inequality above as

$$\begin{aligned} &\frac{(1-t_1/t_{k+1})^\alpha - (1-t_1/t_k)^\alpha}{(1-t_1/t_k)^\alpha - (1-t_1/t_{k-1})^\alpha} \\ &\qquad > \frac{(1-t_j/t_{k+1})^\alpha - (1-t_j/t_k)^\alpha}{(1-t_j/t_k)^\alpha - (1-t_j/t_{k-1})^\alpha}, \quad j \in \{2,\dots,k-1\}, \end{aligned}$$

and define the function

$$h_1(x) = \frac{(1-x/t_{k+1})^\alpha - (1-x/t_k)^\alpha}{(1-x/t_k)^\alpha - (1-x/t_{k-1})^\alpha},$$

then it suffices to show that $h_1'(x) < 0$ for $0 < x < t_{k-1}$. Observing that

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}x} \ln(h_1(x)) = -\frac{\alpha}{x} \Bigg[&\frac{(x/t_{k+1})(1-x/t_{k+1})^{\alpha-1} - (x/t_k)(1-x/t_k)^{\alpha-1}}{(1-x/t_{k+1})^\alpha - (1-x/t_k)^\alpha} \\ &- \frac{(x/t_k)(1-x/t_k)^{\alpha-1} - (x/t_{k-1})(1-x/t_{k-1})^{\alpha-1}}{(1-x/t_k)^\alpha - (1-x/t_{k-1})^\alpha} \Bigg]. \end{aligned}$$

Letting $h_2(x) = (1-x)x^{\alpha-1}, h_3(x) = x^\alpha$, by Cauchy's mean-value theorem, there exists $\eta \in (1-x/t_k, 1-x/t_{k+1})$ and $\xi \in (1-x/t_{k-1}, 1-x/t_k)$ such that

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}x}\left(\ln(h_1(x))\right) &= -\frac{\alpha}{x}\left(\frac{h_2'(\eta)}{h_3'(\eta)} - \frac{h_2'(\xi)}{h_3'(\xi)} \right) = -\frac{\alpha}{x}\left(\left(\frac{\alpha-1}{\alpha\eta} - 1 \right) - \left(\frac{\alpha-1}{\alpha\xi} - 1 \right) \right) \\ &< 0 \end{aligned}$$

because $0 < \xi < \eta$. This implies that $h_1'(x) < 0$ for $0 < x < t_{k-1}$ and finishes inductive step of the induction. Hence (3.9) is proved. $\square$

Remark 3.4. (Alternative definition of $D^\alpha_{\mathcal{P}}$) Let us define, for $k \le n$,

$$b_{\mathcal{P},nk} = \sum_{j=k}^{n} \mathbf{K}^{-1}_{\mathcal{P},nj}.$$

Then, the discrete Caputo derivative (3.3) can also be written as

$$(D^\alpha_{\mathcal{P}}\mathbf{W})_n = \sum_{i=0}^{n} \mathbf{K}^{-1}_{\mathcal{P},ni} W_i = \sum_{i=1}^{n} b_{\mathcal{P},ni}\left(W_i - W_{i-1}\right),$$

which is a discrete version of the definition given in (1.1). As an immediate consequence of Theorems 3.1 and 3.3, we have

$$b_{\mathcal{P},nk} > 0, \quad b_{\mathcal{P},nk} > b_{\mathcal{P},(n+1)k}.$$

Remark 3.5. (Generalization) Notice that, for a general kernel g, property (3.8) remains valid provided $G(t) = \ln(g(t))$ satisfies $G''(t) > 0$.

3.3. *A continuous interpolant*

Given a partition $\mathcal{P}$, a sequence $\mathbf{W} \in \mathcal{H}^N$, and $W_0 \in \mathcal{H}$, we defined the discrete Caputo derivative $\left(D^\alpha_{\mathcal{P}}\mathbf{W}\right)_n$ via (3.3). Motivated by the Volterra type equation (2.17) between a continuous function w and its Caputo derivative $D^\alpha_c w$, it is possible, following Ref. 39, to define, over $\mathcal{P}$, a natural continuous interpolant of W_n by

$$\widehat{W}_{\mathcal{P}}(t) = W_0 + \frac{1}{\Gamma(\alpha)} \int_0^t (t-s)^{\alpha-1} \overline{V}_{\mathcal{P}}(s) \mathrm{d}s, \tag{3.16}$$

where $\overline{V}_{\mathcal{P}}$ is defined by

$$\overline{V}_{\mathcal{P}}(t) = \left(D^\alpha_{\mathcal{P}}\mathbf{W}\right)_{n(t)}. \tag{3.17}$$

By definition, we have that $\widehat{W}_{\mathcal{P}}(t_n) = W_n$. Moreover,

$$\widehat{W}_{\mathcal{P}}(t) = W_0 + \frac{1}{\Gamma(\alpha+1)} \sum_{j=1}^{n-1} ((t-t_{j-1})^\alpha - (t-t_j)^\alpha)(D^\alpha_{\mathcal{P}}\mathbf{W})_j + (t_n - t)^\alpha \left(D^\alpha_{\mathcal{P}}\mathbf{W}\right)_n$$

$$= \sum_{i=0}^{n(t)} W_i \varphi_{\mathcal{P},i}(t), \tag{3.18}$$

where we defined, for $i \in \{1, \dots, N\}$,

$$\varphi_{\mathcal{P},0}(t) = 1 + \frac{1}{\Gamma(\alpha+1)} \sum_{j=1}^{n(t)-1} ((t-t_{j-1})^\alpha - (t-t_j)^\alpha)\mathbf{K}^{-1}_{\mathcal{P},j0}$$
$$+ (t_n - t)^\alpha \mathbf{K}^{-1}_{\mathcal{P},n0},$$

$$\varphi_{\mathcal{P},i}(t) = \frac{1}{\Gamma(\alpha+1)} \sum_{j=i}^{n(t)-1} ((t-t_{j-1})^\alpha - (t-t_j)^\alpha)\mathbf{K}^{-1}_{\mathcal{P},ji} + (t_n - t)^\alpha \mathbf{K}^{-1}_{\mathcal{P},ni}. \tag{3.19}$$

The functions $\{\varphi_{\mathcal{P},i}\}_{i=0}^N$ play the role, in this context, of the standard "hat" basis functions used for piecewise linear interpolation over a partition $\mathcal{P}$. Indeed, they are such that any function with piecewise constant (Caputo) derivative can be written as a linear combination of them. Figure 1 illustrates the behavior of these functions. As expected, and in contrast to the hat basis functions, these functions are nonlocal, in the sense that they have global support. Something worth noticing is also that the figure seems to indicate that, as $\alpha \downarrow 0$, the functions resemble piecewise constants and, in contrast, when $\alpha \uparrow 1$ they tend to the classical hat basis functions.

An important feature of the hat basis functions is that they form a partition of unity. It is easy to check that, for any $t \in [0,T]$ we have $\sum_{i=0}^{n(t)} \varphi_{\mathcal{P},i}(t) = 1$. The following result shows that $\varphi_{\mathcal{P},i}(t) \geq 0$. Thus, for any $t \in [0,T]$, $\widehat{W}_{\mathcal{P}}(t)$ is a convex combination of its nodal values $\{W_j\}_{j=0}^N$. This observation will be crucial to derive an *a posteriori* error estimate in Sec. 5.2.

Proposition 3.6. (Positivity) *Let $\mathcal{P}$ be a partition defined as in* (2.2)*. Let the functions $\{\varphi_{\mathcal{P},i}\}_{i=0}^N$ be defined as in* (3.19)*. Then, for any $i \in \{0,\dots,N\}$ and $t \in [0,T]$, we have $\varphi_{\mathcal{P},i}(t) \geq 0$. In addition, for $t \notin \mathcal{P}$ and $i \in \{0,\dots,n(t)\}$ we have $\varphi_{\mathcal{P},i}(t) > 0$.*

Proof. By definition, for $t = t_n$, we have $\varphi_{\mathcal{P},n}(t_n) = 1$ and $\varphi_{\mathcal{P},i}(t_n) = 0$ for any $i \neq n$. Also, for $i > n(t)$, we see that $\varphi_{\mathcal{P},i}(t) = 0$, and hence it only remains to show that $\varphi_{\mathcal{P},i}(t) > 0$ for $i \leq n(t)$. To show this, consider $W_i = 1$ and $W_j = 0$ for $j \neq i$, a piecewise constant $\overline{V}_{\mathcal{P}}$ and its interpolation $\widehat{W}_{\mathcal{P}}$ defined in (3.16) and (3.17). Then our goal is to show that $\widehat{W}_{\mathcal{P}}(t) > 0$.

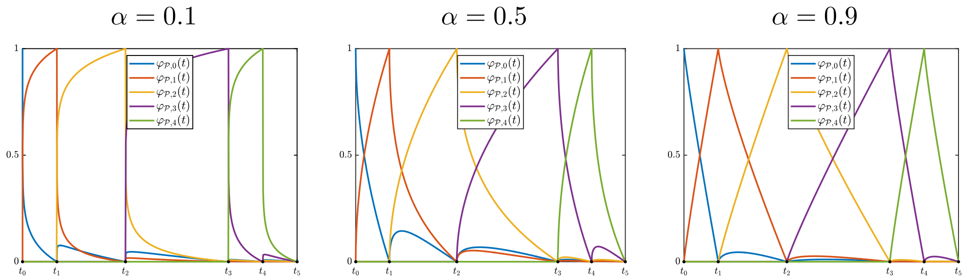

Fig. 1. Given a partition $\mathcal{P}$, the figure shows the nonlocal basis functions $\{\varphi_{\mathcal{P},i}\}_{i=0}^N$ for different values of α. Every function whose Caputo derivative is piecewise constant can be written as a linear combination of these functions. Notice that, for any partition point $\varphi_{\mathcal{P},i}(t_j) = \delta_{ij}$. In addition, Proposition 3.6 shows that these functions form a partition of unity.

If $i = n(t) > 0$, then it is easy to check by definition that $(D^\alpha_{\mathcal{P}}\mathbf{W})_n > 0$ and $(D^\alpha_{\mathcal{P}}\mathbf{W})_j = 0$ for $j \in \{1, \dots, i-1\}$. Therefore we obtain

$$\widehat{W}_{\mathcal{P}}(t) = \frac{1}{\Gamma(\alpha)} \int_0^t (t-s)^{\alpha-1} \overline{V}(s) \mathrm{d}s = \frac{(t_n - t)^\alpha}{\Gamma(\alpha+1)} (D^\alpha_{\mathcal{P}}\mathbf{W})_n > 0.$$

If $i < n(t)$, the proof is not that straightforward. The trick is to insert the time t, which is not on the partition $\mathcal{P}$, to get a new partition $\mathcal{P}' = \mathcal{P} \cup \{t\}$ and then apply Propositions 3.1 and 3.3 in an appropriate way. Let us now work out the details. Let $\mathcal{P}' = \{t'_k\}_{k=0}^{N+1}$ and notice that $t'_{n(t)} = t, t'_{n(t)+1} = t_{n(t)}$. On the basis of this partition we define the vector $\mathbf{W}' \in \mathcal{H}^{N+1}$ via $W'_j = \widehat{W}_{\mathcal{P}}(t'_j)$, then since $\overline{V}_{\mathcal{P}}$ is constant on $(t'_{n(t)-1}, t'_{n(t)+1}] = (t_{n(t)-1}, t_{n(t)}]$, we have

$$(D^\alpha_{\mathcal{P}'}\mathbf{W}')_{n(t)} = (D^\alpha_{\mathcal{P}'}\mathbf{W}')_{n(t)+1} .$$

Since the only possible nonzero components of $\mathbf{W}'$ are $W'_i = W_i = 1$ and $W'_{n(t)} = \widehat{W}_{\mathcal{P}}(t)$, therefore we deduce from the equality above that

$$\begin{aligned} \mathbf{K}^{-1}_{\mathcal{P}',n(t)i} W'_i + \mathbf{K}^{-1}_{\mathcal{P}',n(t)n(t)} W'_{n(t)} &= (D^\alpha_{\mathcal{P}'}\mathbf{W}')_{n(t)} = (D^\alpha_{\mathcal{P}'}\mathbf{W}')_{n(t)+1} \\ &= \mathbf{K}^{-1}_{\mathcal{P}',n(t)+1,i} W'_i + \mathbf{K}^{-1}_{\mathcal{P}',n(t)+1,n(t)} W'_{n(t)}, \end{aligned}$$

which can be rearranged as

$$\mathbf{K}^{-1}_{\mathcal{P}',n(t)+1,i} - \mathbf{K}^{-1}_{\mathcal{P}',n(t)i} = \widehat{W}_{\mathcal{P}}(t) (\mathbf{K}^{-1}_{\mathcal{P}',n(t)n(t)} - \mathbf{K}^{-1}_{\mathcal{P}',n(t)+1,n(t)}).$$

From Theorem 3.3 we see that $\mathbf{K}^{-1}_{\mathcal{P}',n(t)+1,i} - \mathbf{K}^{-1}_{\mathcal{P}',n(t)i} > 0$ and from Theorem 3.1 we see that $\mathbf{K}^{-1}_{\mathcal{P}',n(t)n(t)} - \mathbf{K}^{-1}_{\mathcal{P}',n(t)+1,n(t)} > 0$ as a consequence of $\mathbf{K}^{-1}_{\mathcal{P}',n(t)n(t)} > 0$ and $\mathbf{K}^{-1}_{\mathcal{P}',n(t)+1,n(t)} < 0$. This leads to the fact that $\widehat{W}_{\mathcal{P}}(t) > 0$ and finishes our proof. □

3.4. *Comparison and monotonicity*

Once we have obtained Theorems 3.1 and 3.3, it is easy to see that the properties stated in Theorem 3.3 of Ref. 39 for uniform grids also hold for general partitions. We state these below but we omit the proof.

Proposition 3.7. (Further properties of $D^\alpha_{\mathcal{P}}$) *Let $\mathcal{P}$ be a time partition. The discrete Caputo derivative defined in* (3.3) *satisfies*:

(1) (*Convex functional*) *If $\mathbf{U} = \{U_n\}_{n \geq 0} \subset \mathcal{H}$, and $\Psi : \mathcal{H} \to \mathbb{R} \cup \{+\infty\}$ is convex, then for any $\xi \in \partial\Psi(U_n)$, we have*

$$(D^\alpha_{\mathcal{P}}\Psi(\mathbf{U}))_n \leq \langle (D^\alpha_{\mathcal{P}}\mathbf{U})_n, \xi \rangle . \tag{3.20}$$

(2) (*Discrete comparison*) *Let $f : \mathbb{R}_+ \cup \{0\} \times \mathbb{R} \to \mathbb{R}$ be such that there is $L \geq 0$ for which for all $s \geq 0$ the mapping*

$$z \mapsto f(s,z) - Lz$$

is non-increasing. Assume that $\mathbf{u} = \{u_n\}_{n\ge 0}, \mathbf{v} = \{v_n\}_{n\ge 0}, \mathbf{w} = \{w_n\}_{n\ge 0} \subset \mathbb{R}$ *satisfy* $u_0 \le v_0 \le w_0$, *and*

$$(D^\alpha_{\mathcal{P}}\mathbf{u})_n \le f(t_n, u_n), \quad (D^\alpha_{\mathcal{P}}\mathbf{v})_n = f(t_n, v_n), \quad (D^\alpha_{\mathcal{P}}\mathbf{w})_n \ge f(t_n, w_n).$$

Then, if $\mathbf{K}^{-1}_{\mathcal{P},ii} > L$ *for every* i, *we have* $u_n \le v_n \le w_n$ *for all* $n \ge 0$. *For our scheme,* $\mathbf{K}^{-1}_{\mathcal{P},ii} > L$ *holds provided that* $\tau^\alpha < \Gamma(\alpha+1)/L$ *and thus the statement is true for any partition* $\mathcal{P}$ *if one can take* $L = 0$.

(3) (*Discrete comparison in integral form*) *Let* $f : \mathbb{R}_+ \cup \{0\} \times \mathbb{R} \to \mathbb{R}$ *be such that, for all* $s \ge 0$, *the mapping*

$$z \mapsto f(s, z)$$

is non-decreasing and Lipschitz, with Lipschitz constant L. *Assume that the sequences*

$$\{u_n\}_{n\ge 0}, \quad \{v_n\}_{n\ge 0}, \quad \{w_n\}_{n\ge 0} \subset \mathbb{R}$$

satisfy

$$u_n \le u_0 + (J^\alpha_{\mathcal{P}} f_u)_n, \quad v_n = v_0 + (J^\alpha_{\mathcal{P}} f_v)_n, \quad w_n \ge w_0 + (J^\alpha_{\mathcal{P}} f_w)_n,$$

where $J^\alpha_{\mathcal{P}} f_u$ *is defined by*

$$\begin{aligned}(J^\alpha_{\mathcal{P}} f_u)_n &= \sum_{k=1}^n \frac{f(t_k, u_k)}{\Gamma(\alpha)} \int_{t_{k-1}}^{t_k} (t_n - s)^{\alpha-1} \mathrm{d}s \\ &= \sum_{k=1}^n \frac{(t_n - t_{k-1})^\alpha - (t_n - t_k)^\alpha}{\Gamma(\alpha+1)} f(t_k, u_k).\end{aligned}$$

Then if $\tau^\alpha < \Gamma(\alpha+1)/L$, *we have* $u_n \le v_n \le w_n$.

We comment that the first two properties also hold for other schemes satisfying Theorem 3.1. In addition, we also notice that Theorem 4.1 in Ref. 41, which is stated for uniform grids, also holds for general partitions.

Proposition 3.8. (Discrete monotonicity) *Let* $\mathcal{P}$ *be a time partition, and* $g : \mathbb{R} \to \mathbb{R}$ *be such that there is* $L \ge 0$ *for which*

$$z \mapsto g(z) - Lz$$

is non-increasing. Assume that $\{U_n\}_{n\ge 0} \subset \mathbb{R}$ *satisfies*

$$(D^\alpha_{\mathcal{P}}\mathbf{U})_n = g(U_n), \tag{3.21}$$

where the discrete Caputo derivative $D^\alpha_{\mathcal{P}}$ *is defined in* (3.3). *If* $\mathbf{K}^{-1}_{\mathcal{P},ii} > L$ *for any* i, *then the sequence* $\{U_n\}_{n\ge 0}$ *is monotone. For our scheme this, in particular, requires that* $\tau^\alpha < \Gamma(\alpha+1)/L$ *and thus the statement is true for any partition* $\mathcal{P}$ *if one can take* $L = 0$.

Proof. We first note that, if $g(U_0) = 0$, then $\{U_n\}_{n\ge 0}$ is constant and the statement is trivial. We will prove by induction that if $g(U_0) > 0$, then $\{U_n\}_{n\ge 0}$ is strictly increasing. The proof that $\{U_n\}_{n\geq 0}$ is strictly decreasing when $g(U_0) < 0$ proceeds in a similar way.

We first notice that from (3.21)

$$h_1(U_1) = \mathbf{K}^{-1}_{\mathcal{P},11}(U_1 - U_0) - g(U_1) = 0.$$

From the assumptions, we know that the function h_1 is strictly increasing. Therefore we have $U_1 > U_0$ because

$$h_1(U_0) = 0 - g(U_0) < 0.$$

Suppose now that we have already proved that $U_0 < U_1 < \cdots < U_n$, and we want to show $U_n < U_{n+1}$. From (3.21), it holds that

$$h_{n+1}(U_{n+1}) = \sum_{k=0}^{n} \mathbf{K}^{-1}_{\mathcal{P},(n+1)k}(U_k - U_{n+1}) - g(U_{n+1}) = 0.$$

Since h_{n+1} is also strictly increasing, it suffices to show that

$$\begin{aligned} h_{n+1}(U_n) &= \sum_{k=0}^{n} \mathbf{K}^{-1}_{\mathcal{P},(n+1)k}(U_k - U_n) - g(U_n) \\ &= \sum_{k=0}^{n-1} \mathbf{K}^{-1}_{\mathcal{P}(n+1)k}(U_k - U_n) - g(U_n) < 0. \end{aligned}$$

Recall that

$$g(U_n) = (D^\alpha_{\mathcal{P}}\mathbf{U})_n = \sum_{k=0}^{n-1} \mathbf{K}^{-1}_{\mathcal{P},nk}(U_k - U_n),$$

we thus have

$$h_{n+1}(U_n) = \sum_{k=0}^{n-1} (\mathbf{K}^{-1}_{\mathcal{P},n+1,k} - \mathbf{K}^{-1}_{\mathcal{P},nk})(U_k - U_n) < 0$$

because $\mathbf{K}^{-1}_{\mathcal{P},n+1,k} - \mathbf{K}^{-1}_{\mathcal{P},nk} > 0$, and $U_k - U_n < 0$. This shows that the sequence $\{U_n\}_{n\geq 0}$ is strictly increasing as claimed. □

We remark that the previous result also holds for other schemes satisfying Theorem 3.3. This proposition leads to the following fact, which will be useful in Sec. 4.3.

Proposition 3.9. (Monotonicity under refinement) *Consider a non-increasing function $g : \mathbb{R} \to \mathbb{R}$ and two different time partitions $\mathcal{P}$ and $\widetilde{\mathcal{P}}$, where $\mathcal{P}$ is a refinement of $\widetilde{\mathcal{P}}$. Assume that the sequences $\{\widetilde{U}_n\}_{n\geq 0}, \{U_n\}_{n\geq 0} \subset \mathbb{R}$ satisfy*

$$(D^\alpha_{\widetilde{\mathcal{P}}}\widetilde{\mathbf{U}})_n = g(\widetilde{U}_n), \quad (D^\alpha_{\mathcal{P}}\mathbf{U})_n = g(U_n),$$

and $\widetilde{U}_0 = U_0$. If $g(U_0) < 0$, then we have that

$$\widehat{\widetilde{U}}_{\widetilde{\mathcal{P}}}(t) \geq \widehat{U}_{\mathcal{P}}(t), \quad t \geq 0,$$

where we recall that these interpolants were defined in (3.16). *Similarly, if* $g(U_0) > 0$, *then* $\widehat{\widetilde{U}}_{\widetilde{\mathcal{P}}}(t) \leq \widehat{U}_{\mathcal{P}}(t)$.

Proof. It suffices to prove the result for the case $g(U_0) < 0$ and $\mathcal{P}$ is a refinement obtained by inserting one node into $\widetilde{\mathcal{P}}$. Let $\mathcal{P} = \{t_i\}$, and assume that $t_k \notin \widetilde{\mathcal{P}}$. Define the sequence $\{W_i\}_{i\geq 0}$ using the values of $\widehat{\widetilde{U}}_{\widetilde{\mathcal{P}}}$ at $\mathcal{P}$, i.e.

$$W_i = \widehat{\widetilde{U}}_{\widetilde{\mathcal{P}}}(t_i) = \begin{cases} \widetilde{U}_i, & i < k, \\ \widehat{\widetilde{U}}_{\widetilde{\mathcal{P}}}(t_k), & i = k, \\ \widetilde{U}_{i-1}, & i > k. \end{cases}$$

Since the Caputo derivative of $\widehat{\widetilde{U}}_{\widetilde{\mathcal{P}}}$ is piecewise constant over $\widetilde{\mathcal{P}}$, this implies that

$$\left(D^{\alpha}_{\mathcal{P}}\mathbf{W}\right)_i = \begin{cases} g(W_i), & i \neq k, \\ g(W_{i+1}), & i = k. \end{cases}$$

Notice that, by Theorem 3.8, the sequence $\{W_i\}_{n\geq 0}$ is decreasing. Hence the equation above leads to

$$\left(D^{\alpha}_{\mathcal{P}}\mathbf{W}\right)_n \geq g(W_n).$$

Therefore by the discrete comparison principle of Theorem 3.7, we have

$$W_n \geq U_n.$$

Theorem 3.6 then implies that $\widehat{\widetilde{U}}_{\widetilde{\mathcal{P}}}(t) \geq \widehat{U}_{\mathcal{P}}(t)$. □

4. Time Fractional Gradient Flows: Theory

We have now set the stage for the study of time fractional gradient flows, which were formally described in (1.2). Throughout the remaining of our discussion we shall assume that the initial condition satisfies $u_0 \in D(\Phi)$ and that $f \in L^2_\alpha(0, T; \mathcal{H})$. We begin by commenting that the case $f = 0$ was already studied in Sec. 5 of 39 where they studied so-called *strong solutions*, see Definition 5.4 in Ref. 39. Here we trivially extend their definition to the case $f \neq 0$.

Definition 4.1. (Strong solution) A function $u \in L^1_{loc}([0, T); \mathcal{H})$ is a strong solution to (1.2) if

(i) (Initial condition)

$$\lim_{t\downarrow 0} \fint_0^t \|u(s) - u_0\| \mathrm{d}s = 0.$$

(ii) (Regularity) $D^{\alpha}_c u(t) \in L^1_{loc}([0, T); \mathcal{H})$.

(iii) (Evolution) For almost every $t \in [0, T)$, we have $f(t) - D^{\alpha}_c u(t) \in \partial\Phi(u(t))$.

4.1. *Energy solutions*

Since $\mathcal{H}$ is a Hilbert space, we will mimic the theory for classical gradient flows and introduce the notion of energy solutions for (1.2). To motivate it, suppose that at some $t \in (0,T)$

$$f(t) - D_c^\alpha u(t) \in \partial\Phi(u(t)),$$

then, by definition of the subdifferential, this is equivalent to the *evolution variational inequality* (EVI)

$$\langle D_c^\alpha u(t), u(t) - w\rangle + \Phi(u(t)) - \Phi(w) \leq \langle f(t), u(t) - w\rangle, \quad \forall\, w \in \mathcal{H}. \tag{4.1}$$

Definition 4.2. (Energy solution) The function $u \in L^2(0,T;\mathcal{H})$ is an energy solution to (1.2) if

(i) (Initial condition)

$$\lim_{t\downarrow 0} \fint_0^t \|u(s) - u_0\|^2 \mathrm{d}s = 0.$$

(ii) (Regularity) $D_c^\alpha u \in L^2(0,T;\mathcal{H})$.

(iii) (EVI) For any $w \in L^2(0,T;\mathcal{H})$

$$\int_0^T \left[\langle D_c^\alpha u(t), u(t) - w(t)\rangle + \Phi(u(t)) - \Phi(w(t))\right] \mathrm{d}t \leq \int_0^T \langle f(t), u(t) - w(t)\rangle \mathrm{d}t. \tag{4.2}$$

Notice that, provided $u_0 \in D(\Phi)$ we can set $w(t) = u_0$ in (4.2) and obtain that $\int_0^T \Phi(u(t))\mathrm{d}t < \infty$, which motivates the name for this notion of solution. In addition, as the following result shows, any energy solution is a strong solution.

Proposition 4.3. (Energy versus strong) *An energy solution of* (1.2) *is also a strong solution.*

Proof. Evidently, it suffices to prove that $f(t) - D_c^\alpha u(t) \in \partial\Phi(u(t))$ for almost every $t \in (0,T)$. Let $w_0 \in \mathcal{H}$, $t_0 \in (0,T)$, and choose $h > 0$ sufficiently small so that $(t_0 - h, t_0 + h) \subset (0,T)$. Define

$$w(t) = u(t) - \chi_{(t_0-h,t_0+h)}(u(t) - w_0) \in L^2(0,T;\mathcal{H}),$$

where by χ_S we denote the characteristic function of the set S. This choice of test function on (4.2) gives

$$\fint_{t_0-h}^{t_0+h} \langle D_c^\alpha u(t) - f(t), u(t) - w_0\rangle \mathrm{d}t + \fint_{t_0-h}^{t_0+h} (\Phi(u(t)) - \Phi(w_0))\mathrm{d}t \leq 0.$$

The assumptions of an energy solution guarantee that all terms inside the integrals belong to $L^1(0,T;\mathbb{R})$ so that for almost every t_0 we have, as $h \downarrow 0$, that

$$\langle D_c^\alpha u(t_0) - f(t_0), w_0\rangle + \Phi(u(t_0)) - \Phi(w_0) \leq 0,$$

which is (4.1) and, as we intended to show, is equivalent to the claim. $\square$

Remark 4.4. (Coercivity) By introducing the coercivity modulus of Definition 2.1 one realizes that an energy solution u satisfies, instead of (4.1) and (4.2), the stronger inequalities

$$\begin{aligned}\langle D_c^\alpha u(t), u(t) - w\rangle + \Phi(u(t)) - \Phi(w)& \\ + \sigma(u(t); w) \leq \langle f(t), u(t) - w\rangle,& \quad \forall\, w \in \mathcal{H},\end{aligned} \tag{4.3}$$

and, for any $w \in L^2(0,T;\mathcal{H})$,

$$\begin{aligned}&\int_0^T [\langle D_c^\alpha u(t), u(t) - w(t)\rangle + \Phi(u(t)) - \Phi(w(t)) + \sigma(u(t); w(t))]\mathrm{d}t \\ &\quad \leq \int_0^T \langle f(t), u(t) - w(t)\rangle \mathrm{d}t.\end{aligned} \tag{4.4}$$

4.2. *Existence and uniqueness*

In this section, we will prove the following theorem on the existence and uniqueness of energy solutions to (1.2) in the sense of Theorem 4.2. The main result that we will prove reads as follows.

Theorem 4.5. (Well posedness) *Assume that the energy Φ is convex, l.s.c., and with nonempty effective domain. Let $u_0 \in D(\Phi)$ and $f \in L^2_\alpha(0,T;\mathcal{H})$. In this setting, the fractional gradient flow problem* (1.2) *has a unique energy solution u, in the sense of Theorem* 4.2. *For almost every $t \in (0,T)$, the solution u satisfies that $f(t) - D_c^\alpha u(t) \in \partial\Phi(u(t))$ and for any $t \in [0,T]$ we have*

$$u(t) = u_0 + \frac{1}{\Gamma(\alpha)} \int_0^t (t-s)^{\alpha-1} D_c^\alpha u(s)\mathrm{d}s. \tag{4.5}$$

In addition, $u \in C^{0,\alpha/2}([0,T];\mathcal{H})$ with modulus of continuity

$$\|u(t_2) - u(t_1)\| \leq C|t_2 - t_1|^{\alpha/2}(\|f\|^2_{L^2_\alpha(0,T;\mathcal{H})} + \Phi(u_0) - \Phi_{\inf})^{1/2}, \quad \forall\, t_1, t_2, \in [0,T], \tag{4.6}$$

where the constant C depends only on α.

We point out that our assumptions are weaker than those in Theorem 5.10 of Ref. 39. First, we allow for a nonzero right-hand side. In addition, we do not require Assumption 5.9 of Ref. 39, which is a sort of weak–strong continuity of subdifferentials.

The remainder of this section will be dedicated to the proof of Theorem 4.5. To accomplish this, we follow a similar approach to Sec. 5 in Ref. 39. To show existence of solutions, we consider a sort of fractional minimizing movements scheme. We introduce a partition $\mathcal{P}$ with maximal time step τ and compute the sequence $\mathbf{U} = \{U_n\}_{n=0}^N \subset \mathcal{H}$ as follows. Assume $U_0 \in D(\Phi)$ is given, the nth iterate, for $n \in$

$\{1, \dots, N\}$, is defined recursively via

$$F_n - (D^\alpha_\mathcal{P}\mathbf{U})_n \in \partial\Phi(U_n), \tag{4.7}$$

where

$$F_n = \fint_{t_{n-1}}^{t_n} f(t)\mathrm{d}t. \tag{4.8}$$

We will usually choose $U_0 = u_0$, but other choices of $U_0 \in D(\Phi)$ are also allowed.

From the approximation scheme (4.7) and the expression of the discrete Caputo derivative $(D^\alpha_\mathcal{P}\mathbf{U})_n$ given in (3.3), it is clear that

$$U_n = \operatorname*{arg\,min}_{w\in\mathcal{H}} \left(\Phi(w) - \langle F_n, w\rangle - \frac{1}{2}\sum_{i=0}^{n-1} \mathbf{K}^{-1}_{\mathcal{P},ni}\|w - U_i\|^2 \right). \tag{4.9}$$

Thanks to Theorem 3.1, for $i = 0, \dots, n-1$, we have that $\mathbf{K}^{-1}_{\mathcal{P},ni} < 0$ and as a consequence the functional on the right-hand side of (4.9) is uniformly convex. Combining with the fact that Φ is l.s.c., the functional on the right-hand side has a unique minimizer, and hence U_n is well defined.

Now, in order to define a continuous in time function from $\mathbf{U}$, we use the interpolation introduced in (3.16). Let $\overline{V}_\mathcal{P}(t) = (D^\alpha_\mathcal{P}\mathbf{U})_{n(t)}$. Then we have

$$\widehat{U}_\mathcal{P}(t) = U_0 + \frac{1}{\Gamma(\alpha)}\int_0^t (t-s)^{\alpha-1}\overline{V}_\mathcal{P}(s)\mathrm{d}s. \tag{4.10}$$

Recall that $\overline{F}_\mathcal{P}$ can be defined from $\{F_n\}_{n=1}^N$ using (2.3) and that Theorem 2.4 showed that $\overline{F}_\mathcal{P} \in L^2_\alpha(0,T;\mathcal{H})$ with a norm bounded independently of $\mathcal{P}$. We now obtain some suitable bounds for $\widehat{U}_\mathcal{P}$ and $\overline{V}_\mathcal{P}$.

Lemma 4.6. (*A priori* bounds) *Let $\mathcal{P}$ be any partition. The functions $\widehat{U}_\mathcal{P}$ and $\overline{V}_\mathcal{P}$ satisfy*

$$\begin{aligned}
\sup_{t\in[0,T]} \Phi(\widehat{U}_\mathcal{P}(t)) &\le \Phi(U_0) + \frac{1}{4\Gamma(\alpha)}\|\overline{F}_\mathcal{P}\|^2_{L^2_\alpha(0,T;\mathcal{H})} \\
&\le \Phi(U_0) + C\|f\|^2_{L^2_\alpha(0,T;\mathcal{H})}, \\
\|\overline{V}_\mathcal{P}\|^2_{L^2_\alpha(0,T;\mathcal{H})} &= \sup_{t\in[0,T]}\int_0^t (t-s)^{\alpha-1}\|\overline{V}_\mathcal{P}(s)\|^2\mathrm{d}s \\
&\le C(\|f\|^2_{L^2_\alpha(0,T;\mathcal{H})} + \Phi(U_0) - \Phi_{\inf}),
\end{aligned} \tag{4.11}$$

where the constant C only depends on α.

Proof. Since $F_n - (D^\alpha_\mathcal{P}\mathbf{U})_n \in \partial\Phi(U_n)$, one has

$$\Phi(U_n) - \Phi(U_i) \le \langle F_n - (D^\alpha_\mathcal{P}\mathbf{U})_n, U_n - U_i\rangle.$$

Therefore noticing that $\mathbf{K}_{\mathcal{P},ni}^{-1} < 0$ for $i \in \{0, \ldots, n-1\}$, we get (see also (3.20))

$$\begin{aligned} (D_{\mathcal{P}}^{\alpha}\Phi(\mathbf{U}))_n &= -\sum_{i=0}^{n-1} \mathbf{K}_{\mathcal{P},ni}^{-1}(\Phi(U_n) - \Phi(U_i)) \\ &\leq -\sum_{i=0}^{n-1} \mathbf{K}_{\mathcal{P},ni}^{-1}\langle F_n - (D_{\mathcal{P}}^{\alpha}\mathbf{U})_n \,, U_n - U_i\rangle \\ &= \langle F_n - (D_{\mathcal{P}}^{\alpha}\mathbf{U})_n \,, (D_{\mathcal{P}}^{\alpha}\mathbf{U})_n\rangle, \end{aligned} \tag{4.12}$$

where we denoted $\Phi(\mathbf{U}) = \{\Phi(U_n)\}_{n=0}^{N}$.

We can now proceed to obtain the claimed estimates. To prove the first one, we use that

$$(D_{\mathcal{P}}^{\alpha}\Phi(\mathbf{U}))_n \leq \langle F_n - (D_{\mathcal{P}}^{\alpha}\mathbf{U})_n, (D_{\mathcal{P}}^{\alpha}\mathbf{U})_n\rangle \leq \frac{1}{4}\|F_n\|^2$$

to obtain that for any n,

$$\begin{aligned} \Phi(U_n) &= \Phi(U_0) + \sum_{i=1}^{n} \mathbf{K}_{\mathcal{P},ni}(D_{\mathcal{P}}^{\alpha}\Phi(\mathbf{U}))_i \\ &\leq \Phi(U_0) + \frac{1}{4}\sum_{i=1}^{n} \mathbf{K}_{\mathcal{P},ni}\|F_i\|^2 \\ &= \Phi(U_0) + \frac{1}{4\Gamma(\alpha)}\int_0^{t_n}(t_n - s)^{\alpha-1}\|\overline{F}_{\mathcal{P}}(s)\|^2 \mathrm{d}s \\ &\leq \Phi(U_0) + C\|f\|_{L^2_\alpha(0,T;\mathcal{H})}^2, \end{aligned}$$

where the constant C depends only on α. Now, since Theorem 3.6 has shown that $\widehat{U}_{\mathcal{P}}$ is a convex combination of the values U_n, we have

$$\begin{aligned} \Phi(\widehat{U}_{\mathcal{P}}(t)) &= \Phi\left(\sum_{i=0}^{N} \varphi_{\mathcal{P},i}(t)U_i\right) \leq \sum_{i=0}^{N} \varphi_{\mathcal{P},i}(t)\Phi\left(U_i\right) \\ &\leq \max_n \Phi(U_n) \leq \Phi(U_0) + C\|f\|_{L^2_\alpha(0,T;\mathcal{H})}^2, \end{aligned}$$

which finishes the proof of the first claim.

We now proceed to prove the second claim. Using (4.12) we get

$$\begin{aligned} \Phi_{\inf} &\leq \Phi(\widehat{U}_{\mathcal{P}}(t)) \leq \Phi(U_0) + \frac{1}{\Gamma(\alpha)} \\ &\quad \times \int_0^t (t-s)^{\alpha-1}\langle \overline{F}_{\mathcal{P}}(s) - \overline{V}_{\mathcal{P}}(s), \overline{V}_{\mathcal{P}}(s)\rangle \mathrm{d}s \\ &\leq \Phi(U_0) + \frac{1}{\Gamma(\alpha)}\left(\int_0^t (t-s)^{\alpha-1}\|\overline{F}_{\mathcal{P}}(s)\|^2 \mathrm{d}s\right)^{1/2} \end{aligned}$$

$$\times \left(\int_0^t (t-s)^{\alpha-1} \|\overline{V}_{\mathcal{P}}(s)\|^2 \mathrm{d}s \right)^{1/2}$$
$$- \frac{1}{\Gamma(\alpha)} \int_0^t (t-s)^{\alpha-1} \|\overline{V}_{\mathcal{P}}(s)\|^2 \mathrm{d}s,$$

for any $t \in [0,T]$. This implies that

$$\int_0^t (t-s)^{\alpha-1} \|\overline{V}_{\mathcal{P}}(s)\|^2 \mathrm{d}s \le \|\overline{F}_{\mathcal{P}}\|^2_{L^2_\alpha(0,T;\mathcal{H})} + 2\Gamma(\alpha)(\Phi(U_0) - \Phi_{\inf}),$$

which, using Theorem 2.4, implies the result. □

Remark 4.7. (The function $\widehat{\Phi}$) Notice that, during the course of the proof of the first estimate in (4.11) we also showed that, if we define $\widehat{\Phi}_{\mathcal{P}}(t) = \sum_{i=0}^N \varphi_{\mathcal{P},i}(t)\Phi(U_i)$, then $\widehat{\Phi}(t)$ is the interpolation of $\Phi_{\mathcal{P}}(\mathbf{U})$ with piecewise constant Caputo derivative. Moreover,

$$D_c^\alpha \widehat{\Phi}_{\mathcal{P}}(t) \le \frac{1}{4} \left\|\overline{F}_{\mathcal{P}}(t)\right\|^2.$$

These estimates immediately yield a modulus of continuity estimate on the interpolant $\widehat{U}_{\mathcal{P}}$ which is independent of the partition $\mathcal{P}$.

Lemma 4.8. (Hölder continuity) *Let $\mathcal{P}$ be any partition and $\mathbf{U} \in \mathcal{H}^N$ be the solution to* (4.7) *associated to this partition. For $t_1, t_2 \in [0,T]$ the interpolant $\widehat{U}_{\mathcal{P}}$, defined in* (3.16), *satisfies*

$$\|\widehat{U}(t_2) - \widehat{U}(t_1)\| \le C|t_2 - t_1|^{\alpha/2} (\|f\|^2_{L^2_\alpha(0,T;\mathcal{H})} + \Phi(U_0) - \Phi_{\inf})^{1/2},$$

where the constant C depends only on α.

Proof. As proved in Lemma 5.8 of Ref. 39, $D_c^\alpha w \in L^2_\alpha(0,T;\mathcal{H})$ guarantees $w \in C^{0,\alpha/2}([0,T];\mathcal{H})$. Therefore using $D_c^\alpha \widehat{U} = \overline{V}_\alpha \in L^2_\alpha(0,T;\mathcal{H})$ and the estimate from Theorem 4.6, we obtain the result. □

Next we control the difference between discrete solutions corresponding to different partitions.

Lemma 4.9. (Equicontinuity) *Let, for $i = 1, 2$, $\mathcal{P}_i$ be partitions of $[0,T]$ with maximal step size τ_i, respectively, and denote by $\mathbf{U}^{(i)}$ the associated solutions to* (4.7)*. Let $\widehat{U}_i$ be their interpolations, defined by* (4.10)*, and $\overline{U}_i$ be their piecewise constant interpolations as in* (2.3)*. Assuming that $U_0^{(i)} = U_0$ we have*

$$\|\widehat{U}_1 - \widehat{U}_2\|_{L^\infty(0,T;\mathcal{H})} \le C[\tau_1^{\alpha/2} + \tau_2^{\alpha/2}][\|f\|^2_{L^2_\alpha(0,T;\mathcal{H})} + \Phi(U_0) - \Phi_{\inf}]^{1/2}, \quad (4.13)$$

and

$$\sup_{t\in[0,T]} \int_0^t (t-s)^{\alpha-1} \rho(\overline{U}_1(s), \overline{U}_2(s)) \mathrm{d}s$$
$$\le C(\tau_1^\alpha + \tau_2^\alpha)(\|f\|^2_{L^2_\alpha(0,T;\mathcal{H})} + \Phi(U_0) - \Phi_{\inf}), \quad (4.14)$$

where the constant C only depends on α.

Proof. For almost every $t \in [0,T]$, we have that

$$\langle D_c^\alpha(\widehat{U}_1 - \widehat{U}_2), \widehat{U}_1 - \widehat{U}_2\rangle = \mathrm{I} + \mathrm{II} + \mathrm{III}, \tag{4.15}$$

where

$$\begin{aligned}
\mathrm{I} &= \langle (\overline{F}_2 - D_c^\alpha \widehat{U}_2) - (\overline{F}_1 - D_c^\alpha \widehat{U}_1), \overline{U}_1 - \overline{U}_2\rangle \leq -\rho(\overline{U}_1, \overline{U}_2),\\
\mathrm{II} &= \langle (\overline{F}_2 - D_c^\alpha \widehat{U}_2) - (\overline{F}_1 - D_c^\alpha \widehat{U}_1), (\widehat{U}_1 - \overline{U}_1) - (\widehat{U}_2 - \overline{U}_2)\rangle,\\
\mathrm{III} &= \langle \overline{F}_1 - \overline{F}_2, \widehat{U}_1 - \widehat{U}_2\rangle,
\end{aligned}$$

where to bound I we used that $\overline{F}_i(t) - D_c^\alpha \widehat{U}_i(t) \in \partial\Phi(\overline{U}_i(t))$ and Theorem 2.1. Define now

$$\begin{aligned}
G(t) &= \frac{1}{\Gamma(\alpha)}\int_0^t (t-s)^{\alpha-1}(\overline{F}_1(s) - \overline{F}_2(s))\mathrm{d}s\\
&= \frac{1}{\Gamma(\alpha)}\int_0^t (t-s)^{\alpha-1}(\overline{F}_1(s) - f(s))\mathrm{d}s\\
&\quad - \frac{1}{\Gamma(\alpha)}\int_0^t (t-s)^{\alpha-1}(\overline{F}_2(s) - f(s))\mathrm{d}s,
\end{aligned}$$

so that $D_c^\alpha G(t) = \overline{F}_1(t) - \overline{F}_2(t)$ and by (2.10) of Theorem 2.5 one further has

$$\|G\|_{L^\infty(0,T;\mathcal{H})} \leq C(\tau_1^{\alpha/2} + \tau_2^{\alpha/2})\|f\|_{L^2_\alpha(0,T;\mathcal{H})}, \tag{4.16}$$

where C is a constant that depends only on α. Using these estimates, from (4.15) we deduce that

$$\begin{aligned}
&\langle D_c^\alpha(\widehat{U}_1 - \widehat{U}_2 - G), \widehat{U}_1 - \widehat{U}_2 - G\rangle + \rho(\overline{U}_1, \overline{U}_2)\\
&\quad \leq \mathrm{II} - \langle D_c^\alpha(\widehat{U}_1 - \widehat{U}_2 - G), G\rangle.
\end{aligned} \tag{4.17}$$

Set $w = \widehat{U}_1 - \widehat{U}_2 - G$. By (2.18) we have that

$$\frac{1}{2} D_c^\alpha \|w(t)\|^2 + \rho(\overline{U}_1, \overline{U}_2) \leq \mathrm{II} - \langle D_c^\alpha w, G\rangle,$$

and, using (2.17) and (4.16), we then conclude

$$\begin{aligned}
&\frac{1}{2}\|\widehat{U}_1(t) - \widehat{U}_2(t)\|^2 + \frac{1}{\Gamma(\alpha)}\int_0^t (t-s)^{\alpha-1}\rho(\overline{U}_2(s), \overline{U}_2(s))\mathrm{d}s\\
&\quad \leq \frac{2}{\Gamma(\alpha)}\int_0^t (t-s)^{\alpha-1}(\mathrm{II}(s) - \langle D_c^\alpha w(s), G(s)\rangle)\mathrm{d}s + C(\tau_1^{\alpha/2}\\
&\qquad + \tau_2^{\alpha/2})\|f\|_{L^2_\alpha(0,T;\mathcal{H})}.
\end{aligned}$$

It remains then to estimate the fractional integral on the right-hand side. We estimate each term separately.

First, owing to Theorems 2.4 and 4.6 we have, for $i = 1, 2$, that

$$\|\overline{F}_i - D_c^\alpha \widehat{U}_i\|_{L^2_\alpha(0,T;\mathcal{H})} \leq C(\|f\|^2_{L^2_\alpha(0,T;\mathcal{H})} + \Phi(U_0) - \Phi_{\inf})^{1/2}.$$

Therefore using the Cauchy–Schwarz inequality, for any $t \in [0,T]$, we have

$$\int_0^t (t-s)^{\alpha-1}|\mathrm{II}(s)|\mathrm{d}s$$
$$\leq C(\|f\|^2_{L^2_\alpha(0,T;\mathcal{H})} + \Phi(U_0) - \Phi_{\inf})^{1/2} \sum_{i=1}^{2} \left\| \widehat{U}_i - \overline{U}_i \right\|_{L^2_\alpha(0,T;\mathcal{H})}.$$

Recalling that $\overline{U}_i(t) = \widehat{U}_i(\lceil t \rceil_i)$ we can invoke Theorem 2.6 and, again, Theorem 4.6 to arrive at

$$\int_0^t (t-s)^{\alpha-1}|\mathrm{II}(s)|\mathrm{d}s \leq C(\tau_1^{\alpha/2} + \tau_2^{\alpha/2})(\|f\|^2_{L^2_\alpha(0,T;\mathcal{H})} + \Phi(U_0) - \Phi_{\inf}).$$

Finally, for the remaining term, we use the Cauchy–Schwarz inequality and get

$$\int_0^t (t-s)^{\alpha-1} \left|\left\langle D^\alpha_c w, G\right\rangle(s)\right| \mathrm{d}s$$
$$\leq \left(\int_0^t (t-s)^{\alpha-1} \left\|D^\alpha_c w(s)\right\|^2 \mathrm{d}s\right)^{1/2} \left(\int_0^t (t-s)^{\alpha-1} \left\|G(s)\right\|^2 \mathrm{d}s\right)^{1/2}$$
$$\leq \|D^\alpha_c w\|_{L^2_\alpha(0,T;\mathcal{H})} \|G\|_{L^2_\alpha(0,T;\mathcal{H})}.$$

To estimate the norm of G we apply (2.11) from Theorem 2.5 with $\beta = \alpha$ to obtain

$$\|G\|_{L^2_\alpha(0,T;\mathcal{H})} \leq C(\tau_1^\alpha + \tau_1^\alpha)\|f\|_{L^2_\alpha(0,T;\mathcal{H})}.$$

Furthermore, Theorems 2.4 and 4.6 guarantee that

$$\|D^\alpha_c w\|_{L^2_\alpha(0,T;\mathcal{H})} \leq C(\|f\|^2_{L^2_\alpha(0,T;\mathcal{H})} + \Phi(U_0) - \Phi_{\inf})^{1/2}.$$

Combining all estimates proves the desired result. □

We are finally able to prove Theorem 4.5. We will follow the same approach as in Theorem 5.10 of Ref. 39; we will pass to the limit $\tau_i \downarrow 0$ and study the limit of discrete solutions $\widehat{U}_i$.

Proof of Theorem 4.5. Let us first prove uniqueness of energy solutions. Suppose that we have two energy solutions u_1, u_2 to (1.2). Let $t \in (0,T)$ be arbitrary and $h > 0$ be sufficiently small so that $(t-h, t+h) \subset [0,T]$. Setting as test function, in the EVI that characterizes u_1, the function $w = u_1 - \chi_{(t-h,t+h)}(u_1 - u_2)$ and vice versa, and adding the ensuing inequalities we obtain

$$\int_{t-h}^{t+h} \langle D^\alpha_c u_1(s) - D^\alpha_c u_2(s), u_1(s) - u_2(s)\rangle \mathrm{d}s \leq 0,$$

meaning that $\langle D^\alpha_c u_1(t) - D^\alpha_c u_2(t), u_1(t) - u_2(t)\rangle \leq 0$ for almost every $t \in [0,T]$.

Define $d(t) = \|u_1(t) - u_2(t)\|^2$. Since $u_1, u_2 \in L^2(0,T;\mathcal{H})$ we clearly have $d \in L^1(0,T;\mathbb{R})$. Furthermore,

$$⨍_0^t |d(s)|\mathrm{d}s \leq 2 ⨍_0^t (\|(u_1(s) - u_0)\|^2 + \|(u_2(s) - u_0)\|^2)\mathrm{d}s \to 0,$$

as $t \downarrow 0$, from Theorem 4.2. Using (2.18) we then have

$$D_c^\alpha d(t) \le 2\langle D_c^\alpha u_1(t) - D_c^\alpha u_2(t), u_1(t) - u_2(t)\rangle \le 0$$

in the distributional sense. Combining with the facts that $d \ge 0$ and $\fint_0^t |d(s)|\mathrm{d}s \to 0$ we obtain, by Corollary 3.8 of Ref. 36, $d(t) = 0$. This proves the uniqueness.

We now turn our attention to existence. Let $\{\mathcal{P}_k\}_{k=1}^\infty$ be a sequence of partitions such that $\tau_k \downarrow 0$ as $k \to \infty$. We denote by $\mathbf{U}^{(k)}$ the discrete solution, on partition $\mathcal{P}_k$, given by (4.7) with $U_0^{(k)} = u_0$. The symbols $\widehat{U}_k$, $\overline{V}_k$ and $\overline{F}_k$ carry analogous meaning. Owing to Theorem 4.9 there exists $u \in C([0,T];\mathcal{H})$ such that $\widehat{U}_k$ converges to u in $C([0,T];\mathcal{H})$.

The embedding of Theorem 2.3 and an application of Theorem 4.6 shows that there is a subsequence for which $\overline{V}_{k_j} \rightharpoonup v$ in $L^2(0,T;\mathcal{H})$ as $j \to \infty$. Moreover, we can again appeal to Theorem 4.6 to see that, for every $t \in [0,T]$, the sequence

$$(t-\cdot)^{\frac{\alpha-1}{2}}\overline{V}_{k_j}(\cdot)$$

is uniformly bounded in $L^2(0,t;\mathcal{H})$ so that by passing to a further, not retagged, subsequence

$$(t-\cdot)^{\frac{\alpha-1}{2}}\overline{V}_{k_j}(\cdot) \rightharpoonup (t-\cdot)^{\frac{\alpha-1}{2}}v(\cdot) \quad \text{in } L^2(0,t;\mathcal{H}) \tag{4.18}$$

for any $t \in [0,T]$. This, in addition, shows that $v \in L^2_\alpha(0,T;\mathcal{H})$ so that if we define

$$\widetilde{u}(t) = u_0 + \frac{1}{\Gamma(\alpha)}\int_0^t (t-s)^{\alpha-1}v(s)\mathrm{d}s \tag{4.19}$$

then $D_c^\alpha \widetilde{u} = v$.

Recall that for any $j \in \mathbb{N}$ and any $t \in [0,T]$ we have that

$$\widehat{U}_{k_j}(t) = u_0 + \frac{1}{\Gamma(\alpha)}\int_0^t (t-s)^{\alpha-1}\overline{V}_{k_j}(s)\mathrm{d}s.$$

Since, for an arbitrary $w \in \mathcal{H}$ we have that $(t-\cdot)^{\frac{\alpha-1}{2}}w$ is in $L^2(0,t;\mathcal{H})$, we can use (4.18) to obtain that

$$\begin{aligned}\lim_{j\to\infty}\langle \widehat{U}_{k_j}(t), w\rangle &= \lim_{j\to\infty}\left\langle u_0 + \frac{1}{\Gamma(\alpha)}\int_0^t (t-s)^{\alpha-1}\overline{V}_{k_j}(s)\mathrm{d}s, w\right\rangle \\ &= \left\langle u_0 + \frac{1}{\Gamma(\alpha)}\int_0^t (t-s)^{\alpha-1}v(s)\mathrm{d}s, w\right\rangle = \langle \widetilde{u}(t), w\rangle.\end{aligned}$$

The statement above holds for any $w \in \mathcal{H}$ and all $t \in [0,T]$. Thus,

$$\widehat{U}_{k_j}(t) \rightharpoonup \widetilde{u}(t), \tag{4.20}$$

in $\mathcal{H}$. However, this implies that $\widetilde{u} = u$, as $\widehat{U}_{k_j}$ converges to u in $C([0,T];\mathcal{H})$. Therefore $D_c^\alpha u = v \in L^2_\alpha(0,T;\mathcal{H})$ and, by Theorem 4.6, we have the estimate

$$\|v\|_{L^2_\alpha(0,T;\mathcal{H})} \le C(\|f\|^2_{L^2_\alpha(0,T;\mathcal{H})} + \Phi(U_0) - \Phi_{\inf})^{1/2},$$

for some constant C depending on α. As in the proof of Theorem 4.8 this implies that (4.6) holds. From this, we also see that the initial condition is attained in the required sense.

It remains to show that the EVI (4.2) holds for u. From the construction of discrete solutions, one derives that for any $w \in L^2(0,T;\mathcal{H})$

$$\int_0^T (\Phi(\widehat{U}_{k_j}(t)) - \Phi(w(t)))\mathrm{d}t \le \int_0^T \langle \overline{F}_{k_j}(t) - \overline{V}_{k_j}(t), \widehat{U}_{k_j}(t) - w(t)\rangle \mathrm{d}t. \tag{4.21}$$

We will pass to the limit in this inequality. For the right-hand side, it suffices to observe that $\widehat{U}_{k_j} \to u$ in $C([0,T];\mathcal{H})$, $\overline{V}_{k_j} \rightharpoonup v$ in $L^2(0,T;\mathcal{H})$ and $\overline{F}_{k_j} \to f$ in $L^2(0,T;\mathcal{H})$. Thus,

$$\int_0^T \langle \overline{F}_{k_j}(t) - \overline{V}_{k_j}(t), \widehat{U}_{k_j}(t) - w(t)\rangle \mathrm{d}t \to \int_0^T \langle f(t) - v(t), u(t) - w(t)\rangle \mathrm{d}t.$$

For the left-hand side, the uniform convergence of $\widehat{U}_{k_j}$ and the lower semicontinuity of Φ, give

$$\Phi(u(t)) \le \liminf_{j\to\infty} \Phi\left(\widehat{U}_{k_j}(t)\right),$$

and hence

$$\int_0^T \Phi(u(t)) - \Phi(w(t))\mathrm{d}t \le \int_0^T \langle f(t) - v(t), u(t) - w(t)\rangle \mathrm{d}t.$$

It remains to recall that $D_c^\alpha u = v \in L^2(0,T;\mathcal{H})$ to conclude that, according to Theorem 4.2, u is an energy solution. □

Remark 4.10. (Other notion of solution) The choice of $u \in L^2(0,T;\mathcal{H})$ and $D_c^\alpha u \in L^2(0,T;\mathcal{H})$ in Theorem 4.2 is to guarantee that (4.2) makes sense. It is also necessary in the proof of uniqueness. However, other choices of spaces are also possible. For example, one could consider the following definition instead of Theorem 4.2: $u \in L^\infty(0,T;\mathcal{H})$ is a solution to (1.2) if:

(i) $\lim_{t\downarrow 0} \fint_0^t \|u(s) - u_0\|\mathrm{d}s = 0$;
(ii) $D_c^\alpha u \in L^1(0,T;\mathcal{H})$; and
(iii) for any $w \in L^\infty(0,T;\mathcal{H})$,

$$\int_0^T [\langle D_c^\alpha u(t), u(t) - w(t)\rangle + \Phi(u(t)) - \Phi(w(t))]\mathrm{d}t \le \int_0^T \langle f(t), u(t) - w(t)\rangle \mathrm{d}t. \tag{4.22}$$

Theorem 4.5 also holds for this new definition. However, at least with our techniques, the requirements on the data $u_0 \in D(\Phi)$ and $f \in L^2_\alpha(0,T;\mathcal{H})$ do not change.

4.3. *Asymptotic behavior*

One important, and well-studied, property of the solution to a classical, integer order, gradient flow is its convergence, as time goes to infinity, to a state of equilibrium, i.e. of minimal energy. We refer the interested reader to Chap. III.5 in Ref. 12, Sec. 3.4 of Ref. 8, and Proposition 11.9 in Ref. 54. In addition we comment that, while this paper was under review, Ref. 25 appeared. Here the authors show that, whenever the energy is differentiable, a time fractional gradient flow can be rewritten as a classical gradient flow for a modified energy, one that takes into account the history of the solution. The results of that work could be extended to the case considered here, where the energy is convex but may not be differentiable. Since our solution u satisfies $-\partial\Phi(u) \ni D_c^\alpha u \in L^2_\alpha(0,T;\mathcal{H})$, the modified energy could be constructed in the same way as in Theorem 2 of Ref. 25. The decay of the energy can, in principle, be related to the decay of this modified energy. However, this observation is irrelevant to our investigation in this section.

Let us here study, for the case $f \equiv 0$, the asymptotic behavior of solutions to time fractional gradient flows. We first recall that Proposition 5.11 in Ref. 39 proves the following convergence result under the assumption that the energy Φ is uniformly convex.

Theorem 4.11. (Convergence to equilibrium I) *Assume that the energy Φ is l.s.c, with nonempty effective domain and there is $\mu > 0$ for which*

$$w \mapsto \Phi(w) - \frac{\mu}{2}\|w\|^2,$$

is convex. Let $u^* \in \mathcal{H}$ be the global minimizer of Φ. Then, the energy solution to (1.2) *with $f \equiv 0$, satisfies, for $t > 0$,*

$$\Phi(u(t)) - \Phi_{\inf} \le (\Phi(u_0) - \Phi_{\inf})E_\alpha(-2\mu t^\alpha), \quad \|u(t) - u^*\| \le \|u_0 - u^*\|E_\alpha(-\mu t^\alpha).$$

Let us extend this result by removing the uniform convexity assumption.

Theorem 4.12. (Convergence to equilibrium II) *Let $f \equiv 0$ and assume that the energy Φ is convex, l.s.c., and with nonempty effective domain. Choose $u^* \in \mathcal{H}$ such that*

$$\Phi_{\inf} = \Phi(u^*).$$

Note that u^* may not be unique. Let u be the energy solution to (1.2), *$\mathcal{P}$ be an arbitrary time partition, and $\{U_n\}_{n\ge 0}$ be the solution to* (4.7). *We have, for all $t \ge 0$ and $n \ge 0$,*

$$\|U_n - u^*\| \le \|U_0 - u^*\|, \quad \|u(t) - u^*\| \le \|u_0 - u^*\|.$$

Furthermore, there exists a constant C independent of u, Φ, α such that

$$\Phi(u(t)) - \Phi_{\inf} \le \min\{\Phi(u_0) - \Phi_{\inf}, (\Phi(u_0) - \Phi_{\inf})^{\frac{1}{2}}\,\|u_0 - u^*\|\Gamma(1-\alpha)^{-\frac{1}{2}}\,t^{-\frac{\alpha}{2}}\}. \quad (4.23)$$

This guarantees that $\lim_{t\to\infty}\Phi(u(t)) = \Phi_{\inf}$.

Proof. We split the proof in several steps.

First we note that, since $-\left(D^\alpha_{\mathcal{P}}\mathbf{U}\right)_n \in \partial\Phi(U_n)$, we use (3.20) and the definition of subdifferential set to obtain

$$\frac{1}{2}(D^\alpha_{\mathcal{P}}\|\mathbf{U}-u^*\|^2)_n \le \langle U_n - u^*, (D^\alpha_{\mathcal{P}}\mathbf{U})_n\rangle \le \Phi(u^*) - \Phi(U_n) \le 0. \tag{4.24}$$

Hence, from Theorem 3.7, we get

$$\|U_n - u^*\|^2 \le \|U_0 - u^*\|^2.$$

Take $U_0 = u_0$ and let $\tau \downarrow 0$, we immediately get $\|u(t)-u^*\| \le \|u_0 - u^*\|$ as well.

Let $E_n = \Phi(U_n) - \Phi_{\inf}$. From (4.12) we have

$$(D^\alpha_{\mathcal{P}}\mathbf{E})_n = (D^\alpha_{\mathcal{P}}\Phi(\mathbf{U}))_n \le -\|\partial\Phi(U_n)\|^2,$$

where by $\partial\Phi(U_n)$ we denote any element in this subdifferential. Since

$$\begin{aligned}\Phi(U_n) - \Phi_{\inf} &\le \langle \partial\Phi(U_n), U_n - u^*\rangle \le \|\partial\Phi(U_n)\|\|U_n - u^*\|\\ &\le \|\partial\Phi(U_n)\|\|U_0 - u^*\|,\end{aligned}$$

we have

$$\|\partial\Phi(U_n)\| \ge \frac{E_n}{\|U_0 - u^*\|}, \quad (D^\alpha_{\mathcal{P}}\mathbf{E})_n \le -\frac{E_n^2}{\|U_0-u^*\|^2} = -\nu E_n^2,$$

for $\nu = \|U_0 - u^*\|^{-2}$. Consider the following discrete system with $W_n \ge 0$

$$(D^\alpha_{\mathcal{P}}\mathbf{W})_n = -\nu W_n^2, \quad W(0) = E_0. \tag{4.25}$$

It is easy to show that there exists a unique solution satisfying $0 \le W_n \le E_0$. By the discrete comparison in Theorem 3.7, we have $E_n \le W_n$. Let the time step $\tau \downarrow 0$, we see that $\{W_n\}$ converges to the solution of the fractional ODE

$$D^\alpha_c w = -\nu w^2, \quad w(0) = E_0. \tag{4.26}$$

This can be checked easily because the right-hand side $w \mapsto -\nu w^2$ is Lipschitz for $w \in [0, E_0]$, or one could also view it as a time fractional gradient flow for $\Phi(w) = \frac{\nu|w|^3}{3}$ and refer to the proof of Theorem 4.5.

When letting $\tau \downarrow 0$, owing to Theorem 4.9 we showed in the proof of Theorem 4.5 that $\overline{U}_{\mathcal{P}}$ uniformly converges to u. Hence,

$$\Phi(u(t)) - \Phi_{\inf} \le \liminf_{\tau\downarrow 0, t_n \to t} \Phi(U_n^{(\mathcal{P})}) - \Phi_{\inf} \le \liminf_{\tau\downarrow 0, t_n\to t} W_n^{(\mathcal{P})} \le w(t).$$

Now recall that, owing to Theorem 7.1 of Ref. 66, we have

$$w(t) \le \min\{E_0,\ CE_0^{\frac{1}{2}}\nu^{-\frac{1}{2}}\Gamma(1-\alpha)^{-\frac{1}{2}}\, t^{-\frac{\alpha}{2}}\},$$

which implies the desired result (4.23). $\square$

Remark 4.13. (Rate of decay) The rate of decay for the energy obtained in (4.23) may not be optimal. Although Theorem 7.1 of Ref. 66 shows that the solution to (4.26) decays like $t^{\frac{-\alpha}{2}}$, we know that the energy cannot behave exactly like

$\Phi(u(t)) - \Phi_{\inf} \approx t^{\frac{-\alpha}{2}}$. To see this, we take the fractional integral of order α in (4.24) and obtain

$$\frac{1}{2}\|U_n - u^*\|^2 \le \frac{1}{2}\|U_0 - u^*\|^2 - \frac{1}{\Gamma(\alpha)}\int_0^{t_n} (t_n - s)^{\alpha-1}\left(\Phi(\overline{U}_{\mathcal{P}}(s)) - \Phi_{\inf}\right) \mathrm{d}s.$$

This implies that

$$\frac{1}{\Gamma(\alpha)}\int_0^{t_n} (t_n - s)^{\alpha-1}(\Phi(\overline{U}_{\mathcal{P}}(s)) - \Phi_{\inf})\mathrm{d}s \le \frac{1}{2}\|U_0 - u^*\|^2 < +\infty.$$

By passing to the limit, this also holds for the continuous solution. If $\Phi(u(t)) - \Phi_{\inf} \approx t^{\frac{-\alpha}{2}}$ for large t, then

$$\int_0^t (t-s)^{\alpha-1}\left(\Phi(u(t)) - \Phi_{\inf}\right) \mathrm{d}s \to +\infty,$$

as $t \to +\infty$, which cannot be true.

Using Theorem 3.9 we also have convergence of the energy for our discrete solution.

Theorem 4.14. (Convergence to equilibrium III) *Under the same assumptions of Theorem* 4.12, *we have for any partition* $\mathcal{P}$ *that*

$$\Phi(U_n) - \Phi_{\inf} \le \min\{\Phi(U_0) - \Phi_{\inf}, (\Phi(U_0) - \Phi_{\inf})^{\frac{1}{2}}\Gamma(\alpha+1)^{\frac{1}{2}}\|U_0 - u^*\|t_n^{\frac{-\alpha}{2}}\},$$

which implies that $\lim_{t_n \to \infty} \Phi(U_n) = \Phi_{\inf}$.

Proof. From the proof of Theorem 4.12, we see that for $E_n = \Phi(U_n) - \Phi_{\inf}$ we have $E_n \le W_n$, where $W_n \ge 0$ was defined in (4.25). To get an upper bound for W_n, we consider a new partition $\widetilde{\mathcal{P}}$ with only two points $\widetilde{t}_0 = 0, \widetilde{t}_1 = t_n$. By Theorem 3.9, we have

$$E_n \le \widetilde{W}_1,$$

where $\widetilde{W}_1 \ge 0$ satisfies

$$(D^{\alpha}_{\widetilde{\mathcal{P}}}\widetilde{\mathbf{W}})_1 = \frac{\Gamma(\alpha+1)}{t_n^{\alpha}}(\widetilde{W}_1 - W_0) = -\nu\widetilde{W}_1^2,$$

and $\nu = \|U_0 - u^*\|^{-2}$. Since $W_0 = E_0 \ge 0$, the equation for $\widetilde{W}_1$ immediately implies that

$$\widetilde{W}_1 \le W_0, \quad \nu\widetilde{W}_1^2 \le \frac{\Gamma(\alpha+1)}{t_n^{\alpha}}W_0.$$

Combining these inequalities with $E_n \le \widetilde{W}_1$ finishes the proof. □

We comment that, both in Theorems 4.12 and 4.14, without uniform convexity, the equilibrium u^* point may not be unique. Hence, we cannot state convergence of solutions to a particular equilibrium point. The best, and immediate, result that one can obtain is the weak convergence, up to subsequences, to an equilibrium point. For this reason, instead, let us show that the solution cannot oscillate too much for large times. This is obtained in several steps.

We begin with an enhanced stability result, whose proof is inspired by the techniques of Ref. 62.

Lemma 4.15. (Enhanced stability) *Assume that* $f \equiv 0$, *and that* Φ *is convex, l.s.c., and with nonempty effective domain. Then,* u *the energy solution to* (1.2) *satisfies*

$$\|D_c^{\frac{\alpha+1}{2}} u\|^2_{L^2(0,T;\mathcal{H})} \le \frac{1}{\sin(\frac{\alpha\pi}{2})}(\Phi(u_0) - \Phi_{\mathrm{inf}}). \tag{4.27}$$

Proof. The proof proceeds by obtaining estimates for discrete solutions defined via (4.7). Let $\mathcal{P}$ be a time partition. Using the convexity of Φ and the definition of subdifferential we have, for $t \in (t_{n-1}, t_n]$,

$$\Phi(U_n) - \Phi(U_{n-1}) \le -\langle (D^\alpha_{\mathcal{P}} \mathbf{U})_n, U_n - U_{n-1}\rangle = -\langle D^\alpha_c \widehat{U}_{\mathcal{P}}(t), U_n - U_{n-1}\rangle.$$

The inequality above can be rewritten as

$$\Phi(U_n) - \Phi(U_{n-1}) \le \int_{t_{n-1}}^{t_n} -\langle D^\alpha_c \widehat{U}_{\mathcal{P}}(t), \widehat{U}'_{\mathcal{P}}(t)\rangle \,\mathrm{d}t,$$

which leads, after summation from $n = 1$ to $n = N$, to

$$\Phi(U_N) - \Phi(U_0) \le -\int_0^T \langle D^\alpha_c \widehat{U}_{\mathcal{P}}(t), \widehat{U}'_{\mathcal{P}}(t)\rangle \,\mathrm{d}t.$$

The inequality (see, for example, Lemma 2.1 in Ref. 62, or Lemma 3.1 in Ref. 48 for a proof)

$$\begin{aligned}
&\frac{1}{\Gamma(\beta)} \int_0^T \int_0^t (t-s)^{\beta-1} \langle h(s), h(t)\rangle \mathrm{d}s\mathrm{d}t \\
&\qquad \ge \cos\left(\frac{\beta\pi}{2}\right) \int_0^T \left\| \frac{1}{\Gamma(\beta/2)} \int_0^t (t-s)^{\beta/2-1} h(s)\mathrm{d}s \right\|^2 \mathrm{d}t
\end{aligned}$$

can be used with $\beta = 1 - \alpha$ and $h(t) = \widehat{U}'_{\mathcal{P}}$ to obtain that

$$\begin{aligned}
&\sin\left(\frac{\alpha\pi}{2}\right) \left\| D_c^{\frac{\alpha+1}{2}} \widehat{U}_{\mathcal{P}} \right\|^2_{L^2(0,T;\mathcal{H})} \\
&\qquad \le \int_0^T \langle D^\alpha_c \widehat{U}_{\mathcal{P}}(t), \widehat{U}'_{\mathcal{P}}(t)\rangle \,\mathrm{d}t \le \Phi(U_0) - \Phi(U_N) \le \Phi(U_0) - \Phi_{\mathrm{inf}}.
\end{aligned}$$

Indeed, since $\widehat{U}_{\mathcal{P}}$ is absolutely continuous, all definitions of the Caputo derivative, in particular (1.1) are valid.

It is now easy to see that if we use this stability and pass to the limit in the proof of Theorem 4.5, then we obtain (4.27), as we intended to show. □

With this at hand we can bound the oscillation of the solution for large times.

Theorem 4.16. (Oscillation estimate) *In the setting of Theorem* 4.15, *let u be the solution to* (1.2). *For any $\varepsilon > 0$, there exists $T_0 > 0$ such that*

$$|u|_{C^{\alpha/2}(T_0,\infty;\mathcal{H})} \leq \varepsilon.$$

In other words, for any $t > T_0$ and $s > 0$, we have

$$\|u(t) - u(t+s)\| \leq \varepsilon s^{\alpha/2}.$$

Proof. Since Theorem 4.15 has shown that $D_c^{\frac{1+\alpha}{2}} u \in L^2(0,\infty;\mathcal{H})$, the solution to (1.2) exists for all times $t > 0$. Moreover, there is $T_1 > 0$ such that

$$\frac{1}{\Gamma(\frac{1-\alpha}{2})\alpha^{1/2}} \|D_c^{\frac{\alpha+1}{2}} u\|_{L^2(T_1,\infty;\mathcal{H})} \leq \frac{\varepsilon}{3}.$$

Notice now that there is a constant $C > 0$ such that, for every $t > T_1$, and all $s > 0$ we have

$$(t - T_1)^{\alpha-1} - (t - T_1 + s)^{\alpha-1} \leq C(t - T_1)^{\frac{\alpha}{2}-1} s^{\alpha/2}.$$

Since $\frac{\alpha}{2} - 1 < 0$ we can choose now $T_0 > T_1$ for which

$$\frac{C}{\Gamma(\frac{1-\alpha}{2})}(T_0 - T_1)^{\frac{\alpha}{2}-1} \int_0^{T_1} \|D_c^{\frac{\alpha+1}{2}} u(r)\| \mathrm{d}r \leq \frac{\varepsilon}{3}.$$

Let $t > T_0$. We have the representation

$$u(t) = u(0) + \frac{1}{\Gamma(\frac{1-\alpha}{2})} \int_0^t (t-r)^{\frac{\alpha-1}{2}} D_c^{\frac{1+\alpha}{2}} u(r) \mathrm{d}r.$$

We decompose $u(t+s) - u(t)$ into three parts

$$\begin{aligned} u(t+s) - u(t) = & \frac{1}{\Gamma(\frac{1-\alpha}{2})} \int_t^{t+s} (t+s-r)^{\frac{\alpha-1}{2}} D_c^{\frac{1+\alpha}{2}} u(r) \mathrm{d}r \\ & - \frac{1}{\Gamma(\frac{1-\alpha}{2})} \int_{T_1}^t \left((t-r)^{\frac{\alpha-1}{2}} - (t+s-r)^{\frac{\alpha-1}{2}} \right) D_c^{\frac{1+\alpha}{2}} u(r) \mathrm{d}r \\ & - \frac{1}{\Gamma(\frac{1-\alpha}{2})} \int_0^{T_1} \left((t-r)^{\frac{\alpha-1}{2}} - (t+s-r)^{\frac{\alpha-1}{2}} \right) D_c^{\frac{1+\alpha}{2}} u(r) \mathrm{d}r \\ = & \frac{1}{\Gamma(\frac{1-\alpha}{2})} \left(I_1 + I_2 + I_3 \right). \end{aligned}$$

Clearly, by the Cauchy–Schwarz inequality, we have

$$\begin{aligned} \|I_1\| & \leq \left(\int_t^{t+s} (t+s-r)^{\alpha-1} \mathrm{d}r \int_t^{t+s} \|D_c^{\frac{1+\alpha}{2}} u(r)\|^2 \mathrm{d}r \right)^{1/2} \\ & \leq \frac{s^{\alpha/2}}{\alpha^{1/2}} \|D_c^{\frac{1+\alpha}{2}} u\|_{L^2(t,t+s;\mathcal{H})} \end{aligned}$$

and

$$\|I_2\| \le \left(\int_{T_1}^t \left((t-r)^{\frac{\alpha-1}{2}} - (t+s-r)^{\frac{\alpha-1}{2}}\right)^2 \mathrm{d}r \int_{T_1}^t \|D_c^{\frac{1+\alpha}{2}} u(r)\|^2 \mathrm{d}r\right)^{1/2}$$

$$\le \left(\int_{T_1}^t (t-r)^{\alpha-1} - (t+s-r)^{\alpha-1} \mathrm{d}r\right)^{1/2} \|D_c^{\frac{\alpha+1}{2}} u\|_{L^2(T_1,t;\mathcal{H})}$$

$$\le \frac{s^{\alpha/2}}{\alpha^{1/2}} \|D_c^{\frac{\alpha+1}{2}} u\|_{L^2(T_1,t;\mathcal{H})}.$$

The choice of T_1 guarantees that

$$\frac{1}{\Gamma(\frac{1-\alpha}{2})}(\|I_1\| + \|I_2\|) \le \frac{2}{3}\varepsilon s^{\alpha/2}.$$

For I_3, notice that

$$\|I_3\| \le \left((T_0-T_1)^{\alpha-1} - (T_0-T_1+s)^{\alpha-1}\right) \int_0^{T_1} \|D_c^{\frac{\alpha+1}{2}} u(r)\| \mathrm{d}r$$

$$\le C(T_0-T_1)^{\frac{\alpha}{2}-1} s^{\frac{\alpha}{2}} \int_0^{T_1} \|D_c^{\frac{\alpha+1}{2}} u(r)\| \mathrm{d}r.$$

Thus, the choice of T_0 implies

$$\frac{1}{\Gamma(\frac{1-\alpha}{2})}\|I_3\| \le \frac{\varepsilon}{3} s^{\alpha/2}.$$

Combining the estimates for I_1, I_2, I_3 proves the result. □

5. Time Fractional Gradient Flows: Numerics

Since the existence of an energy solution was proved by a rather constructive approach, namely a fractional minimizing movements scheme, it makes sense to provide error analyses for this scheme. We will provide an *a priori* error estimate which, in light of the smoothness $u \in C^{0,\alpha/2}([0,T];\mathcal{H})$ proved in Theorem 4.5, is optimal. In addition, in the spirit of Ref. 51 we will provide an *a posteriori* error analysis.

5.1. *A priori error analysis*

The *a priori* error estimate reads as follows. We comment that this result gives us a better rate compared to Theorem 5.10 of Ref. 39.

Theorem 5.1. (*A priori* I) *Let u be the energy solution of* (1.2). *Given a partition $\mathcal{P}$, of maximal step size τ, let $\mathbf{U} \in \mathcal{H}^N$ be the discrete solution defined by* (4.7) *starting from $U_0 \in D(\Phi)$. Let $\widehat{U}_{\mathcal{P}}$ and $\overline{U}_{\mathcal{P}}$ be defined as in* (4.10) *and* (2.3), *respectively. Then we have,*

$$\|u - \widehat{U}_{\mathcal{P}}\|_{L^\infty(0,T;\mathcal{H})} \le \|u_0 - U_0\| + C\tau^{\alpha/2}(\|f\|^2_{L^2_\alpha(0,T;\mathcal{H})} + \Phi_0 - \Phi_{\inf})^{1/2}, \quad (5.1)$$

and

$$\sup_{t\in[0,T]} \int_0^t (t-s)^{\alpha-1}\rho(u(s),\overline{U}_{\mathcal{P}}(s))\mathrm{d}s$$
$$\leq \|u_0 - U_0\|^2 + C\tau^\alpha(\|f\|^2_{L^2_\alpha(0,T;\mathcal{H})} + \Phi_0 - \Phi_{\inf}), \tag{5.2}$$

where $\Phi_0 = \max\{\Phi(U_0), \Phi(u_0)\}$, *and the constant* C *depends only on* α.

Proof. The proof can be obtained by following the same procedure employed in the proof of Theorem 4.9. In the current situation, however, instead of comparing two discrete solutions we compare the exact and discrete ones. The only difference is that we allow $U_0 \neq u_0$ here, but this presents no essential difficulty. For brevity, we skip the details. □

5.2. *A posteriori error analysis*

Let us now provide an *a posteriori* error estimate between the discretization in (4.7) and the solution of (1.2). We will also show how, from this *a posteriori* error estimator, an *a priori* error estimate can be derived. Let us first introduce the *a posteriori* error estimator.

Definition 5.2. (Error estimator) Let $\mathcal{P}$ be a partition of $[0,T]$ as in (2.2), and $\mathbf{U} \in \mathcal{H}^N$ denote the discrete solution given by (4.7). We define the error estimator function as

$$\mathcal{E}_{\mathcal{P}}(t) = \mathcal{E}_{\mathcal{P},1}(t) + \mathcal{E}_{\mathcal{P},2}(t), \tag{5.3}$$

where

$$\mathcal{E}_{\mathcal{P},1}(t) = \langle D_c^\alpha \widehat{U}_{\mathcal{P}}(t) - \overline{F}_{\mathcal{P}}(t), \widehat{U}_{\mathcal{P}}(t) - \overline{U}_{\mathcal{P}}(t)\rangle, \quad \mathcal{E}_{\mathcal{P},2}(t) = \Phi(\widehat{U}_{\mathcal{P}}(t)) - \Phi(\overline{U}_{\mathcal{P}}(t)).$$

Notice that the quantity $\mathcal{E}_{\mathcal{P}}(t)$ is nonnegative because $\overline{F}_{\mathcal{P}}(t) - D_c^\alpha \widehat{U}_{\mathcal{P}}(t) = F_{n(t)} - (D_{\mathcal{P}}^\alpha \mathbf{U})_{n(t)} \in \partial\Phi(U_{n(t)}) = \partial\Phi(\overline{U}_{\mathcal{P}}(t))$. It is also, in principle, computable since it only depends on data, and the discrete solution $\mathbf{U}$. It is then a suitable candidate for an *a posteriori* error estimator.

The derivation of an *a posteriori* error estimate begins with the observation that, for any $w \in \mathcal{H}$, we have

$$\begin{aligned}
&\langle D_c^\alpha \widehat{U}_{\mathcal{P}}(t) - f(t), \widehat{U}_{\mathcal{P}}(t) - w\rangle + \Phi(\widehat{U}_{\mathcal{P}}(t)) - \Phi(w) \\
&\quad = \mathcal{E}_{\mathcal{P}}(t) + \langle \overline{F}_{\mathcal{P}}(t) - D_c^\alpha \widehat{U}_{\mathcal{P}}(t), w - \overline{U}_{\mathcal{P}}(t)\rangle + \Phi(\overline{U}_{\mathcal{P}}(t)) - \Phi(w) \\
&\qquad + \langle f(t) - \overline{F}_{\mathcal{P}}(t), w - \widehat{U}_{\mathcal{P}}(t)\rangle \\
&\quad \leq \mathcal{E}_{\mathcal{P}}(t) + \langle f(t) - \overline{F}_{\mathcal{P}}(t), w - \widehat{U}_{\mathcal{P}}(t)\rangle - \sigma(\overline{U}_{\mathcal{P}}(t); w).
\end{aligned} \tag{5.4}$$

In other words, the function $\widehat{U}_{\mathcal{P}}$ solves an EVI similar to (4.3) but with additional terms on the right-hand side. We can then compare the EVIs by a now standard approach, that is, set $w = u(t)$ in (5.4) and $w = \widehat{U}_{\mathcal{P}}(t)$ in (4.3), respectively, to see that

$$\begin{aligned} &\langle D_c^\alpha \left(\widehat{U}_{\mathcal{P}} - u\right)(t), \widehat{U}_{\mathcal{P}}(t) - u(t)\rangle + \sigma(\overline{U}_{\mathcal{P}}(t); u(t)) + \sigma(u(t); \widehat{U}_{\mathcal{P}}(t)) \\ &\quad \le \mathcal{E}_{\mathcal{P}}(t) + \langle f(t) - \overline{F}_{\mathcal{P}}(t), u(t) - \widehat{U}_{\mathcal{P}}(t)\rangle \end{aligned} \tag{5.5}$$

for almost every $t \in [0, T]$. Consider the following notions of error:

$$E = \left(\sup_{t\in[0,T]} \{E_{\mathcal{H}}^2(t) + E_\sigma^2(t)\}\right)^{1/2}, \quad E_{\mathcal{H}}(t) = \|u(t) - \widehat{U}_{\mathcal{P}}(t)\|,$$

$$E_\sigma(t) = \left(\frac{2}{\Gamma(\alpha)} \int_0^t (t-s)^{\alpha-1}[\sigma(u(s); \widehat{U}_{\mathcal{P}}(s)) + \sigma(\overline{U}_{\mathcal{P}}(s); u(s))]\mathrm{d}s\right)^{1/2}. \tag{5.6}$$

We have the following error estimate for E.

Theorem 5.3. (*A posteriori*) *Let u be the energy solution of* (1.2). *Let $\mathcal{P}$ be a partition of $[0, T]$ defined as in* (2.2) *and let $\mathbf{U} \in \mathcal{H}^N$ be the discrete solution given by* (4.7) *starting from $U_0 \in D(\Phi)$. Let E and $\mathcal{E}_{\mathcal{P}}$ be defined in* (5.6) *and* (5.3), *respectively, The following a posteriori error estimate holds*

$$E \le \left(\|u_0 - U_0\|^2 + \frac{2}{\Gamma(\alpha)}\|\mathcal{E}_{\mathcal{P}}\|_{L^1_\alpha(0,T;\mathcal{H})}\right)^{1/2} + \frac{2}{\Gamma(\alpha)}\|f - \overline{F}_{\mathcal{P}}\|_{L^1_\alpha(0,T;\mathcal{H})}. \tag{5.7}$$

Proof. From (2.18) we infer

$$\begin{aligned} \frac{1}{2} D_c^\alpha \|\widehat{U}_{\mathcal{P}} - u\|^2(t) &\le \langle D_c^\alpha(\widehat{U}_{\mathcal{P}} - u)(t), \widehat{U}_{\mathcal{P}}(t) - u(t)\rangle \\ &\le \mathcal{E}_{\mathcal{P}}(t) + \langle f(t) - \overline{F}_{\mathcal{P}}(t), u(t) - \widehat{U}_{\mathcal{P}}(t)\rangle \\ &\quad - \sigma(\overline{U}_{\mathcal{P}}(t); u(t)) - \sigma(u(t); \widehat{U}_{\mathcal{P}}(t)). \end{aligned}$$

The claimed *a posteriori* error estimate (5.7) follows from Theorem 2.8 by setting

$$\lambda = 0, \quad a(t) = \|(\widehat{U}_{\mathcal{P}} - u)(t)\|, \quad b(t) = 2(\sigma(\overline{U}_{\mathcal{P}}(t); u(t)) + \sigma(u(t); \widehat{U}_{\mathcal{P}}(t))),$$
$$c(t) = 2\mathcal{E}_{\mathcal{P}}(t), \quad d(t) = \|(f - \overline{F}_{\mathcal{P}})(t)\|. \qquad \square$$

5.3. *Rate of convergence*

Although we have already established an optimal *a priori* rate of convergence for our scheme in Theorem 5.1, in this section we study the sharpness of the *a posteriori* error estimator $\mathcal{E}_{\mathcal{P}}$ by obtaining the same convergence rates through it. We comment that neither in Theorem 5.1 nor in our discussion here, we require any relation between time steps. We will also consider some cases when the rate of convergence can be improved.

5.3.1. *Rate of convergence for energy solutions*

Let us now use the estimator $\mathcal{E}_{\mathcal{P}}$ to derive a convergence rate or order $\mathcal{O}(\tau^{\alpha/2})$ for the error E, defined in (5.6), when $f \in L^2_\alpha(0,T;\mathcal{H})$. Notice that such regularity *a priori* does not give any order of convergence for $\|f - \overline{F}_{\mathcal{P}}\|_{L^1_\alpha(0,T;\mathcal{H})}$ in (5.7). Observe also that the rate that we obtain is consistent with classical gradient flow theories, where an order $\mathcal{O}(\tau^{1/2})$ is proved provided that $u_0 \in D(\Phi)$ and $f \in L^2(0,T;\mathcal{H})$; see Sec. 3.2 in Ref. 51.

We first bound $\|\mathcal{E}_{\mathcal{P}}\|_{L^1_\alpha(0,T;\mathcal{H})}$.

Theorem 5.4. (Bound on $\|\mathcal{E}_{\mathcal{P}}\|_{L^1_\alpha(0,T;\mathcal{H})}$) *Under the assumption that* $U_0 \in D(\Phi)$, *the estimator* $\mathcal{E}_{\mathcal{P}}$, *defined in* (5.3), *satisfies*

$$\|\mathcal{E}_{\mathcal{P}}\|_{L^1_\alpha(0,T;\mathcal{H})} \leq C\tau^\alpha(\|f\|^2_{L^2_\alpha(0,T;\mathcal{H})} + \Phi(U_0) - \Phi_{\inf}), \tag{5.8}$$

where the constant C depends only on α.

Proof. We bound the contributions $\mathcal{E}_{\mathcal{P},1}$ and $\mathcal{E}_{\mathcal{P},2}$ separately. The bound of $\mathcal{E}_{\mathcal{P},1}$ follows without change that of the term II of (4.15) in Theorem 4.9. Thus,

$$\|\mathcal{E}_{\mathcal{P},1}\|_{L^1_\alpha(0,T;\mathcal{H})} \leq C\tau^\alpha(\|f\|^2_{L^2_\alpha(0,T;\mathcal{H})} + \Phi(U_0) - \Phi_{\inf}). \tag{5.9}$$

To bound $\mathcal{E}_{\mathcal{P},2}$, we recall the function $\widehat{\Phi}_{\mathcal{P}}$, defined in Theorem 4.7, and its properties. Define also $\overline{\Phi}_{\mathcal{P}}(t) = \Phi(\overline{U}_{\mathcal{P}}(t))$. We have

$$\begin{aligned}
\mathcal{E}_{\mathcal{P},2}(t) &= \Phi(\widehat{U}_{\mathcal{P}}(t)) - \Phi(\overline{U}_{\mathcal{P}}(t)) \leq \widehat{\Phi}_{\mathcal{P}}(t) - \overline{\Phi}_{\mathcal{P}}(t) \\
&= \frac{1}{\Gamma(\alpha)}\left(\int_0^t (t-s)^{\alpha-1} D_c^\alpha \widehat{\Phi}_{\mathcal{P}}(s)\mathrm{d}s \right.\\
&\quad \left. - \int_0^{\lceil t\rceil_{\mathcal{P}}} (\lceil t\rceil_{\mathcal{P}} - s)^{\alpha-1} D_c^\alpha \widehat{\Phi}_{\mathcal{P}}(s)\mathrm{d}s\right) \\
&= \frac{1}{\Gamma(\alpha)}\left(\int_0^t [(t-s)^{\alpha-1} - (\lceil t\rceil_{\mathcal{P}} - s)^{\alpha-1}] D_c^\alpha \widehat{\Phi}_{\mathcal{P}}(s)\mathrm{d}s \right.\\
&\quad \left. - \int_t^{\lceil t\rceil_{\mathcal{P}}} (\lceil t\rceil_{\mathcal{P}} - s)^{\alpha-1} D_c^\alpha \widehat{\Phi}_{\mathcal{P}}(s)\mathrm{d}s\right) \\
&\leq \frac{1}{4\Gamma(\alpha)} \int_0^t [(t-s)^{\alpha-1} - (\lceil t\rceil_{\mathcal{P}} - s)^{\alpha-1}] \|\overline{F}_{\mathcal{P}}(s)\|^2 \mathrm{d}s \\
&\quad - \frac{1}{\Gamma(\alpha)} \int_t^{\lceil t\rceil_{\mathcal{P}}} (\lceil t\rceil_{\mathcal{P}} - s)^{\alpha-1} D_c^\alpha \widehat{\Phi}_{\mathcal{P}}(s)\mathrm{d}s \\
&= \frac{1}{4\Gamma(\alpha)} \int_0^t [(t-s)^{\alpha-1} - (\lceil t\rceil_{\mathcal{P}} - s)^{\alpha-1}] \|\overline{F}_{\mathcal{P}}(s)\|^2 \mathrm{d}s \\
&\quad - \frac{1}{\Gamma(\alpha+1)} (\lceil t\rceil_{\mathcal{P}} - t)^\alpha D_c^\alpha \widehat{\Phi}_{\mathcal{P}}(t) \\
&= \mathrm{I}_1(t) - \mathrm{I}_2(t).
\end{aligned}$$

On the one hand, proceeding as in the proof of Theorem 2.6 we obtain

$$\sup_{r\in[0,T]}\int_0^r (r-t)^{\alpha-1}\mathrm{I}_1(t)\mathrm{d}t \le C_3\tau^\alpha \|\overline{F}_{\mathcal{P}}\|^2_{L^2_\alpha(0,T;\mathcal{H})}.$$

On the other hand, using

$$\begin{aligned}
-\mathrm{I}_2(t) &\le \frac{-1}{\Gamma(\alpha+1)}(\lceil t\rceil_{\mathcal{P}}-t)^\alpha\left(D_c^\alpha\widehat{\Phi}_{\mathcal{P}}(t)-\frac{1}{4}\left\|\overline{F}_{\mathcal{P}}(t)\right\|^2\right)\\
&\le \frac{\tau^\alpha}{\Gamma(\alpha+1)}\left(\frac{1}{4}\left\|\overline{F}_{\mathcal{P}}(t)\right\|^2 - D_c^\alpha\widehat{\Phi}_{\mathcal{P}}(t)\right)
\end{aligned}$$

we have for any $r\in[0,T]$ that

$$\begin{aligned}
&-\int_0^r (r-t)^{\alpha-1}\mathrm{I}_2(t)\mathrm{d}t\\
&\quad\le \frac{\tau^\alpha}{\Gamma(\alpha+1)}\int_0^r (r-t)^{\alpha-1}\left(\frac{1}{4}\left\|\overline{F}(t)\right\|^2 - D_c^\alpha\widehat{\Phi}_{\mathcal{P}}(t)\right)\mathrm{d}t\\
&\quad= \frac{\tau^\alpha}{4\Gamma(\alpha+1)}\int_0^r (r-t)^{\alpha-1}\|\overline{F}_{\mathcal{P}}(t)\|^2\mathrm{d}t - \frac{\tau^\alpha}{\alpha}(\widehat{\Phi}_{\mathcal{P}}(r)-\Phi(U_0))\\
&\quad\le \frac{\tau^\alpha}{4\Gamma(\alpha+1)}\|\overline{F}_{\mathcal{P}}\|^2_{L^2_\alpha(0,T;\mathcal{H})} + \frac{\tau^\alpha}{\alpha}(\Phi(U_0)-\Phi_{\mathrm{inf}}).
\end{aligned}$$

Therefore combining the estimates for I_1 and I_2 we have proved that

$$\sup_{r\in[0,T]}\int_0^r (r-t)^{\alpha-1}\mathcal{E}_{\mathcal{P},2}(t)\mathrm{d}t \le C_4\tau^\alpha(\|f\|^2_{L^2_\alpha(0,T;\mathcal{H})}+\Phi(U_0)-\Phi_{\mathrm{inf}}),$$

which together with (5.9) proves (5.8) because $\mathcal{E}_{\mathcal{P}}$ is nonnegative. □

We next take advantage of Theorem 2.5, and derive a rate for E without additional smoothness assumptions on the right-hand side f.

Theorem 5.5. (*A priori* II) *Let u be the energy solution of* (1.2). *Let $\mathcal{P}$ be a partition of* $[0,T]$ *defined as in* (2.2) *and* $\mathbf{U}\in\mathcal{H}^N$ *be the discrete solution given by* (4.7) *starting from* $U_0\in D(\Phi)$. *Let E be defined in* (5.6). *Then we have*

$$E\le \|u_0-U_0\| + C\tau^{\alpha/2}(\|f\|^2_{L^2_\alpha(0,T;\mathcal{H})}+\Phi(U_0)-\Phi_{\mathrm{inf}})^{1/2},$$

where the constant C depends only on α.

Proof. We follow closely the approach and notation in Theorem 4.9. Define

$$G(t)=\frac{1}{\Gamma(\alpha)}\int_0^t (t-s)^{\alpha-1}\left(f(s)-\overline{F}_{\mathcal{P}}(s)\right)\mathrm{d}s$$

and note that, by Theorem 2.5, G satisfies

$$\tau^{\alpha/2}\|G\|_{L^\infty(0,T;\mathcal{H})}+\|G\|_{L^2_\alpha(0,T;\mathcal{H})}\le C_1\tau^\alpha\|f\|_{L^2_\alpha(0,T;\mathcal{H})}, \tag{5.10}$$

where the constant depends only on α. Set $e = u - \widehat{U}_{\mathcal{P}}$ and note that (5.5) can be rewritten as

$$\begin{aligned}\langle D_c^\alpha (e-G)(t), (e-G)(t)\rangle &+ \sigma(\overline{U}_{\mathcal{P}}(t); u(t)) + \sigma(u(t); \widehat{U}_{\mathcal{P}}(t)) \\ &\le \mathcal{E}_{\mathcal{P}}(t) - \langle D_c^\alpha (e-G)(t), G(t)\rangle.\end{aligned}$$

Notice the resemblance with (4.17). We can thus proceed as in Theorem 4.9, and use Theorem 5.4, to deduce that, for some constant C, depending only on α

$$\|u - \widehat{U}_{\mathcal{P}} - G\|^2(t) + E_\sigma(t) \le \|u_0 - U_0\|^2 + C_3\tau^\alpha(\|f\|^2_{L^2_\alpha(0,T;\mathcal{H})} + \Phi(U_0) - \Phi_{\inf}).$$

Estimate (5.10) then implies the result. □

5.3.2. *Rate of convergence for smooth energies*

Let us show that, at least for smoother energies, it is possible to obtain a better rate of convergence. We will, essentially, assume that the energy is locally $C^{1+\beta}$ for $\beta \in (0,1]$. More specifically in this section we consider energies that satisfy the following. There exists $\beta \in (0,1]$ such that for every $R > 0$, there is a constant $C_{\beta,R} > 0$ for which, whenever $w_1, w_2 \in B_R$ and $\xi_1 \in \partial\Phi(w_1)$,

$$\Phi(w_2) - \Phi(w_1) - \langle \xi_1, w_2 - w_1\rangle \le C_{\beta,R}\|w_2 - w_1\|^{1+\beta}, \tag{5.11}$$

where B_R denotes the ball of radius R in $\mathcal{H}$. Notice that, by Theorem 4.8, all the discrete solutions $\widehat{U}_{\mathcal{P}}$ are uniformly bounded in $C([0,T];\mathcal{H})$. Thus, we can fix $\bar{R} > 0$ depending only on the data such that, for any partition $\mathcal{P}$ and all $t \in [0,T]$, $\widehat{U}_{\mathcal{P}}(t) \in B_{\bar{R}}$. Therefore, (5.11) implies that

$$\Phi(w_2) - \Phi(w_1) - \langle \xi_1, w_2 - w_1\rangle \le C_\beta\|w_2 - w_1\|^{1+\beta}, \tag{5.12}$$

for some constant $C_\beta = C_{\beta,\bar{R}}$ and all $w_1, w_2 \in \widehat{U}_{\mathcal{P}}([0,T])$, $\xi_1 \in \partial\Phi(w_1)$.

A particular example to which this situation applies is the following. Let $\mathcal{H} = \mathbb{R}^d$ and $\Phi(w) = \frac{1}{p}|w|^p$ with $p > 1$. In this case, (5.12) holds with $\beta = 1$ for $p \ge 2$ and $\beta = p-1$ for $p \in (1,2)$. For $p < 2$, to reach $\beta = 1$, we must assume that u and $\widehat{U}_{\mathcal{P}}$ stay uniformly away from zero. This example can, of course, be generalized.

In this setting, we have the following improved estimate for $\|\mathcal{E}_{\mathcal{P}}\|_{L^1_\alpha(0,T;\mathcal{H})}$.

Theorem 5.6. (Improved bound) *Assume that the energy* Φ *satisfies* (5.12). *Let* u *be the energy solution to* (1.2), *and denote by* $\mathcal{P}$ *a partition of* $[0,T]$ *defined as in* (2.2). *Denote by* $\widehat{U}_{\mathcal{P}}$ *the solution of* (4.7) *starting from* $U_0 \in D(\Phi)$. *In this setting, the estimator* $\mathcal{E}_{\mathcal{P}}$ *defined in* (5.3) *satisfies*

$$\begin{aligned}\|\mathcal{E}_{\mathcal{P}}\|_{L^1_\alpha(0,T;\mathcal{H})} &\le CT^{\alpha(1-\beta)/2}\tau^{\alpha(\beta+1)} \\ &\quad \times (\|f\|^2_{L^2_\alpha(0,T;\mathcal{H})} + \Phi(U_0) - \Phi_{\inf})^{(\beta+1)/2},\end{aligned} \tag{5.13}$$

for some constant C *that depends on* α, β, *and the problem data.*

Proof. Owing to (5.12), the estimator $\mathcal{E}_{\mathcal{P}}$ can be bounded from above by

$$\begin{aligned}\mathcal{E}_{\mathcal{P}}(t) &= \langle D_c^\alpha \widehat{U}_{\mathcal{P}}(t) - \overline{F}_{\mathcal{P}}(t), \widehat{U}_{\mathcal{P}}(t) - \overline{U}_{\mathcal{P}}(t)\rangle + \Phi(\widehat{U}_{\mathcal{P}}(t)) - \Phi(\overline{U}_{\mathcal{P}}(t)) \\ &\le C_\beta \|\widehat{U}_{\mathcal{P}}(t) - \overline{U}_{\mathcal{P}}(t)\|^{1+\beta}.\end{aligned}$$

Applying Theorem 2.6 with $p = 1+\beta$ we have

$$\begin{aligned}\|\mathcal{E}_{\mathcal{P}}\|_{L^1_\alpha(0,T;\mathcal{H})} &\le \sup_{r\in[0,T]} C_\beta \int_0^r (r-t)^{\alpha-1} \|\widehat{U}_{\mathcal{P}}(t) - \overline{U}_{\mathcal{P}}(t)\|^{1+\beta} \mathrm{d}t \\ &\le C\tau^{\alpha(1+\beta)} \|D_c^\alpha \widehat{U}\|^{1+\beta}_{L^{1+\beta}_\alpha(0,T;\mathcal{H})},\end{aligned}$$

for some constant C that depends on α, β and the problem data. Since $1+\beta \in (1,2]$, Theorem 4.6 and the embedding

$$\|w\|_{L^{1+\beta}_\alpha(0,T;\mathcal{H})} \le \|w\|_{L^2_\alpha(0,T;\mathcal{H})} \left(\frac{T^\alpha}{\alpha}\right)^{(1-\beta)/(2(1+\beta))},$$

imply that

$$\|D_c^\alpha \widehat{U}_{\mathcal{P}}\|^{1+\beta}_{L^{1+\beta}_\alpha(0,T;\mathcal{H})} \le C_2 T^{\alpha(1-\beta)/2} (\|f\|^2_{L^2_\alpha(0,T;\mathcal{H})} + \Phi(U_0) - \Phi_{\text{inf}})^{(1+\beta)/2},$$

and this implies the claim. □

Now, in order to obtain a convergence rate using (5.7), we still need to control $\|f - \overline{F}_{\mathcal{P}}\|_{L^1_\alpha(0,T;\mathcal{H})}$. To do so, we invoke inequality (2.5) and see that

$$\|f - \overline{F}_{\mathcal{P}}\|_{L^1_\alpha(0,T;\mathcal{H})} \le \left(\frac{q-1}{q\alpha-1}\right)^{(q-1)/q} T^{\alpha-1/q} \|f - \overline{F}_{\mathcal{P}}\|_{L^q(0,T;\mathcal{H})}$$

for $q > 1/\alpha$. Thus, if $f \in W^{\alpha(1+\beta)/2,q}(0,T;\mathcal{H})$, then we have

$$\|f - \overline{F}_{\mathcal{P}}\|_{L^q(0,T;\mathcal{H})} \le C\tau^{\alpha(1+\beta)/2} |f|_{W^{\alpha(1+\beta)/2,q}(0,T;\mathcal{H})}$$

and hence

$$\|f - \overline{F}_{\mathcal{P}}\|_{L^1_\alpha(0,T;\mathcal{H})} \le CT^{\alpha-1/q} \tau^{\alpha(1+\beta)/2} |f|_{W^{\alpha(1+\beta)/2,q}(0,T;\mathcal{H})} \tag{5.14}$$

for some constant C that depends on α and q. Combining this with Theorem 5.6, the following convergence rate is a direct consequence of Theorem 5.3.

Theorem 5.7. (Improved rate: smooth energies) *Assume that the energy* Φ *satisfies* (5.12). *Let* u *be the energy solution to* (1.2), *and denote by* $\mathcal{P}$ *a partition of* $[0,T]$ *defined as in* (2.2). *Denote by* $\widehat{U}_{\mathcal{P}}$ *the solution of* (4.7) *starting from* $U_0 \in D(\Phi)$. *In this setting, if there is* $q > 1/\alpha$ *for which* $f \in W^{\alpha(\beta+1)/2,q}(0,T;\mathcal{H})$ *then the error* E, *defined in* (5.6), *satisfies*

$$\begin{aligned}E \le \|u_0 - U_0\| + C\tau^{\alpha(\beta+1)/2} [&(\|f\|^2_{L^2_\alpha(0,T;\mathcal{H})} + \Phi(U_0) - \Phi_{\text{inf}})^{(\beta+1)/4} \\ &+ |f|_{W^{\alpha(\beta+1)/2,q}(0,T;\mathcal{H})}],\end{aligned}$$

where the constant C *depends on* α, β, q, T, *and the problem data.*

5.3.3. *Rate of convergence for linear problems*

Let us now show how for certain classes of linear problems an improved rate of convergence can be obtained. We first assume that we have a Gelfand triple,

$$\mathcal{V} \hookrightarrow \mathcal{H} \hookrightarrow \mathcal{V}'$$

and that

$$\Phi(w) = \begin{cases} \dfrac{1}{2}\mathfrak{a}(w,w), & w \in \mathcal{V}, \\ +\infty, & w \notin \mathcal{V}, \end{cases} \tag{5.15}$$

where $\mathfrak{a} : \mathcal{V} \times \mathcal{V} \to \mathbb{R}$ is a nonnegative, symmetric, bounded, and semicoercive bilinear form. In this setting, (4.1) becomes

$$\langle D_c^\alpha u, w\rangle + \mathfrak{a}(u,w) = \langle f, w\rangle, \quad \forall\, w \in \mathcal{V}.$$

Notice that the bilinear form induces an operator $\mathfrak{A} : \mathcal{V} \to \mathcal{V}'$ given by

$$\langle \mathfrak{A}v, w\rangle_{\mathcal{V},\mathcal{V}'} = \mathfrak{a}(v,w), \quad \forall\, v, w \in \mathcal{V},$$

which implies that, for almost every $t \in (0,T)$, we have a problem in $\mathcal{V}'$ which reads

$$D_c^\alpha u(t) + \mathfrak{A}u(t) = f(t).$$

So that, $u_0 \in D(\partial\Phi)$ is equivalent to $\mathfrak{A}u_0 \in \mathcal{H}$. The bilinear form $\mathfrak{a}$ also induces a semi-norm on $\mathcal{V}$

$$[w]_{\mathcal{V}} = \mathfrak{a}(w,w)^{1/2}.$$

We further assume that $f \in L^2_\alpha(0,T;[\cdot]_{\mathcal{V}})$. More essentially we also require $u_0 \in D(\partial\Phi)$.

The motivation for an improved rate of convergence is then the following, at this stage formal, calculation. From (2.18) we have

$$\begin{aligned} \frac{1}{2} D_c^\alpha \|\mathfrak{A}u(t)\|^2 &\leq \langle D_c^\alpha \mathfrak{A}u(t), \mathfrak{A}u(t)\rangle = \langle \mathfrak{A}u(t), \mathfrak{A}D_c^\alpha u(t)\rangle \\ &= \langle f(t) - D_c^\alpha u(t), \mathfrak{A}D_c^\alpha u(t)\rangle \\ &= \mathfrak{a}(f(t), D_c^\alpha u(t)) - [D_c^\alpha u(t)]^2_{\mathcal{V}} \\ &\leq [f(t)]_{\mathcal{V}}[D_c^\alpha u(t)]_{\mathcal{V}} - [D_c^\alpha u(t)]^2_{\mathcal{V}}, \end{aligned}$$

which then shows via (2.17) that

$$\begin{aligned} &\frac{\Gamma(\alpha)}{2}\|\mathfrak{A}u(t)\|^2 + \int_0^t (t-s)^{\alpha-1}[D_c^\alpha u(s)]^2_{\mathcal{V}}\,\mathrm{d}s \\ &\quad \leq \frac{\Gamma(\alpha)}{2}\|\mathfrak{A}u_0\|^2 + \left(\int_0^t (t-s)^{\alpha-1}[f(s)]^2_{\mathcal{V}}\,\mathrm{d}s\right)^{1/2} \\ &\qquad \times \left(\int_0^t (t-s)^{\alpha-1}[D_c^\alpha u(s)]^2_{\mathcal{V}}\,\mathrm{d}s\right)^{1/2}. \end{aligned}$$

This implies that

$$[D_c^\alpha u]^2_{L^2_\alpha(0,T;[\cdot]_\mathcal{V})} = \sup_{t\in[0,T]} \int_0^t (t-s)^{\alpha-1} \left[D_c^\alpha u(s)\right]^2_\mathcal{V} \,\mathrm{d}s \le \Gamma(\alpha)\|\mathfrak{A}u_0\|^2 + \|f\|^2_{L^2_\alpha(0,T;[\cdot]_\mathcal{V})},$$

which says that $D_c^\alpha u$ is uniformly bounded in $L^2_\alpha(0,T;[\cdot]_\mathcal{V})$.

To make these considerations rigorous, we consider the discrete problem (4.7), which in this case reduces to

$$(D_\mathcal{P}^\alpha \mathbf{U})_n + \mathfrak{A}U_n = F_n.$$

Then the computations can be followed verbatim to obtain that

$$\begin{aligned}
&\frac{\Gamma(\alpha)}{2}\|\mathfrak{A}\widehat{U}_\mathcal{P}(t)\|^2 + \int_0^t (t-s)^{\alpha-1}[D_c^\alpha \widehat{U}_\mathcal{P}(s)]^2 \mathrm{d}s \\
&\quad\le \frac{\Gamma(\alpha)}{2}\|\mathfrak{A}U_0\|^2 + \left(\int_0^t (t-s)^{\alpha-1}[D_c^\alpha \widehat{U}_\mathcal{P}(s)]^2 \mathrm{d}s\right)^{1/2} \\
&\qquad\times \left(\int_0^t (t-s)^{\alpha-1}[\overline{F}_\mathcal{P}(s)]^2 \mathrm{d}s\right)^{1/2}
\end{aligned}$$

and

$$\begin{aligned}
[D_c^\alpha \widehat{U}_\mathcal{P}]^2_{L^2_\alpha(0,T;[\cdot]_\mathcal{V})} &= \sup_{t\in[0,T]} \int_0^t (t-s)^{\alpha-1}[D_c^\alpha \widehat{U}_\mathcal{P}(s)]^2_\mathcal{V} \mathrm{d}s \\
&\le \Gamma(\alpha)\|\mathfrak{A}U_0\|^2 + \|\overline{F}_\mathcal{P}\|^2_{L^2_\alpha(0,T;[\cdot]_\mathcal{V})}. \qquad (5.16)
\end{aligned}$$

Similar to Theorem 2.4, we know that

$$\|\overline{F}_\mathcal{P}\|_{L^2_\alpha(0,T;[\cdot]_\mathcal{V})} \le C\|f\|_{L^2_\alpha(0,T;[\cdot]_\mathcal{V})}$$

and hence $D_c^\alpha \widehat{U}_\mathcal{P}$ is uniformly bounded $L^2_\alpha(0,T;[\cdot]_\mathcal{V})$.

With this additional regularity, we can obtain an improved rate of convergence. To see this, we will use that Φ is, essentially, quadratic to observe that in this case the error estimator, defined in (5.3) reduces to

$$\mathcal{E}_\mathcal{P} = \frac{1}{2}\mathfrak{a}(\widehat{U}_\mathcal{P} - \overline{U}_\mathcal{P}, \widehat{U}_\mathcal{P} - \overline{U}_\mathcal{P}) = \frac{1}{2}[\widehat{U}_\mathcal{P} - \overline{U}_\mathcal{P}]^2_\mathcal{V}. \qquad (5.17)$$

These ingredients together give us the following improved estimate.

Theorem 5.8. (Improved rate: linear problems) *Assume that the energy Φ is given by* (5.15), *that the initial data satisfies $\mathfrak{A}u_0 \in \mathcal{H}$, and that $f \in L^2_\alpha(0,T;[\cdot]_\mathcal{V})$. Let u be the energy solution to* (1.2), *and denote by $\mathcal{P}$ a partition of $[0,T]$ defined as in* (2.2). *Denote by $\widehat{U}_\mathcal{P}$ the solution to* (4.7) *starting from $U_0 \in \mathcal{H}$, such that $\mathfrak{A}U_0 \in \mathcal{H}$. In this setting, we have that*

$$\|\mathcal{E}_\mathcal{P}\|_{L^1_\alpha(0,T;\mathcal{H})} \le C\tau^{2\alpha}(\|\mathfrak{A}U_0\|^2 + \|f\|^2_{L^2_\alpha(0,T;[\cdot]_\mathcal{V})}), \qquad (5.18)$$

where the constant C depends only on α. This, immediately, implies that

$$E \leq \|u_0 - U_0\| + C\tau^\alpha(\|\mathfrak{A}U_0\| + \|f\|_{L^2_\alpha(0,T;[\cdot]_\mathcal{V})}) + \|f - \overline{F}_\mathcal{P}\|_{L^1_\alpha,\mathcal{H})},$$

so that if, in addition, we further have $f \in W^{\alpha,q}(0,T;\mathcal{H})$ for some $q > 1/\alpha$, then

$$E \leq \|u_0 - U_0\| + C\tau^\alpha(\|\mathfrak{A}U_0\| + \|f\|_{L^2_\alpha(0,T;[\cdot]_\mathcal{V})} + |f|_{W^{\alpha,q}(0,T;\mathcal{H})}), \tag{5.19}$$

where the constant C depends only on α, q and T.

Proof. Owing to Theorem 5.3 and Eq. (5.14), the convergence rate (5.19) follows directly from (5.18) in the same way as Theorem 5.7. We only need to prove (5.18) and bound $\|\mathcal{E}_\mathcal{P}\|_{L^1_\alpha(0,T;\mathcal{H})}$. Using (5.17), for every $r \in (0,T]$ we have

$$2\int_0^r (r-t)^{\alpha-1}\mathcal{E}_\mathcal{P}(t)\mathrm{d}t = \int_0^r (r-t)^{\alpha-1}[\widehat{U}_\mathcal{P} - \overline{U}_\mathcal{P}]^2_\mathcal{V}(t)\mathrm{d}t.$$

Now, we invoke Theorem 2.6 with $p = 2$ and the semi-norm $[\cdot]_\mathcal{V}$ to obtain that

$$\int_0^r (r-t)^{\alpha-1}[\widehat{U}_\mathcal{P} - \overline{U}_\mathcal{P}]^2_\mathcal{V}(t)\mathrm{d}t \leq C\tau^{2\alpha}[D^\alpha_c \widehat{U}_\mathcal{P}]^2_{L^2_\alpha(0,T;[\cdot]_\mathcal{V})}.$$

By (5.16), we have that $D^\alpha_c\widehat{U}_\mathcal{P} \in L^2_\alpha(0,T;[\cdot]_\mathcal{V})$ uniformly in $\mathcal{P}$ and thus arrive at

$$\int_0^r (r-t)^{\alpha-1}[\widehat{U}_\mathcal{P} - \overline{U}_\mathcal{P}]^2_\mathcal{V}(t)\mathrm{d}t \leq C\tau^{2\alpha}(\|\mathfrak{A}U_0\|^2 + \|f\|^2_{L^2_\alpha(0,T;[\cdot]_\mathcal{V})}).$$

This implies the desired bound

$$\|\mathcal{E}_\mathcal{P}\|_{L^1_\alpha(0,T;\mathcal{H})} \leq C\tau^{2\alpha}(\|\mathfrak{A}U_0\|^2 + \|f\|^2_{L^2_\alpha(0,T;[\cdot]_\mathcal{V})})$$

for $\|\mathcal{E}_\mathcal{P}\|_{L^1_\alpha(0,T;\mathcal{H})}$ and finishes the proof. □

6. Lipschitz Perturbations

In this section, inspired by the results of Ref. 3, we consider the analysis and approximation of a fractional gradient flow with a Lipschitz perturbation. Namely, we consider the following problem:

$$\begin{cases} D^\alpha_c u(t) + \partial\Phi(u(t)) + \Psi(t, u(t)) \ni f(t), & t \in (0,T], \\ u(0) = u_0. \end{cases} \tag{6.1}$$

We assume that the perturbation function $\Psi : (0,T] \times \mathcal{H} \to \mathcal{H}$ satisfies

(1) (Carathéodory) For every $w \in \mathcal{H}$ the mapping $t \mapsto \Psi(t,w)$ is strongly measurable on $(0,T)$ with values in $\mathcal{H}$. Moreover, there exists $\mathfrak{L} > 0$ such that for almost every $t \in (0,T)$ and every $w_1, w_2 \in \mathcal{H}$ we have

$$\|\Psi(t,w_1) - \Psi(t,w_2)\| \leq \mathfrak{L}\|w_1 - w_2\|.$$

(2) (Integrability) There is $w_0 \in L^2_\alpha(0,T;\mathcal{H})$ for which

$$t \mapsto \Psi(t, w_0(t)) \in L^2_\alpha(0,T;\mathcal{H}).$$

We immediately comment that our assumptions can fit the case where Φ is merely λ–convex. Moreover, these assumptions also guarantee the existence of $\psi \in L^2_\alpha(0,T;\mathbb{R})$ for which

$$\|\Psi(t,w)\| \leq \psi(t) + \mathfrak{L}\|w\|, \quad \forall\, w \in \mathcal{H}.$$

Consequently $w \mapsto \Psi(\cdot, w(\cdot))$ is Lipschitz continuous in $L^2_\alpha(0,T;\mathcal{H})$.

We introduce the notion of *energy solution* of (6.1).

Definition 6.1. (Energy solution) A function $u \in L^2(0,T;\mathcal{H})$ is an energy solution to (6.1) if

(i) (Initial condition)

$$\lim_{t\downarrow 0} \fint_0^t \|u(s) - u_0\|^2 \mathrm{d}s = 0.$$

(ii) (Regularity) $D^\alpha_c u \in L^2(0,T;\mathcal{H})$.

(iii) (Evolution) For almost every $t \in (0,T)$ we have

$$D^\alpha_c u(t) + \partial\Phi(u(t)) + \Psi(t,u(t)) \ni f(t).$$

Evidently, an energy solution to (6.1) satisfies, for almost every $t \in (0,T)$ and all $w \in \mathcal{H}$, the EVI

$$\begin{aligned} &\langle D^\alpha_c u(t), u(t) - w\rangle + \langle \Psi(t,u(t)), u(t) - w\rangle + \Phi(u(t)) - \Phi(w) \\ &\quad \leq \langle f(t), u(t) - w\rangle. \end{aligned} \tag{6.2}$$

6.1. *Existence, uniqueness, and stability*

Our main result in this direction is the following.

Theorem 6.2. (Well posedness) *Assume that the energy Φ is convex, l.s.c., and with nonempty effective domain. Assume the mapping Ψ satisfies conditions* (1) *and* (2) *stated above. Let $u_0 \in D(\Phi)$ and $f \in L^2_\alpha(0,T;\mathcal{H})$, then there is a unique energy solution to* (6.1) *in the sense of Theorem* 6.1. *Moreover, we have that this solution satisfies*

$$\|D^\alpha_c u\|_{L^2_\alpha(0,T;\mathcal{H})} \leq C,$$

where the constant depends only on the problem data α, T, u_0, f, Φ, and Ψ.

Proof. We begin by proving existence. We essentially follow the idea used for the classical ODEs. A similar argument was also used in the proof of Theorem 4.4 in Ref. 36.

For $w \in L^2_\alpha(0,T;\mathcal{H})$ we denote by $\mathfrak{S}(w) \in L^2_\alpha(0,T;\mathcal{H})$ the energy solution to

$$D^\alpha_c u(t) + \partial\Phi(u(t)) \ni f(t) - \Psi(t,w(t)), \quad a.e.\ t \in (0,T], \quad u(0) = u_0.$$

Our assumptions and the results of Theorem 4.5 guarantee that this mapping is well defined, and moreover, $\mathfrak{S}(w) \in L^\infty(0,T;\mathcal{H})$. We want to show that there exists a fixed point w such that $\mathfrak{S}(w) = w$. If $u_i = \mathfrak{S}(w_i)$ for $i = 1, 2$, then for almost every t we have

$$\frac{1}{2} D_c^\alpha \|u_1(t) - u_2(t)\|^2 \leq -\langle \Psi(t, w_1(t)) - \Psi(t, w_2(t)), u_1(t) - u_2(t) \rangle.$$

This readily implies that

$$\begin{aligned}\|u_1(t) - u_2(t)\|^2 &\leq \frac{\mathfrak{L}}{\Gamma(\alpha)} \int_0^t (t-s)^{\alpha-1} \|w_1(s) - w_2(s)\| \|u_1(s) - u_2(s)\| \mathrm{d}s \\ &\leq \frac{\mathfrak{L} \|u_1 - u_2\|_{L^\infty(0,t;\mathcal{H})}}{\Gamma(\alpha)} \int_0^t (t-s)^{\alpha-1} \|w_1(s) - w_2(s)\| \mathrm{d}s\end{aligned}$$

which as a consequence yields that, for every $t \in [0, T]$,

$$\|u_1 - u_2\|_{L^\infty(0,t;\mathcal{H})} \leq \frac{\mathfrak{L}}{\Gamma(\alpha)} \|w_1 - w_2\|_{L^1_\alpha(0,t;\mathcal{H})}.$$

We claim that by induction, we can further obtain the following stability result

$$\|\mathfrak{S}^n(w_1) - \mathfrak{S}^n(w_2)\|_{L^\infty(0,t;\mathcal{H})} \leq \frac{\mathfrak{L}^n \, t^{\alpha n}}{\Gamma(\alpha n + 1)} \|w_1 - w_2\|_{L^\infty(0,t;\mathcal{H})} \tag{6.3}$$

for any $t \in [0, T]$ and positive integer n. In fact, for $n = 1$, we simply have

$$\begin{aligned}\|u_1 - u_2\|_{L^\infty(0,t;\mathcal{H})} &\leq \frac{\mathfrak{L}}{\Gamma(\alpha)} \|w_1 - w_2\|_{L^1_\alpha(0,t;\mathcal{H})} \\ &\leq \frac{\mathfrak{L} \, t^\alpha}{\Gamma(\alpha + 1)} \|w_1 - w_2\|_{L^\infty(0,t;\mathcal{H})}.\end{aligned}$$

Furthermore, if (6.3) holds for $n = k$, then for $n = k + 1$

$$\begin{aligned}\|\mathfrak{S}^{k+1}(w_1) - \mathfrak{S}^{k+1}(w_2)\|_{L^\infty(0,t;\mathcal{H})} &\leq \frac{\mathfrak{L}}{\Gamma(\alpha)} \|\mathfrak{S}^k(w_1) - \mathfrak{S}^k(w_2)\|_{L^1_\alpha(0,t;\mathcal{H})} \\ &\leq \frac{\mathfrak{L}}{\Gamma(\alpha)} \sup_{0 \leq r \leq t} \int_0^r (r-s)^{\alpha-1} \\ &\quad \times \frac{\mathfrak{L}^k \, s^{\alpha k}}{\Gamma(\alpha k + 1)} \|w_1 - w_2\|_{L^\infty(0,t;\mathcal{H})} \mathrm{d}s \\ &= \frac{\mathfrak{L}^{k+1} t^{\alpha(k+1)}}{\Gamma(\alpha(k+1) + 1)} \|w_1 - w_2\|_{L^\infty(0,t;\mathcal{H})},\end{aligned}$$

which proves (6.3). Now consider $w_0 \in L^2_\alpha(0, T; \mathcal{H})$ and the sequence of functions defined via $w_n = \mathfrak{S}^n(w_0)$. It is easy to see that, for $n \geq 1$, we have $w_n \in L^\infty(0, T; \mathcal{H})$, and $\sum_{n=1}^\infty \|w_n - w_{n+1}\|_{L^\infty(0,T;\mathcal{H})}$ converges because

$$\sum_{n=0}^\infty \frac{\mathfrak{L}^n \, t^{\alpha n}}{\Gamma(\alpha n + 1)} = E_\alpha(\mathfrak{L} t^\alpha).$$

This shows that $w_n \to u$ in $L^\infty(0, T; \mathcal{H})$ for some u. Since $w_{n+1} = \mathfrak{S}(w_n)$, it follows immediately that $u = \mathfrak{S}(u)$. This proves the existence of solutions.

As for uniqueness, assume that we have two solutions u_1 and u_2, for almost every t, we have

$$\frac{1}{2} D_c^\alpha \|u_1(t) - u_2(t)\|^2 \leq -\langle \Psi(t, u_1(t)) - \Psi(t, u_2(t)), u_1(t) - u_2(t) \rangle$$
$$\leq \mathfrak{L}\, \|u_1(t) - u_2(t)\|^2.$$

Combining with the fact that $u_1(0) = u_2(0) = u_0$, one obtains that $\|u_1(t) - u_2(t)\|^2 = 0$ for almost every t, which proves uniqueness.

Finally, the estimate on the Caputo derivative trivially follows from the iteration scheme. We skip the details. □

For diversity in our arguments, we present an alternative proof. The arguments here are inspired by those of Theorem 5.1 in Ref. 3.

Proof. (**alternative proof of Theorem 6.2**) Let us, for $\mu > \mathfrak{L}^{1/\alpha}$, define

$$\|w\|_\mu^2 = \sup_{t \in [0,T]} e^{-\mu t} \int_0^t (t-s)^{\alpha-1} \|w(s)\|^2 \mathrm{d}s,$$

which by the obvious inequalities $e^{-\mu T} \leq e^{-\mu t} \leq 1$, defines an equivalent norm in $L^2_\alpha(0, T; \mathcal{H})$.

Let $\mathfrak{S} : L^2_\alpha(0, T; \mathcal{H}) \to L^2_\alpha(0, T; \mathcal{H})$ be as before. As shown, if $u_i = \mathfrak{S}(w_i)$ for $i = 1, 2$, then for every t we have

$$\|u_1(t) - u_2(t)\|^2 \leq \frac{\mathfrak{L}}{\Gamma(\alpha)} \int_0^t (t-s)^{\alpha-1} \|w_1(s) - w_2(s)\| \|u_1(s) - u_2(s)\| \mathrm{d}s,$$

which as a consequence yields that, for every $r \in [0, T]$,

$$e^{-\mu r} \int_0^r (r-t)^{\alpha-1} \|u_1(r) - u_2(r)\|^2 \mathrm{d}r \leq \frac{\mathfrak{L} e^{-\mu r}}{\Gamma(\alpha)} \mathrm{I}(r),$$

where

$$\mathrm{I}(r) = \int_0^r (r-t)^{\alpha-1} \int_0^t (t-s)^{\alpha-1} \|w_1(s) - w_2(s)\| \|u_1(s) - u_2(s)\| \mathrm{d}s \mathrm{d}t.$$

Obvious manipulations then yield

$$\mathrm{I}(r) \leq \|u_1 - u_2\|_\mu \|w_1 - w_2\|_\mu \int_0^r (r-t)^{\alpha-1} e^{\mu t} \mathrm{d}t,$$

which implies

$$e^{-\mu r} \int_0^r (r-t)^{\alpha-1} \|u_1(r) - u_2(r)\|^2 \mathrm{d}r \leq \frac{\mathfrak{L}}{\Gamma(\alpha)} \int_0^r (r-t)^{\alpha-1} e^{-\mu(r-t)} \mathrm{d}t$$
$$\leq \frac{\mathfrak{L}}{\mu^\alpha} < 1,$$

so that $\mathfrak{S}$ is a contraction with respect to the norm $\|\cdot\|_\mu$. We conclude then by invoking the contraction mapping principle. This unique fixed point, evidently, is an energy solution in the sense of Theorem 6.1.

Uniqueness and stability follow as before. □

6.2. *Discretization*

Let us now present the numerical scheme for problem (6.1). We follow the previous notations and conventions regarding discretization so that, for any partition $\mathcal{P}$ of $[0,T]$ defined as in (2.2), we can also consider the discrete solution defined recursively via

$$F_n - (D^\alpha_{\mathcal{P}}\mathbf{U})_n - \Psi_n(U_n) \in \partial\Phi(U_n), \tag{6.4}$$

where F_n is defined in (4.8) and $\Psi_n : \mathcal{H} \to \mathcal{H}$ is defined by

$$\Psi_n(w) = \fint_{t_{n-1}}^{t_n} \Psi(t,w)\mathrm{d}t.$$

Clearly, for every n, Ψ_n is Lipschitz continuous with Lipschitz constant $\mathfrak{L}$. Using the definition of $D^\alpha_{\mathcal{P}}$ in (3.3) and $\mathbf{K}^{-1}_{\mathcal{P},nn} = (\mathbf{K}_{\mathcal{P},nn})^{-1} = \Gamma(\alpha+1)\tau_n^{-\alpha}$, we can rewrite (6.4) as

$$\Gamma(\alpha+1)\tau_n^{-\alpha}U_n + \Psi_n(U_n) + \partial\Phi(U_n) \ni F_n - \sum_{i=0}^{n-1}\mathbf{K}^{-1}_{\mathcal{P},ni}U_i.$$

Hence the discrete scheme can be recursively well defined provided $\mathfrak{L}\tau^\alpha < \Gamma(\alpha+1)$. For this reason, moving forward, we will implicitly operate under this assumption.

It is possible to show that the discrete solutions in (6.4) satisfy

$$\|D^\alpha_c \widehat{U}_{\mathcal{P}}\|_{L^2_\alpha(0,T;\mathcal{H})} \le C, \tag{6.5}$$

with a constant that depends on problem data but is independent of the partition $\mathcal{P}$. To see this, we follow the arguments of either proof of Theorem 6.2, and realize that while the operator $\mathfrak{S}$ may depend on $\mathcal{P}$, the estimates that we obtain do not.

6.3. *Error estimates*

Let us now show how to derive error estimates for the problem with Lipschitz perturbation (6.1). We recall that the energy solution u to this problem satisfies (6.2). In addition, for simplicity, we will operate under the assumption that the perturbation does not depend explicitly on time, i.e. $\Psi(t,w) = \Psi(w)$ for all $w \in \mathcal{H}$. The general case only lengthens the discussion but brings nothing substantive to it, as the additional terms that appear can be controlled via arguments used to control terms of the form

$$f(t) - \overline{F}_{\mathcal{P}}(t).$$

Similar to the discussion before, we define the error estimator

$$\mathcal{E}_{\mathcal{P},\mathfrak{L}}(t) = \mathcal{E}_{\mathcal{P}}(t) + \langle \Psi(\overline{U}_{\mathcal{P}}(t)), \widehat{U}_{\mathcal{P}}(t) - \overline{U}_{\mathcal{P}}(t)\rangle, \tag{6.6}$$

which, as before, is nonnegative. In addition, for any $w \in \mathcal{H}$ we have

$$
\begin{aligned}
&\langle D_c^\alpha \widehat{U}_\mathcal{P}(t) + \Psi(\widehat{U}_\mathcal{P}(t)) - f(t), \widehat{U}_\mathcal{P}(t) - w\rangle + \Phi(\widehat{U}_\mathcal{P}(t)) - \Phi(w) \\
&\quad = \mathcal{E}_{\mathcal{P},\mathfrak{L}}(t) + \langle \overline{F}_\mathcal{P}(t) - \Psi(\overline{U}_\mathcal{P}(t)) - D_c^\alpha \widehat{U}_\mathcal{P}(t), w - \overline{U}_\mathcal{P}(t)\rangle + \Phi(\overline{U}_\mathcal{P}(t)) - \Phi(w) \\
&\qquad + \langle \Psi(\overline{U}_\mathcal{P}(t)) - \Psi(\widehat{U}_\mathcal{P}(t)) + f(t) - \overline{F}_\mathcal{P}(t), w - \widehat{U}_\mathcal{P}(t)\rangle \\
&\quad \le \mathcal{E}_{\mathcal{P},\mathfrak{L}}(t) + \langle \Psi(\overline{U}_\mathcal{P}(t)) - \Psi(\widehat{U}_\mathcal{P}(t)) + f(t) - \overline{F}_\mathcal{P}(t), w - \widehat{U}_\mathcal{P}(t)\rangle \\
&\qquad - \sigma(\overline{U}_\mathcal{P}(t); w).
\end{aligned}
$$

Setting $w = u(t)$ in the inequality above and setting $w = \widehat{U}(t)$ in (6.2) lead to

$$
\begin{aligned}
&\langle D_c^\alpha (\widehat{U}_\mathcal{P} - u)(t), \widehat{U}_\mathcal{P}(t) - u(t)\rangle + \sigma(\overline{U}_\mathcal{P}(t); u(t)) + \sigma(u(t); \widehat{U}_\mathcal{P}(t)) \\
&\quad \le \mathcal{E}_{\mathcal{P},\mathfrak{L}}(t) + \langle \Psi(\overline{U}_\mathcal{P}(t)) - \Psi(\widehat{U}_\mathcal{P}(t)) + f(t) - \overline{F}_\mathcal{P}(t), u(t) - \widehat{U}_\mathcal{P}(t)\rangle \\
&\qquad + \langle \Psi(\widehat{U}_\mathcal{P}(t)) - \Psi(u(t)), u(t) - \widehat{U}_\mathcal{P}(t)\rangle
\end{aligned} \tag{6.7}
$$

for almost every $t \in (0, T)$. This implies the following error estimates.

Theorem 6.3. (*A posteriori*: Lipschitz perturbations) *Let u be the unique energy solution of* (6.1). *Let $\mathcal{P}$ be a partition of* $[0, T]$ *defined as in* (2.2) *and let $\mathbf{U} \in \mathcal{H}^N$ be the discrete solution given by* (6.4) *starting from $U_0 \in D(\Phi)$. Let E and $\mathcal{E}_{\mathcal{P},\mathfrak{L}}$ be defined in* (5.6) *and* (6.6), *respectively, The following a posteriori error estimate holds*

$$
\begin{aligned}
E \le &\left(\|u_0 - U_0\|^2 + \frac{2}{\Gamma(\alpha)} \|\mathcal{E}_{\mathcal{P},\mathfrak{L}}\|_{L^1_\alpha(0,T;\mathcal{H})} \right)^{1/2} (E_\alpha(2\mathfrak{L}T^\alpha))^{1/2} \\
&+ \frac{2}{\Gamma(\alpha)} \left(\|f - \overline{F}_\mathcal{P}\|_{L^1_\alpha(0,T;\mathcal{H})} + \mathfrak{L}\|\overline{U}_\mathcal{P} - \widehat{U}_\mathcal{P}\|_{L^1_\alpha(0,T;\mathcal{H})} \right) E_\alpha(2\mathfrak{L}T^\alpha).
\end{aligned} \tag{6.8}
$$

Proof. We argue as in the proof of (5.3). To make formulas shorter we omit the coercivity terms. From (2.18) and (6.7) we infer

$$
\begin{aligned}
\frac{1}{2} D_c^\alpha \|\widehat{U}_\mathcal{P} - u\|^2(t) &\le \left\langle D_c^\alpha (\widehat{U}_\mathcal{P} - u\right)(t), \widehat{U}_\mathcal{P}(t) - u(t)\rangle \\
&\le \mathcal{E}_{\mathcal{P},\mathfrak{L}}(t) + \langle \Psi(t, \overline{U}_\mathcal{P}(t)) - \Psi(t, \widehat{U}_\mathcal{P}(t)) \\
&\quad + f(t) - \overline{F}_\mathcal{P}(t), u(t) - \widehat{U}_\mathcal{P}(t)\rangle + \mathfrak{L}\|\widehat{U}_\mathcal{P}(t) - u(t)\|^2(t) \\
&\le \mathcal{E}_{\mathcal{P},\mathfrak{L}}(t) + (\mathfrak{L}\|\overline{U}_\mathcal{P}(t) - \widehat{U}_\mathcal{P}(t)\| + \|f(t) - \overline{F}_\mathcal{P}(t)\|) \\
&\quad \times \|\widehat{U}_\mathcal{P}(t) - u(t)\| + \mathfrak{L}\|\widehat{U}_\mathcal{P}(t) - u(t)\|^2.
\end{aligned} \tag{6.9}
$$

Then the error estimate (6.8) follows from Theorem 2.8 with

$$
\lambda = \mathfrak{L}, \quad a(t) = \|(\widehat{U}_\mathcal{P} - u)(t)\|, \quad b = 0, \quad c = 2\mathcal{E}_{\mathcal{P},\mathfrak{L}}(t)
$$

and

$$
d(t) = \mathfrak{L}\|\overline{U}_\mathcal{P}(t) - \widehat{U}_\mathcal{P}(t)\| + \|(f - \overline{F}_\mathcal{P})(t)\|. \qquad \square
$$

We also comment here that by Theorem 2.6

$$\|\overline{U}_{\mathcal{P}} - \widehat{U}_{\mathcal{P}}\|_{L^1_\alpha(0,T;\mathcal{H})} \leq C\tau^\alpha \|D^\alpha_c \widehat{U}_{\mathcal{P}}\|_{L^1_\alpha(0,T;\mathcal{H})} \leq \frac{CT^{\alpha/2}}{\alpha^{1/2}} \tau^\alpha \|D^\alpha_c \widehat{U}_{\mathcal{P}}\|_{L^2_\alpha(0,T;\mathcal{H})},$$

where the constant C only depends on α. In addition, the norm on the right-hand side is bounded independently of the partition $\mathcal{P}$; see (6.5). Hence the convergence rates proved in Theorems 5.5 and 5.7 also hold for problems with a Lipschitz perturbation. Since the proofs are almost identical, we only state the theorems below without proofs.

Theorem 6.4. (Convergence rate: Lipschitz perturbations) *Let u be the energy solution of* (6.1)*. Let $\mathcal{P}$ be a partition of $[0,T]$ defined as in* (2.2) *and $\mathbf{U} \in \mathcal{H}^N$ be the discrete solution given by* (6.4) *starting from $U_0 \in D(\Phi)$. Let E be defined in* (5.6)*. Then we have*

$$E \leq \|u_0 - U_0\|(E_\alpha(2\mathfrak{L}T^\alpha))^{1/2} + C\tau^{\alpha/2}(\|f\|_{L^2_\alpha(0,T;\mathcal{H})} + \|D^\alpha_c \widehat{U}_{\mathcal{P}}\|_{L^2_\alpha(0,T;\mathcal{H})}),$$

where the constant C depends only on $\alpha, \mathfrak{L}$ and T, but not on $\mathcal{P}$.

Theorem 6.5. (Improved rate: smooth energies and Lipschitz perturbations) *Assume that the energy Φ satisfies* (5.12)*. Let u be the energy solution to* (6.1)*, and denote by $\mathcal{P}$ a partition of $[0,T]$ defined as in* (2.2)*. Denote by $\widehat{U}_{\mathcal{P}}$ the solution of* (6.4) *starting from $U_0 \in D(\Phi)$. In this setting, if there is $q > 1/\alpha$ for which $f \in W^{\alpha(\beta+1)/2,q}(0,T;\mathcal{H})$ then the error E, defined in* (5.6)*, satisfies*

$$\begin{aligned} E \leq{}& \|u_0 - U_0\|(E_\alpha(2\mathfrak{L}T^\alpha))^{1/2} + C_1\tau^\alpha \|D^\alpha_c \widehat{U}_{\mathcal{P}}\|_{L^2_\alpha(0,T;\mathcal{H})} \\ &+ C_2\tau^{\alpha(\beta+1)/2}[(\|f\|_{L^2_\alpha(0,T;\mathcal{H})} + \|D^\alpha_c \widehat{U}_{\mathcal{P}}\|_{L^2_\alpha(0,T;\mathcal{H})})^{(\beta+1)/2} \\ &+ |f|_{W^{\alpha(\beta+1)/2,q}(0,T;\mathcal{H})}], \end{aligned} \tag{6.10}$$

where the constants C_1 and C_2 depend only on $\alpha, \beta, q, \mathfrak{L}$, T, and the problem data, but are independent of $\mathcal{P}$.

Finally we consider the setting of Sec. 5.3.3 with a Lipschitz perturbation. Similar to (6.5), we can show that $\|D^\alpha_c \widehat{U}_{\mathcal{P}}\|_{L^2_\alpha(0,T;[\cdot]_{\mathcal{V}})}$ is bounded uniformly with respect to the partition $\mathcal{P}$. For this reason, an improved error estimate analogous to Theorem 5.8 can be proved in this case.

Theorem 6.6. (Improved rate: quadratic energies and Lipschitz perturbations) *Assume that the energy Φ is given by* (5.15)*, that the initial data satisfies $\mathfrak{A}u_0 \in \mathcal{H}$, and that $f \in L^2_\alpha(0,T;[\cdot]_{\mathcal{V}})$. Let u be the energy solution to* (6.1)*, and denote by $\mathcal{P}$ a partition of $[0,T]$ defined as in* (2.2)*. Denote by $\widehat{U}_{\mathcal{P}}$ the solution to* (6.4) *starting from $U_0 \in \mathcal{H}$, such that $\mathfrak{A}U_0 \in \mathcal{H}$. In this setting, we have that*

$$\begin{aligned} E \leq{}& \|u_0 - U_0\|(E_\alpha(2\mathfrak{L}T^\alpha))^{1/2} + C\|f - \overline{F}\|_{L^1_\alpha(0,T;\mathcal{H})} \\ &+ C\tau^\alpha(\|\mathfrak{A}U_0\| + \|f\|_{L^2_\alpha(0,T;[\cdot]_{\mathcal{V}})} + \|D^\alpha_c \widehat{U}_{\mathcal{P}}\|_{L^2_\alpha(0,T;[\cdot]_{\mathcal{V}})} \\ &+ \|D^\alpha_c \widehat{U}_{\mathcal{P}}\|_{L^2_\alpha(0,T;\mathcal{H})}), \end{aligned} \tag{6.11}$$

where the constant C depends only on $\alpha, \mathfrak{L}$ and T.

7. Applications

Here we present a, far from exhaustive, list of example problems to which all our developments can be applied. As a rule, we will present the problem, how to cast it into our framework, provide some references where such a problem has been studied before, and briefly describe how our results contribute to this problem.

As we mentioned above, the list that we present here is by no means exhaustive. In the case of PDEs, we only consider second-order equations on domains. Thus, for instance, the space-time fractional parabolic problem studied in Refs. 50 and 11 can be cast in our framework but we do not discuss it here. We refer the reader to, for instance, Ref. 40 for applications of this problem. Another omission is the case of equations on graphs; see, for instance, Ref. 46.

7.1. *Fractional differential equations with discontinuous right hand side*

Given some mapping $A : S \subset \mathbb{R} \to \mathbb{R}$ and $u_0 \in \mathbb{R}$, we consider the fractional differential equation: Find $u : [0, T] \to \mathbb{R}$ such that

$$D_c^\alpha u(t) + Au(t) = f(t), \quad t \in [0, T], \quad u(0) = u_0. \tag{7.1}$$

Most of the theory regarding existence and uniqueness of solutions to fractional differential equations of this form requires, see Chaps. 6 and 7 of Ref. 20, that the mapping

$$\mathbb{R}^2 \ni (t, u) \mapsto F(t, u) = f(t) - Au \in \mathbb{R} \tag{7.2}$$

is at least continuous. Consequently a discontinuous mapping A is not admitted by the theory. Here we show under which conditions we can cover this scenario.

Set $\mathcal{H} = \mathbb{R}$. Assume that S is nonempty, convex, and closed; Assume also that A is non-decreasing. Therefore, after possible modification on a Lebesgue null set, this mapping is l.s.c. According to Theorem 1.1 of Ref. 35, the function

$$\Phi(u) = \begin{cases} \displaystyle\int_w^u A v \mathrm{d}v, & u \in S, \\ +\infty, & u \notin S, \end{cases} \qquad w \in S$$

is absolutely continuous and convex. Moreover, $\partial\Phi(v) = \overline{A}v$, where $\overline{A}$ denotes the maximal monotone extension of A. It is in this sense then that (7.1) can be cast into the setting of (1.2).

Integer order ODEs with discontinuous F, see (7.2), have been studied in Refs. 8 and 57. For instance, in Sec. 4.4 of Ref. 57, classical applications of ODEs to systems with friction and others have been presented. Thus, our theory and numerics cover, for instance, the case of fractional order systems with friction.

7.2. *The subdiffusion equation*

The subdiffusion, or fractional heat, equation reads: Let $d \geq 1$, $\Omega \subset \mathbb{R}^d$ be a bounded domain, $u_0 : \Omega \to \mathbb{R}$, and $f : \Omega \times (0, T] \to \mathbb{R}$. We seek for $u : \bar{\Omega} \times [0, T] \to \mathbb{R}$ that solves

$$\begin{cases} D_c^\alpha u(x,t) - \Delta u(x,t) = f(x,t), & (x,t) \in \Omega \times (0,T), \\ u(x,t) = 0, & (x,t) \in \partial\Omega \times (0,T), \\ u(x,0) = u_0(x), & x \in \Omega, \end{cases} \tag{7.3}$$

where Δ is the Laplace operator. This problem has been studied in many instances, and we refer the reader to Refs. 47 and 30 for applications.

Let us show how this problem fits our framework. We set $\mathcal{H} = L^2(\Omega)$, so that

$$\Phi(v) = \frac{1}{2} \int_\Omega |\nabla v(x)|^2 \mathrm{d}x, \quad D(\Phi) = H_0^1(\Omega),$$

and

$$\partial\Phi(v) = -\Delta v, \quad D(\partial\Phi) = \{v \in D(\Phi) : \Delta v \in L^2(\Omega)\}.$$

We point out that this problem also fits the framework of Sec. 5.3.3 by setting $\mathcal{V} = H_0^1(\Omega)$ and

$$\mathfrak{a}(v,w) = \int_\Omega \nabla v(x) \cdot \nabla w(x) \mathrm{d}x, \quad \mathfrak{A}v = -\Delta v.$$

A variation using a more general, but symmetric, linear elliptic differential operator is immediate.

For this type of problems, we have developed an unconditionally stable time-stepping scheme over arbitrary time partitions. We developed a reliable *a posteriori* error estimator for such a scheme.

7.3. *Time fractional quasilinear parabolic problems*

We can generalize the subdiffusion equation to a nonlinear problem. Assume that $F : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ is a convex, Carathéodory function, that is continuously differentiable with respect to its second argument, and such that together with $\mathbf{A} = D_2 F$ (its derivative with respect to the second variable), satisfy so-called p-coercivity, and p-growth conditions. In other words, there is $p > 1$ for which,

$$F(x, \boldsymbol{\xi}) \geq \alpha_0 |\boldsymbol{\xi}|^p - \alpha_1, \quad |\mathbf{A}(x, \boldsymbol{\xi})| \leq \alpha_2 (1 + |\boldsymbol{\xi}|^{p-1}), \quad \forall\, x \in \Omega, \boldsymbol{\xi} \in \mathbb{R}^d.$$

Here $\alpha_i > 0$ for $i = 0, \ldots, 2$. The time fractional PDE

$$D_c^\alpha u(x,t) - \nabla \cdot \mathbf{A}(x, \nabla u(x,t)) = f(x,t), \quad (x,t) \in \Omega \times (0,T),$$

supplemented with suitable initial and boundary conditions can be cast as in (1.2) by setting $\mathcal{H} = L^2(\Omega)$

$$\Phi(v) = \int_\Omega F(x, \nabla v(x)) \mathrm{d}x, \quad D(\Phi) = L^2(\Omega) \cap W_0^{1,p}(\Omega).$$

Problems of this kind have been studied in Refs. 49, 27 and 63, see also Refs. 70, 71, 73 and 45. Notice that besides convexity and the p-growth conditions, no additional structure is assumed on F or $\mathbf{A}$. Nevertheless, we have provided a theory for solutions of this problem. In addition, as in the linear case, we have developed an unconditionally stable time-stepping scheme over arbitrary time partitions. We developed a reliable *a posteriori* error estimator for such a problem.

7.4. *Time fractional parabolic obstacle problem*

The subdiffusion equation, and its nonlinear variants, can be generalized to the nonsmooth case. Here we present but one possibility: a time fractional parabolic obstacle problem. With the same notation as before choose $g \in W^{1,p}(\Omega)$ such that $g \leq 0$ on $\partial\Omega$, and define

$$\mathfrak{K} = \{v \in W_0^{1,p}(\Omega) : v(x) \geq g(x) \text{ a.e. } x \in \Omega\}.$$

Then, we set $\mathcal{H} = L^2(\Omega)$ and

$$\Phi(v) = \int_\Omega F(x, \nabla v(x))\mathrm{d}x + I_{\mathfrak{K}}(v), \quad D(\Phi) = \mathfrak{K} \cap L^2(\Omega),$$

where

$$I_{\mathfrak{K}}(v) = \begin{cases} 0, & v \in \mathfrak{K}, \\ +\infty, & v \notin \mathfrak{K}, \end{cases}$$

is the indicator function of the admissible set $\mathfrak{K}$. Time fractional obstacle problems have appeared, for instance, in Refs. 49, 27 and 63.

To our knowledge, our work is the first to tackle time fractional obstacle problems, and their theory, in a variational and energy setting. The aforementioned references are concerned with viscosity solutions. In addition, all the numerical developments that we have presented here are new for this problem.

7.5. *The time fractional porous medium equation*

The porous medium equation, see Ref. [64], is a prototypical example of a degenerate parabolic equation. Time fractional versions of it have appeared, for instance, in Refs. 7, 68, 17 and 22. In its simplest version this problem reads

$$D_c^\alpha u(x,t) - \Delta\beta(u(x,t)) = f(x,t),$$

supplemented by suitable initial and boundary conditions. Usually $\beta(t) = t^m$ with $m > 1$, but a general monotone function is also admissible.

To fit this problem into our framework we consider $\mathcal{H} = H^{-1}(\Omega)$ endowed with the inner product

$$\langle v, w \rangle = \langle v, (-\Delta)^{-1} w \rangle_{H^{-1}, H_0^1}, \quad \forall\, v, w \in H^{-1}(\Omega).$$

Let now

$$\phi(t) = \int_0^t \beta(s)\mathrm{d}s,$$

and

$$\Phi(v) = \int_\Omega \phi(v(x))\mathrm{d}x, \quad D(\Phi) = \{v \in L^2(\Omega) : \phi(v) \in L^1(\Omega)\},$$

so that, since the ambient space is $\mathcal{H} = H^{-1}(\Omega)$,

$$\partial\Phi(v) = -\Delta\beta(v).$$

To our knowledge, the only reference that numerically treats this problem is Ref. 52, where the developments are confined to one spatial dimension. Once again, our time-stepping scheme is unconditionally stable, and we have provided a reliable *a posteriori* error estimator for it.

7.6. *Time fractional diffusion reaction equations*

Variants of the subdiffusion equation, like the diffusion reaction equations studied in Ref. 21, can be studied by allowing Lipschitz perturbations. For instance, the semilinear problem

$$D_c^\alpha u(x,t) - \Delta u(x,t) + g(u(x,t)) = f(x,t)$$

fits into our framework under the assumption that the function g is Lipschitz, or that it can be split into the sum of a convex function and a Lipschitz perturbation.

7.7. *Time fractional Allen–Cahn equations*

Finally, if $g = G'$, where G is of double well type, that is,

$$G(r) = \begin{cases} (r-1)^2, & r > 1, \\ \dfrac{1}{4}\left(1-r^2\right)^2, & |r| \le 1, \\ (r+1)^2, & r < -1, \end{cases} \tag{7.4}$$

then we obtain the time fractional Allen–Cahn equation, which has been studied in Refs. 62, 53, 43 and 42. Notice that we can also consider this equation with constraints by considering

$$\begin{aligned} \Phi(v) &= \frac{1}{2}\int_\Omega |\nabla v(x)|^2\mathrm{d}x + I_{\mathfrak{K}}(v), \\ \mathfrak{K} &= \left\{v \in L^2(\Omega) : v(x) \in [-1,1] \text{ a.e. } x \in \Omega\right\}, \\ G(r) &= -\frac{1}{2}r^2. \end{aligned}$$

8. Numerical Illustrations

In this section, we present some simple numerical examples aimed at illustrating, and extending, our theory. All the computations were done with an in-house code that was written in MATLAB©.

8.1. *Practical a posteriori estimators*

We begin by commenting that, unlike the *a posteriori* estimators for the classical gradient flow proposed in Ref. 51, our *a posteriori* estimator $\mathcal{E}_{\mathcal{P}}$ is not constant on each subinterval of our partition $\mathcal{P}$; see (5.3). Here we mention more computationally friendly alternatives, and their properties.

First, we define an estimator that is piecewise constant in time via

$$\mathcal{D}_{\mathcal{P}}(t) = \max_{s\in[\lfloor t\rfloor_{\mathcal{P}},\lceil t\rceil_{\mathcal{P}}]} \{\langle D_c^\alpha \widehat{U}_{\mathcal{P}}(s) - \overline{F}(s), \widehat{U}_{\mathcal{P}}(s) - \overline{U}_{\mathcal{P}}(s)\rangle + \Phi(\widehat{U}_{\mathcal{P}}(s)) - \Phi(\overline{U}_{\mathcal{P}}(s))\}.$$

This is clearly an upper bound for $\mathcal{E}_{\mathcal{P}}(t)$.

One may also consider the simpler indicator

$$\widetilde{\mathcal{E}}_{\mathcal{P},n} = \langle (D_{\mathcal{P}}^\alpha \mathbf{U})_n - F_n, U_{n-1} - U_n\rangle + \Phi(U_{n-1}) - \Phi(U_n), \quad n = 1,\dots,N. \tag{8.1}$$

Although it is not always true that $\mathcal{E}_{\mathcal{P}}(t) \le \widetilde{\mathcal{E}}_{\mathcal{P},n(t)}$, this indicator is convenient to use in practice and gives reasonable results. In fact, this is the one that we implemented in the numerical examples of Sec. 8.3.

8.2. *A linear example*

As a first simple example we consider the fractional heat equation (7.3) with $\Omega = (0,\pi)$. Since, in this domain, the operator $-\Delta$ has eigenvalues and eigenfunctions

$$\lambda_k = k^2, \quad \varphi_k(x) = \sin(kx), \quad k \in \mathbb{N},$$

the fractional heat equation (7.3) has the solution

$$u(x,t) = \sum_{k=1}^\infty u_{0,k} E_\alpha(-\lambda_k t^\alpha)\varphi_k(x)$$

provided that the initial condition u_0 has the representation

$$u_0(x) = \sum_{k=1}^\infty u_{0,k}\varphi_k(x).$$

Since our main focus here is on time discretization, we simply consider a spectral discretization for space and use $m = 100$ modes in our experiments. We set $T = 1$. To quantify the error, we measure both

$$e_{\text{end}} = \|u(T) - U_N\|_{L^2(\Omega)}, \quad e_{\text{inf}} = \max_i \|u(t_i) - U_i\|_{L^2(\Omega)},$$

where the latter is a proxy for the error $\|u - \widehat{U}_{\mathcal{P}}\|_{L^\infty(0,T;\mathcal{H})}$.

Table 1. Convergence rate for the approximation of (7.3) using scheme (4.7) over a uniform partition of size τ. As predicted by Theorem 5.5, the rate is $\mathcal{O}(\tau^{\frac{\alpha}{2}})$ for e_{inf}.

$\alpha = 0.3$				
τ	e_{inf}	rate	e_{end}	rate
5.00e-02	1.09e-02	—	1.71e-03	—
2.50e-02	9.74e-03	0.166	9.03e-04	0.921
1.25e-02	8.72e-03	0.159	4.70e-04	0.940
6.25e-03	7.84e-03	0.153	2.43e-04	0.954
3.13e-03	7.07e-03	0.150	1.25e-04	0.964
1.56e-03	6.37e-03	0.149	6.35e-05	0.971
7.81e-04	5.75e-03	0.149	3.23e-05	0.977
3.91e-04	5.18e-03	0.150	1.63e-05	0.982
1.95e-04	4.67e-03	0.150	8.25e-06	0.985
9.77e-05	4.21e-03	0.150	4.16e-06	0.988
$\alpha = 0.7$				
τ	e_{inf}	rate	e_{end}	rate
5.00e-02	2.72e-02	—	5.97e-03	—
2.50e-02	2.13e-02	0.350	3.03e-03	0.979
1.25e-02	1.67e-02	0.350	1.53e-03	0.988
6.25e-03	1.31e-02	0.350	7.68e-04	0.993
3.13e-03	1.03e-02	0.350	3.85e-04	0.996
1.56e-03	8.08e-03	0.350	1.93e-04	0.997
7.81e-04	6.34e-03	0.350	9.65e-05	0.998
3.91e-04	4.98e-03	0.350	4.83e-05	0.999
1.95e-04	3.90e-03	0.350	2.41e-05	0.999
9.77e-05	3.06e-03	0.351	1.21e-05	1.000

We first consider the case where the initial condition is such that $u_0 \in H_0^1(\Omega) = D(\Phi)$, but $u_0 \notin H_0^1(\Omega) \cap H^2(\Omega) = D(\Delta)$. An example of this is,

$$u_{0,k} = k^{-1.5+\delta}, \tag{8.2}$$

with $0 < \delta \ll 1$. For computations we set $\delta = 10^{-4}$. The results are summarized in Table 1.

From Table 1, we observe that $e_{\text{inf}} = \mathcal{O}(\tau^{\frac{\alpha}{2}})$ and $e_{\text{end}} = \mathcal{O}(\tau)$. The convergence rate for e_{inf} is consistent with that proved in Theorem 5.5. It seems that the convergence rate for e_{end} is better. This warrants further investigation.

We next consider an initial value $u_0 \in H_0^1(\Omega) \cap H^2(\Omega)$. Namely,

$$u_{0,k} = k^{-2.5+\delta},$$

with $0 < \delta \ll 1$. In this case, both the assumptions in Theorem 5.7 (for $\beta = 1$) and Theorem 5.8 are satisfied, so the convergence rate for e_{inf} must be $\mathcal{O}(\tau^{\alpha})$. This is consistent with the experiments in Table 2, where $e_{\text{inf}} = \mathcal{O}(\tau^{\alpha})$ and $e_{\text{end}} = \mathcal{O}(\tau)$, where we chose $\delta = 10^{-4}$.

Table 2. Convergence rate for the approximation of (7.3) using scheme (4.7) over a uniform partition of size τ. As predicted by Sec. 5.3.2, the rate is $\mathcal{O}(\tau^\alpha)$ for e_{inf}.

$\alpha = 0.3$				
τ	e_{inf}	rate	e_{end}	rate
5.00e-02	8.44e-03	—	1.64e-03	—
2.50e-02	6.88e-03	0.296	8.66e-04	0.919
1.25e-02	5.57e-03	0.305	4.52e-04	0.939
6.25e-03	4.50e-03	0.308	2.33e-04	0.953
3.13e-03	3.63e-03	0.307	1.20e-04	0.963
1.56e-03	2.94e-03	0.305	6.11e-05	0.971
7.81e-04	2.38e-03	0.303	3.10e-05	0.977
3.91e-04	1.93e-03	0.302	1.57e-05	0.981
1.95e-04	1.57e-03	0.301	7.94e-06	0.985
9.77e-05	1.27e-03	0.301	4.00e-06	0.988
$\alpha = 0.7$				
τ	e_{inf}	rate	e_{end}	rate
5.00e-02	1.05e-02	—	5.81e-03	—
2.50e-02	6.48e-03	0.702	2.95e-03	0.977
1.25e-02	3.99e-03	0.701	1.49e-03	0.987
6.25e-03	2.45e-03	0.700	7.49e-04	0.992
3.13e-03	1.51e-03	0.700	3.76e-04	0.995
1.56e-03	9.30e-04	0.700	1.88e-04	0.997
7.81e-04	5.73e-04	0.700	9.42e-05	0.998
3.91e-04	3.52e-04	0.700	4.71e-05	0.999
1.95e-04	2.17e-04	0.700	2.36e-05	0.999
9.77e-05	1.34e-04	0.700	1.18e-05	1.000

8.3. *Adaptive time stepping*

We now illustrate the use of the *a posteriori* error estimator $\mathcal{E}_{\mathcal{P}}$, given in (5.3) to drive the selection of the size of the time step. For a given tolerance ε we, at every step, choose the local time step τ_n to guarantee that

$$\frac{2T^\alpha}{\Gamma(\alpha+1)}\widetilde{\mathcal{E}}_{\mathcal{P},n} \le \varepsilon^2,$$

where $\widetilde{\mathcal{E}}_{\mathcal{P},n}$ is given in (8.1). Then, by Theorem 5.3, we expect that

$$\|u - \widehat{U}_{\mathcal{P}}\|_{L^\infty(0,T;\mathcal{H})} \le \varepsilon,$$

provided the approximation error $\|f - \overline{F}_{\mathcal{P}}\|_{L^1_\alpha(0,T;\mathcal{H})}$ is negligible. Notice that to drive the process we are using the simpler estimator $\widetilde{\mathcal{E}}_{\mathcal{P}}$; see the discussion in Sec. 8.1.

We consider the linear problem (7.3) with $\alpha = \frac{1}{2}$, u_0 given by (8.2) and set $\varepsilon = 10^{-2}$. Figure 2 shows the local time step $\tau(t)$ for $t \in [0,T]$. As expected, due to the weak singularity of u at $t = 0$, the time step must be rather small for small

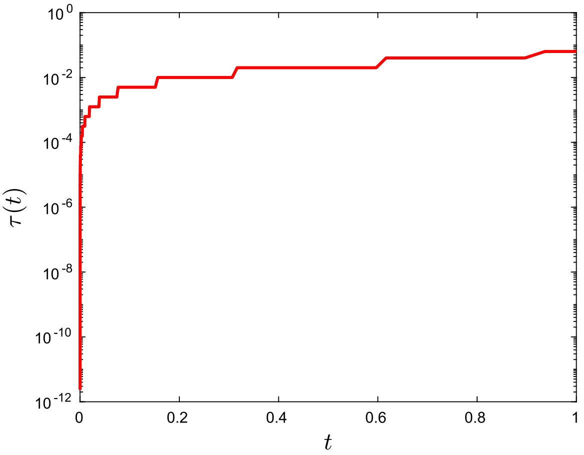

Fig. 2. Adaptive time stepping for problem (7.3) with $T = 1, \alpha = \frac{1}{2}$ is used to achieve a tolerance of $\varepsilon = 10^{-2}$. The adaptive solver uses 455 time intervals with minimum time step 2.328×10^{-12} and max time step 6.339×10^{-2}.

times. For larger times, however, the solution is smoother and larger local time steps can be taken. With this process we obtain that

$$\|u - \widehat{U}_{\mathcal{P}}\|_{L^\infty(0,T;\mathcal{H})} \approx 1.89 \times 10^{-3},$$

and this requires $N = 455$ time subintervals. For comparison, choosing a uniform time step of $\tau \approx 2.44 \times 10^{-5}$ we require $N = 40,960$ time intervals. This achieves an error of $e_{\text{inf}} \approx 3.0 \times 10^{-3}$, which is slightly higher than that obtained with our adaptive procedure. This clearly shows the advantages and possibilities for this strategy.

8.4. *The time fractional Allen–Cahn equation*

We now, depart from the linear theory and present several nonlinear examples with more complexity. We first examine a space-fractional variant of the time fractional Allen–Cahn equation mentioned in Sec. 7.7. To be specific, we consider the following equation:

$$D_c^\alpha u(x,t) + \lambda_1(-\Delta)^s u(x,t) + \lambda_2 g(u(x,t)) = 0 \tag{8.3}$$

for $x \in \Omega = (0,1)^d \subset \mathbb{R}^d$ and $t \in (0,T)$, where $g = G'$ with G defined in (7.4) and coefficients $\lambda_1, \lambda_2 > 0$. This is obtained by replacing the time and space derivatives in the classical model by their nonlocal counterparts. We impose periodic boundary conditions and consider $\mathcal{H} = L^2(\Omega)$,

$$\mathbb{H}^s(\Omega) = \{w \in L^2(\Omega) : |w|_{\mathbb{H}^s(\Omega)} < \infty\}, \tag{8.4}$$

where

$$|w|_{\mathbb{H}^s(\Omega)} = \left(\sum_{k\in\mathbb{Z}^d} |\widehat{w}_k|^2(2\pi|k|)^{2s}\right)^{1/2}, \quad \widehat{w}_k = \int_\Omega e^{-2\pi i k\cdot x} w(x)\mathrm{d}x.$$

Then this problem fits our framework with

$$\Phi(v) = \frac{\lambda}{2}|v|^2_{\mathbb{H}^s(\Omega)}, \quad \Psi(t,v) = g(v), \quad D(\Phi) = \mathbb{H}^s(\Omega).$$

To solve this problem numerically, we discretize in space by a collocation method. Let $M \in \mathbb{N}$ be the number of points in one spatial direction, $h = 1/M$ and introduce the grid domain and the space of grid functions

$$\Omega_M = \{x \in [0,1)^d \,|\, x_i \in h\mathbb{R}, 1 \le i \le d\}, \quad \mathbb{H}_M = \{v_M : \Omega_M \to \mathbb{R}\}.$$

Notice that $\mathbb{H}_M$ is a discretization of $\mathcal{H}$. To approximate the fractional Laplacian, we introduce the discrete Fourier transform. Let

$$\mathbb{Z}_M = \{r \in \mathbb{Z} \,|\, \lfloor (M-1)/2 \rfloor \le r \le \lceil (M-1)/2 \rceil\},$$

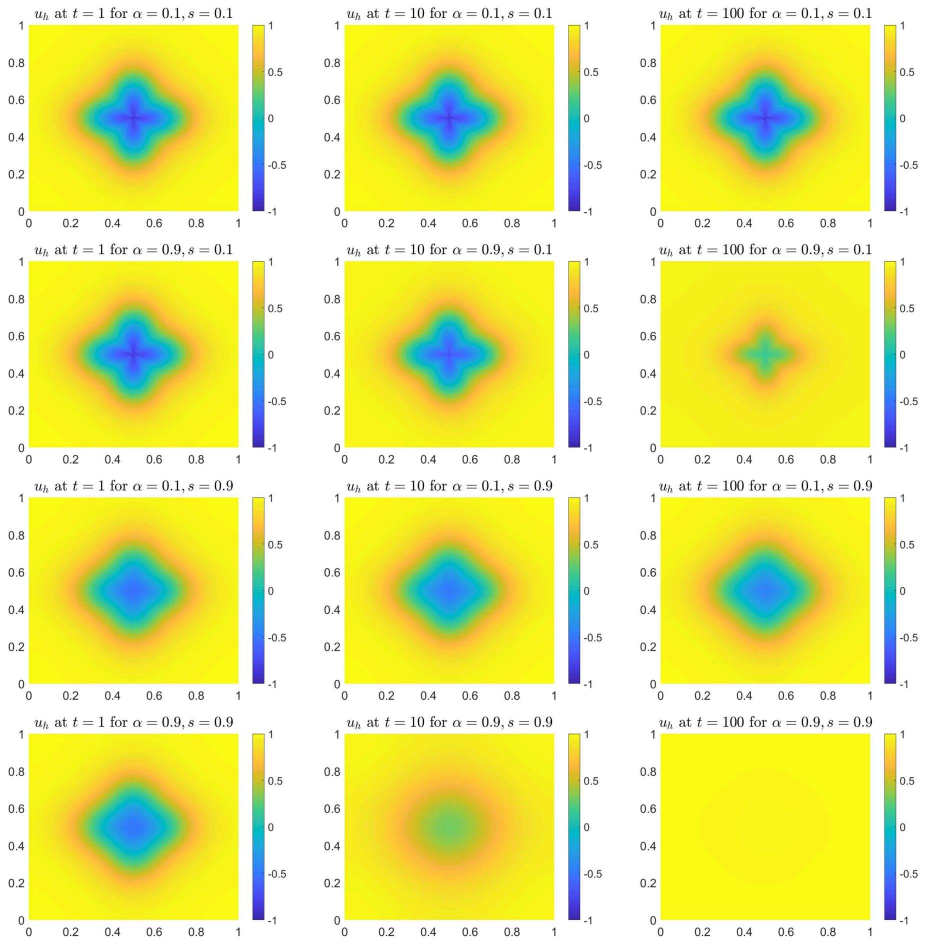

Fig. 3. Snapshots of discrete solutions of the time fractional Allen–Cahn equation (8.3) for $\alpha = 0.1, 0.9, s = 0.1, 0.9$.

and for $r \in \mathbb{Z}_M^d$, define

$$\widehat{w}_M(r) = h^d \sum_{x \in \Omega_M} w_M(x) e^{-2\pi i r \cdot x}.$$

Then the discrete fractional Laplacian $(-\Delta_M)^s$ is defined as

$$[(-\Delta_M)^s w_M](x) = \sum_{r \in \mathbb{Z}_M^2} (2\pi |r|)^{2s} \widehat{w}_M(r) e^{2\pi i r \cdot x}, \quad x \in \Omega_M.$$

To explore the behavior of the solutions, we consider an example similar to the one in Sec. 5.1 of Ref. 62. Let $d = 2$ and define the initial data

$$u_0 = \tanh\left(\frac{1}{\sqrt{2}\varepsilon_0}\left(2r - \frac{5}{16} - \frac{\cos(\theta)}{16}\right)\right),$$

where $\varepsilon_0 > 0$ and r, θ are polar coordinates centered at $(x_1 - 0.5, x_2 - 0.5)$, i.e.

$$x_1 = r\cos(\theta), \quad x_2 = r\sin(\theta), \quad r \geq 0, \ \ \theta \in [0, 2\pi),$$

Table 3. Convergence rate for the time fractional Allen–Cahn equation (8.3) over a uniform partition of size τ with $s = 0.999$. The rates for e_{inf} are close to $\mathcal{O}(\tau^{\alpha/2})$ predicted in Sec. 5.3.2.

$\alpha = 0.3$				
τ	e_{inf}	rate	e_{end}	rate
2.50e-03	9.11e-04	—	3.11e-04	—
1.25e-03	7.29e-04	0.321	1.72e-04	0.859
6.25e-04	5.81e-04	0.328	9.23e-05	0.894
3.13e-04	4.60e-04	0.336	4.88e-05	0.919
1.56e-04	3.65e-04	0.336	2.55e-05	0.938
7.81e-05	2.92e-04	0.320	1.32e-05	0.951
3.91e-05	2.39e-04	0.287	6.77e-06	0.961
1.95e-05	2.02e-04	0.244	3.46e-06	0.969
9.77e-06	1.76e-04	0.203	1.76e-06	0.975
$\alpha = 0.7$				
τ	e_{inf}	rate	e_{end}	rate
2.50e-03	7.24e-04	—	3.84e-04	—
1.25e-03	5.42e-04	0.416	1.97e-04	0.961
6.25e-04	4.04e-04	0.426	1.00e-04	0.977
3.13e-04	3.04e-04	0.411	5.06e-05	0.987
1.56e-04	2.33e-04	0.386	2.54e-05	0.992
7.81e-05	1.82e-04	0.355	1.28e-05	0.995
3.91e-05	1.45e-04	0.326	6.39e-06	0.997
1.95e-05	1.16e-04	0.322	3.20e-06	0.998
9.77e-06	9.01e-05	0.365	1.60e-06	0.999

for $x \in \Omega$. For this initial data, we have $u_0 \notin H^1(\Omega)$ and $u_0 \in H^{1-\delta}(\Omega)$ for any $\delta > 0$. For $\varepsilon_0 = 0.2$, we run experiments with $M = 64$ and uniform time step $\tau = 2^{-6}$ for different α and s. The parameters λ_1, λ_2 are set to

$$\lambda_1 = \begin{cases} \gamma\varepsilon^{2s-1}, & s \in (1/2, 1), \\ \gamma|\log\varepsilon|^{-1}, & s = 1/2, \\ \gamma, & s \in (0, 1/2), \end{cases} \qquad \lambda_2 = \lambda_1 \varepsilon^{-2s}, \tag{8.5}$$

with $\varepsilon = 0.2, \gamma = 0.01$. The scalings in (8.5) are informed in Ref. 55. Figure 3 shows snapshots of the phase field function at times $T = 1, 10, 100$ for different choices of α and s. We observe that a bigger α indicates faster convergence to the stable solution and larger s implies more smoothing in the phase field function u.

We also investigate the convergence rates with respect to the time step τ. We let $N^{(k)} = 2^{-k} \times 10$ and compute the discrete solutions $U^{(k)}$ on the uniform partitions of $(0, T)$ with $N^{(k)}$ intervals with step size $\tau^{(k)}$. Since the exact solution is unknown, to obtain an approximation of the convergence rates, we define

$$e_{\mathrm{end},k} = \|U^{(k-1)}_{N^{(k-1)}} - U^{(k)}_{N^{(k)}}\|_{L^2(\Omega)}, \quad e_{\mathrm{inf},k} = \max_{0 \le i \le N^{(k-1)}} \|U^{(k-1)}_i - U^{(k)}_{2i}\|_{L^2(\Omega)}$$

Table 4. Convergence rate for the time fractional Allen–Cahn equation (8.3) over a uniform partition of size τ with $s = 0.499$. The rates for e_{inf} are better than $\mathcal{O}(\tau^\alpha)$ predicted in Sec. 5.3.2.

$\alpha = 0.3$				
τ	e_{inf}	rate	e_{end}	rate
2.50e-03	4.78e-04	—	3.94e-04	—
1.25e-03	2.98e-04	0.680	2.17e-04	0.861
6.25e-04	1.88e-04	0.668	1.17e-04	0.895
3.13e-04	1.20e-04	0.651	6.17e-05	0.919
1.56e-04	7.72e-05	0.631	3.22e-05	0.938
7.81e-05	5.11e-05	0.596	1.67e-05	0.951
3.91e-05	3.46e-05	0.562	8.55e-06	0.961
1.95e-05	2.40e-05	0.529	4.37e-06	0.969
9.77e-06	1.70e-05	0.498	2.22e-06	0.976
$\alpha = 0.7$				
τ	e_{inf}	rate	e_{end}	rate
2.50e-03	1.05e-04	—	1.05e-04	—
1.25e-03	5.39e-05	0.967	5.39e-05	0.967
6.25e-04	2.74e-05	0.978	2.74e-05	0.978
3.13e-04	1.38e-05	0.986	1.38e-05	0.986
1.56e-04	6.95e-06	0.991	6.95e-06	0.991
7.81e-05	3.49e-06	0.995	3.49e-06	0.995
3.91e-05	1.75e-06	0.996	1.75e-06	0.996
1.95e-05	8.76e-07	0.998	8.76e-07	0.998
9.77e-06	4.38e-07	0.999	4.38e-07	0.999

and measure

$$\mathrm{rate}_{\mathrm{end},k} = \frac{\log(e_{\mathrm{end},k-1}) - \log(e_{\mathrm{end},k})}{\log(\tau^{(k-1)}) - \log(\tau^{(k)})}, \quad \mathrm{rate}_{\mathrm{inf},k} = \frac{\log(e_{\mathrm{inf},k-1}) - \log(e_{\mathrm{inf},k})}{\log(\tau^{(k-1)}) - \log(\tau^{(k)})}.$$

We set $T = 0.1, \lambda_1 = 0.1, \lambda_2 = 1, \varepsilon_0 = 1, M = 64$ and display the convergence rates for $\alpha = 0.3, 0.7$ and $s = 0.999, 0.499$ in Tables 3 and 4, respectively. For $s = 0.999$, we have $u_0 \in D(\Phi)$, but $u_0 \notin D(\partial\Phi)$. From Table 3, we observe that $e_{\mathrm{end}} \approx \mathcal{O}(\tau)$ and e_{inf} converges slightly better than the convergence rate $\mathcal{O}(\tau^{\alpha/2})$ proved in Theorem 6.4. For $s = 0.499$, the initial data $u_0 \in D(\Phi)$, and the assumptions in Theorem 6.5 (for $\beta = 1$) and Theorem 6.6 are satisfied. We observe in Table 4 better rates for e_{inf} than the theoretical rates $\mathcal{O}(\tau^{\alpha})$.

8.5. *A time fractional parabolic obstacle problem*

As a final example, we consider the space-fractional version of the parabolic obstacle problem presented in Sec. 7.4. We set $\Omega = (0,1)^d, \mathcal{H} = L^2(\Omega)$ and

$$\Phi(v) = \frac{1}{2}|v|^2_{\mathbb{H}^s(\Omega)} + I_{\mathfrak{K}}(v),$$

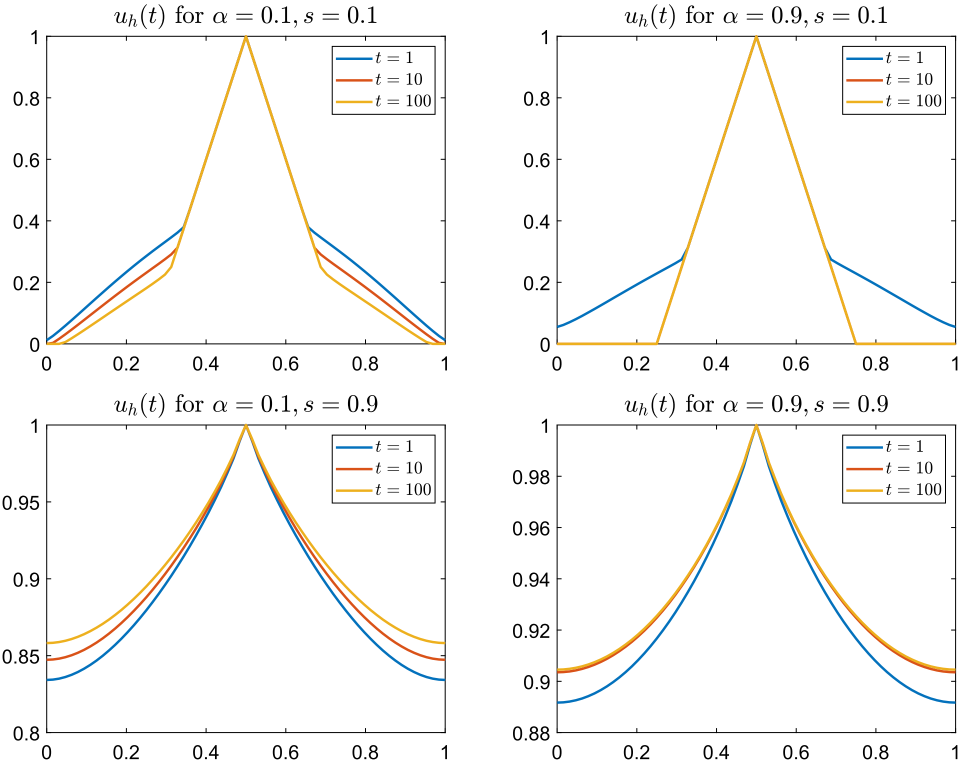

Fig. 4. Snapshots of discrete solutions of the space-time fractional parabolic obstacle problem for $\alpha = 0.1, 0.9, s = 0.1, 0.9$.

where $\mathbb{H}^s(\Omega)$ is defined in (8.4) and

$$\mathfrak{K} = \{v \in \mathbb{H}^s(\Omega) : v(x) \geq g(x) \text{ a.e. } x \in \Omega\}.$$

Consider $d = 1$ and an obstacle g, initial data u_0 and function $f(x,t)$ given by

$$g(x) = \max\{1 - 4|x - 0.5|, 0\}, \quad u_0(x) = \sin(\pi x), \quad f(x,t) = -0.5.$$

We use a collocation method as in Sec. 8.4, set $M = 64$, and a uniform time step $\tau = 2^{-6}$. The computed solutions u_h for different α, s at different times are presented in Fig. 4. As expected, we see that a bigger α indicates faster convergence to the stable solution and smaller s allows u to have sharp transitions and makes u closer to the obstacle.

Acknowledgments

The work of the authors is supported by NSF grant DMS-2111228. AJS is also partially supported by NSF grant DMS-1720213.

References

1. E. Affili and E. Valdinoci, Decay estimates for evolution equations with classical and fractional time-derivatives, *J. Differential Equations* **266** (2019) 4027–4060.
2. R. Agarwal and B. Ahmad, Existence theory for anti-periodic boundary value problems of fractional differential equations and inclusions, *Comput. Math. Appl.* **62** (2011) 1200–1214.
3. G. Akagi, Fractional flows driven by subdifferentials in Hilbert spaces, *Israel J. Math.* **234** (2019) 809–862.
4. M. Allen, Hölder regularity for nondivergence nonlocal parabolic equations, *Calc. Var. Partial Differential Equations* **57** (2018) 29.
5. M. Allen, A nondivergence parabolic problem with a fractional time derivative, *Differential Integral Equations* **31** (2018) 215–230.
6. M. Allen, L. Caffarelli and A. Vasseur, A parabolic problem with a fractional time derivative, *Arch. Ration. Mech. Anal.* **221** (2016) 603–630.
7. M. Allen, L. Caffarelli and A. Vasseur, Porous medium flow with both a fractional potential pressure and fractional time derivative, *Chin. Ann. Math. Ser. B* **38** (2017) 45–82.
8. J.-P. Aubin and A. Cellina, *Differential Inclusions: Set-valued Maps and Viability Theory*, Grundlehren der mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], Vol. 264 (Springer, 1984).
9. I. Benedetti, V. Obukhovskii and V. Taddei, On noncompact fractional order differential inclusions with generalized boundary condition and impulses in a Banach space, *J. Funct. Spaces* **2015** (2015) 10.
10. A. Bernardis, F. Martín-Reyes, P. Stinga and J. Torrea, Maximum principles, extension problem and inversion for nonlocal one-sided equations, *J. Differential Equations* **260** (2016) 6333–6362.
11. A. Bonito, W. Lei and J. Pasciak, Numerical approximation of space-time fractional parabolic equations, *Comput. Methods Appl. Math.* **17** (2017) 679–705.
12. H. Brézis, *Opérateurs Maximaux Monotones et Semi-Groupes de Contractions dans les Espaces de Hilbert*, North-Holland Mathematics Studies, Vol. 5 (North-Holland Publishing Co., 1973).

13. L. Caffarelli and L. Silvestre, An extension problem related to the fractional Laplacian, *Comm. Partial Differential Equations* **32** (2007) 1245–1260.
14. M. Caputo, Linear models of dissipation whose Q is almost frequency independent-II, *Geophys. J. R. Astron. Soc.* **13** (1967) 529–539.
15. A. Cernea, On a fractional differential inclusion arising from real estate asset securitization and HIV models, *Ann. Univ. Buchar. Math. Ser.* **4(LXII)** (2013) 447–453.
16. F. Clarke, *Optimization and Nonsmooth Analysis*, Classics in Applied Mathematics, Vol. 5, 2nd edn. (Society for Industrial and Applied Mathematics (SIAM), 1990).
17. E. Daus, M. Gualdani, J. Xu, N. Zamponi and X. Zhang, Non-local porous media equations with fractional time derivative, *Nonlinear Anal.* **211** (2021) 35.
18. B. de Andrade and T. Cruz, Regularity theory for a nonlinear fractional reaction–diffusion equation, *Nonlinear Anal.* **195** (2020) 14.
19. D. del Castillo-Negrete, Fractional diffusion models of nonlocal transport, *Phys. Plasmas* **13** (2006) 082308, 16.
20. K. Diethelm, *The Analysis of Fractional Differential Equations: An Application-Oriented Exposition Using Differential Operators of Caputo Type*, Lecture Notes in Mathematics, Vol. 2004 (Springer, 2010).
21. S. Dipierro, B. Pellacci, E. Valdinoci and G. Verzini, Time-fractional equations with reaction terms: Fundamental solutions and asymptotics, *Discrete Contin. Dyn. Syst.* **41** (2021) 257–275.
22. J.-D. Djida, J. Nieto and I. Area, Nonlocal time-porous medium equation: Weak solutions and finite speed of propagation, *Discrete Contin. Dyn. Syst. Ser. B* **24** (2019) 4031–4053.
23. X. Feng and M. Sutton, A new theory of fractional differential calculus, *Anal. Appl.* (*Singap.*) **19** (2021) 715–750.
24. Y. Feng, L. Li, J.-G. Liu and X. Xu, Continuous and discrete one dimensional autonomous fractional ODEs, *Discrete Contin. Dyn. Syst. Ser. B* **23** (2018) 3109–3135.
25. M. Fritz, U. Khristenko and B. Wohlmuth, Equivalence between a time-fractional and an integer-order gradient flow: The memory effect reflected in the energy, *Adv. Nonlinear Anal.* **12** (2023) 23.
26. I. Gel'fand and G. Šilov, *Obobshchennye funksii i deĭstviya iad nimi*, Obobščennye funkcii, Vypusk 1. (Gosudarstv. Izdat. Fiz.-Mat. Lit., Moscow, 1958), (in Russian).
27. Y. Giga, Q. Liu and H. Mitake, On a discrete scheme for time fractional fully nonlinear evolution equations, *Asymptot. Anal.* **120** (2020) 151–162.
28. R. Gorenflo, A. Kilbas, F. Mainardi and S. Rogosin, *Mittag–Leffler Functions, Related Topics and Applications*, Springer Monographs in Mathematics (Springer, 2014).
29. L. Grafakos, *Classical Fourier Analysis*, Graduate Texts in Mathematics, Vol. 249, 3rd edn. (Springer, 2014).
30. B. Jin, R. Lazarov and Z. Zhou, Numerical methods for time-fractional evolution equations with nonsmooth data: A concise overview, *Comput. Methods Appl. Mech. Eng.* **346** (2019) 332–358.
31. T. Ke, N. Thang and L. T. P. Thuy, Regularity and stability analysis for a class of semilinear nonlocal differential equations in Hilbert spaces, *J. Math. Anal. Appl.* **483** (2020) 23.
32. J. Kemppainen, J. Siljander, V. Vergara and R. Zacher, Decay estimates for time-fractional and other non-local in time subdiffusion equations in $\mathbb{R}^d$, *Math. Ann.* **366** (2016) 941–979.
33. M. Krasnoschok, V. Pata, S. Siryk and N. Vasylyeva, A subdiffusive Navier–Stokes–Voigt system, *Physica D* **409** (2020) 13.

34. M. Krasnoschok, V. Pata and N. Vasylyeva, Semilinear subdiffusion with memory in multidimensional domains, *Math. Nachr.* **292** (2019) 1490–1513.
35. M. A. Krasnosel'skiĭ and J. B. Rutickiĭ, *Convex Functions and Orlicz Spaces* (P. Noordhoff Ltd., Groningen, 1961), translated from the first Russian edition by Leo F. Boron.
36. L. Li and J.-G. Liu, A generalized definition of Caputo derivatives and its application to fractional ODEs, *SIAM J. Math. Anal.* **50** (2018) 2867–2900.
37. L. Li and J.-G. Liu, A note on deconvolution with completely monotone sequences and discrete fractional calculus, *Quart. Appl. Math.* **76** (2018) 189–198.
38. L. Li and J.-G. Liu, Some compactness criteria for weak solutions of time fractional PDEs, *SIAM J. Math. Anal.* **50** (2018) 3963–3995.
39. L. Li and J.-G. Liu, A discretization of Caputo derivatives with application to time fractional SDEs and gradient flows, *SIAM J. Numer. Anal.* **57** (2019) 2095–2120.
40. L. Li, J.-G. Liu and L. Wang, Cauchy problems for Keller–Segel type time-space fractional diffusion equation, *J. Differential Equations* **265** (2018) 1044–1096.
41. L. Li and D. Wang, Complete monotonicity-preserving numerical methods for time fractional ODEs, *Commun. Math. Sci.* **19** (2021) 1301–1336.
42. H.-L. Liao, T. Tang and T. Zhou, A second-order and nonuniform time-stepping maximum-principle preserving scheme for time-fractional Allen–Cahn equations, *J. Comput. Phys.* **414** (2020) 109473, 16.
43. H.-L. Liao, T. Tang and T. Zhou, An energy stable and maximum bound preserving scheme with variable time steps for time fractional Allen–Cahn equation, *SIAM J. Sci. Comput.* **43** (2021) A3503–A3526.
44. Y. Lin, X. Li and C. Xu, Finite difference/spectral approximations for the fractional cable equation, *Math. Comp.* **80** (2011) 1369–1396.
45. W. Liu, M. Röckner and J. da Silva, Quasi-linear (stochastic) partial differential equations with time-fractional derivatives, *SIAM J. Math. Anal.* **50** (2018) 2588–2607.
46. V. Mehandiratta, M. Mehra and G. Leugering, Optimal control problems driven by time-fractional diffusion equations on metric graphs: Optimality system and finite difference approximation, *SIAM J. Control Optim.* **59** (2021) 4216–4242.
47. R. Metzler and J. Klafter, The random walk's guide to anomalous diffusion: A fractional dynamics approach, *Phys. Rep.* **339** (2000) 77.
48. K. Mustapha and D. Schötzau, Well-posedness of *hp*-version discontinuous Galerkin methods for fractional diffusion wave equations, *IMA J. Numer. Anal.* **34** (2014) 1426–1446.
49. T. Namba, On existence and uniqueness of viscosity solutions for second order fully nonlinear PDEs with Caputo time fractional derivatives, *NoDEA Nonlinear Differential Equations Appl.* **25** (2018) 39.
50. R. Nochetto, E. Otárola and A. J. Salgado, A PDE approach to space-time fractional parabolic problems, *SIAM J. Numer. Anal.* **54** (2016) 848–873.
51. R. Nochetto, G. Savaré and C. Verdi, *A posteriori* error estimates for variable time-step discretizations of nonlinear evolution equations, *Commun. Pure Appl. Math.* **53** (2000) 525–589.
52. L. Plociniczak, Numerical method for the time-fractional porous medium equation, *SIAM J. Numer. Anal.* **57** (2019) 638–656.
53. C. Quan, T. Tang and J. Yang, How to define dissipation-preserving energy for time-fractional phase-field equations, *CSIAM Trans. Appl. Math.* **1** (2020) 478–490.
54. T. Roubíček, *Nonlinear Partial Differential Equations with Applications*, International Series of Numerical Mathematics, Vol. 153, 2nd edn. (Birkhäuser/Springer, 2013).

55. O. Savin and E. Valdinoci, Γ-convergence for nonlocal phase transitions, *Ann. Inst. H. Poincaré Anal. Non Linéaire* **29** (2012) 479–500.
56. W. Schirotzek, *Nonsmooth Analysis*, Universitext (Springer, 2007).
57. G. Smirnov, *Introduction to the Theory of Differential Inclusions*, Graduate Studies in Mathematics, Vol. 41 (Amer. Math. Soc., 2002).
58. P. Stinga and J. Torrea, Extension problem and Harnack's inequality for some fractional operators, *Commun. Partial Differential Equations* **35** (2010) 2092–2122.
59. M. Stynes, Too much regularity may force too much uniqueness, *Fract. Calc. Appl. Anal.* **19** (2016) 1554–1562.
60. M. Stynes, Fractional-order derivatives defined by continuous kernels are too restrictive, *Appl. Math. Lett.* **85** (2018) 22–26.
61. M. Stynes, Singularities, in *Handbook of Fractional Calculus with Applications*, Vol. 3 (De Gruyter, 2019), pp. 287–305.
62. T. Tang, H. Yu and T. Zhou, On energy dissipation theory and numerical stability for time-fractional phase-field equations, *SIAM J. Sci. Comput.* **41** (2019) A3757–A3778.
63. E. Topp and M. Yangari, Existence and uniqueness for parabolic problems with Caputo time derivative, *J. Differential Equations* **262** (2017) 6018–6046.
64. J. Vázquez, *The Porous Medium Equation: Mathematical Theory*, Oxford Mathematical Monographs (Oxford Univ. Press, 2007).
65. V. Vergara and R. Zacher, Lyapunov functions and convergence to steady state for differential equations of fractional order, *Math. Z.* **259** (2008) 287–309.
66. V. Vergara and R. Zacher, Optimal decay estimates for time-fractional and other nonlocal subdiffusion equations via energy methods, *SIAM J. Math. Anal.* **47** (2015) 210–239.
67. A. Vityuk, Existence of solutions of a differential inclusion of fractional order with an upper-semicontinuous right-hand side, *Ukraïn. Mat. Zh.* **51** (1999) 1562–1565.
68. P. Wittbold, P. Wolejko and R. Zacher, Bounded weak solutions of time-fractional porous medium type and more general nonlinear and degenerate evolutionary integro-differential equations, *J. Math. Anal. Appl.* **499** (2021) 20.
69. R. Zacher, A weak Harnack inequality for fractional differential equations, *J. Integral Equations Appl.* **19** (2007) 209–232.
70. R. Zacher, Weak solutions of abstract evolutionary integro-differential equations in Hilbert spaces, *Funkcial. Ekvac.* **52** (2009) 1–18.
71. R. Zacher, Global strong solvability of a quasilinear subdiffusion problem, *J. Evol. Equ.* **12** (2012) 813–831.
72. R. Zacher, A De Giorgi–Nash type theorem for time fractional diffusion equations, *Math. Ann.* **356** (2013) 99–146.
73. R. Zacher, A weak Harnack inequality for fractional evolution equations with discontinuous coefficients, *Ann. Sc. Norm. Super. Pisa Cl. Sci.* (5) **12** (2013) 903–940.
74. R. Zacher, Time fractional diffusion equations: Solution concepts, regularity, and long-time behavior, in *Handbook of Fractional Calculus with Applications*, Vol. 2 (De Gruyter, 2019), pp. 159–179.
75. Y. Zhang, Numerical treatment of the modified time fractional Fokker–Planck equation, *Abstr. Appl. Anal.* **2014** (2014) 10.