A DISCRETIZATION OF CAPUTO DERIVATIVES WITH APPLICATION TO TIME FRACTIONAL SDEs AND GRADIENT FLOWS*

LEI LI† AND JIAN-GUO LIU‡

Abstract. We consider a discretization of Caputo derivatives resulted from deconvolving a scheme for the corresponding Volterra integral. Properties of this discretization, including signs of the coefficients, comparison principles, and stability of the corresponding implicit schemes, are proved by its linkage to Volterra integrals with completely monotone kernels. We then apply the backward scheme corresponding to this discretization to two time fractional dissipative problems, and these implicit schemes are helpful for the analysis of the corresponding problems. In particular, we show that the overdamped generalized Langevin equation with fractional noise has a unique limiting measure for strongly convex potentials and we establish the convergence of numerical solutions to the strong solutions of time fractional gradient flows. The proposed scheme and schemes derived using the same philosophy can be useful for many other applications as well.

Key words. deconvolution, fractional SDE, fractional gradient flows, implicit scheme, completely monotone sequence

AMS subject classifications. 65C30, 65L20

DOI. 10.1137/19M123854X

1. Introduction. Continuous time fractional calculus has been used widely in physics and engineering for the memory effect, viscoelasticity, porous media, etc. [1, 2, 3]. Among them, the Caputo and Riemann–Liouville definitions are very popular for power law memory kernels [4, 1, 2, 5]. Caputo's definition of fractional derivatives was first introduced in [4] to study the memory effect of energy dissipation for some anelastic materials and soon became a useful modeling tool in engineering and physical sciences for nonlocal interactions in time (see [6, 7, 8]). Compared with Riemann–Liouville derivatives, Caputo derivatives remove the singularities at the origin and are suitable for initial value problems [5]. There are other models that have power law kernels with certain cutoffs (especially exponential cutoffs) so that they can give transitions between different behaviors [9, 10, 11, 12, 13]. In this paper, we are interested in discretizing gradient type time fractional dissipative problems. Moreover, we desire to use numerical discretization to investigate the properties of solutions of time continuous dissipative problems. The first problem is the time fractional stochastic differential equation (FSDE) of dissipative type, which is the overdamped limit of the generalized Langevin equation with fractional noise. Another problem is the time fractional gradient flows in a separable Hilbert space.

In some complex systems, the Langevin equations cannot give accurate predictions and the interaction between the system and the surrounding heat bath can no longer be modeled by white noise [14, 6]. The generalized Langevin equation (GLE)

$$\dot{X} = v,$$

(1.1)

$$m\dot{v} = -\nabla V(X) - \int_{t_0}^{t} \gamma(t-s)v(s)\,ds + \eta(t)$$

was then proposed by Mori [15] and Kubo [14] to describe such complex systems with memory. In this model, $-\int_{t_0}^{t}\gamma(t-s)v(s)\,ds$ is the friction acting on the system we consider and t_0 is the point where the memory is counted from (usually 0 or $-\infty$). The friction is the mean effect of the interaction between the system and the heat bath. The last term $\eta(t)$ is the noise which is the fluctuation part of the interaction between the system and the heat bath. Later, the GLE was recovered by dimension reduction from Ford–Kac and Kac–Zwanzig models using Mori–Zwanzig projection [16, 17, 18, 19]. In the GLE models, the noise η and the kernel for the friction $\gamma(\cdot)$ satisfy the so-called fluctuation-dissipation theorem (FDT)

(1.2) $$\mathbb{E}(\eta(t)\eta(t+\tau)) = kT\gamma(|\tau|) \ \forall \tau \in \mathbb{R}.$$

Intuitively, the random force and the frictional kernel all originate from the interaction between the system and the surrounding environment. When the energy balance is reached, they must be related for the system to achieve the correct temperature. In [6], Kou and Xie considered the GLE with fractional Gaussian noise to explain the subdiffusive behaviors for a protein molecule in solution. Later, this model was studied by many authors [20, 21]. The fractional Gaussian noise is the distributional derivative of the fractional Brownian motion B_H (see [22] and section 4.1 for more details)

(1.3) $$\eta = \sigma\dot{B}_H(t).$$

Using FDT (1.2) and considering the overdamped limit, we obtain the FSDE as the overdamped GLE (see section 4.1 for a simple derivation and the rigorous definition):

(1.4) $$D_c^{2-2H}X = -\nabla V(X) + \sigma dB_H,$$

where D_c^{α} is the Caputo derivative (see section 2 for more explanation). In [3], the FSDE has been studied theoretically. If the force is linear, it was shown that the process converges in law to a unique limiting measure. Moreover, if the FDT is satisfied, the limiting measure is the Gibbs measure. The general potential V cases seem hard to justify. In [23], numerical methods have been designed for the overdamped GLE and the numerical results there give positive evidence. One of our goals in this paper is to use the numerical schemes to prove that the limiting measure is unique if the potential V is strongly convex.

Though there might not be strong physical interpretation, the time fractional gradient flow is of its own mathematical interest and can be used for new phase field models (see [24, 25] for the phase field models). In particular, consider a separable Hilbert space H and a functional $\phi : H \to \mathbb{R}$ that is lower semicontinuous. The time fractional gradient flow we consider is

(1.5) $$D_c^{\alpha}u \in -\partial\phi(u), \quad u(0) = u_0,$$

where the Frechét subdifferential $\partial\phi$ at u is a set defined as

(1.6) $$\partial\phi(u) := \left\{ \xi \in H : \liminf_{w \to u} \frac{\phi(w) - \phi(u) - \langle \xi, w - u \rangle}{|w - u|} \geq 0 \right\}.$$

If $\partial\phi$ contains a single point ξ, then we define $\text{grad}\phi(u) := \xi$. We aim to investigate the discretization using our scheme in this paper and establish error estimates. For related fractional gradient flow, one can see [9], where the memory kernel takes the form $t^{-\gamma}e^{-\mu t}$ with exponential decay. If ϕ is convex, then $\partial\phi$ is accretive, and some related Volterra equations have been discussed in [26, 27], where the existence of generalized solutions have been established using the Yosida approximations. The equation we will consider is not included in these papers.

Numerical discretizations of time fractional differential equations and related equations have already been investigated by many authors [28, 29, 30, 31, 32, 33, 34, 35, 36]. In particular, the authors of [30, 31] applied the L^1 schemes, which approximate the Caputo derivative directly, for several dissipative problems. In [32], some corrections are made for the first $k-1$ steps so that the nonsmoothness at $t = 0$ does not pollute the desired accuracy of the schemes. In [35, 36], some spectral methods have been developed for fractional differential equations. Moreover, in [37], some comparison principles for the discrete fractional equations have been established. Unfortunately, using these discretizations to study the FSDE and time fractional gradient flows is not appropriate because the time continuous problems are not well understood yet. Our approach is to consider the discretization of the integral formulation first and apply the deconvolution (see [38]) to obtain the discretization of the Caputo derivatives in differential form. Since the integral formulation is more suitable for passing the limit, we are then able to conclude the important results regarding the time continuous problems and establish some error estimates. Note that the new scheme is not just discretization of Volterra integrals since some important properties will be proved based on the deconvolved sequence, which seems very hard using the discretization of the integral form. Besides the problems considered in this paper, the scheme proposed here or schemes derived using the same philosophy may be applied for other problems [13, 39].

The rest of the paper is organized as follows. In section 2, we give the basic notation and propose the discretization of the Caputo derivatives using deconvolution. In section 3, we prove some important properties of the new discretization. Sections 4 and 5 are devoted to the FSDE and time fractional gradient flows. In particular, we show that the overdamped GLE with fractional noise has a unique limiting measure for strongly convex potentials; we also establish some error estimates for the strong solutions of time fractional gradient flows.

2. Notation and setup for the discretization. Let B be a Banach space. Consider the following equation for a mapping: $X : [0, T] \to B$:

$$(2.1) \qquad D_c^\alpha X(t) = f(t),$$

where $f : [0, T] \to B$ is some mapping. D_c^α represents the Caputo derivative of order $\alpha \in (0, 1)$ ([1, 2]). If $X(\cdot)$ is regular enough, for example, absolutely continuous, the Caputo derivative traditionally is defined as

$$(2.2) \qquad D_c^\alpha X(t) = \frac{1}{\Gamma(1-\alpha)} \int_0^t \frac{\dot{X}(s)}{(t-s)^\alpha}\, ds.$$

In [5, 40], a generalized definition of Caputo derivative based on convolution groups was proposed. To explain this generalized defintion, we first recall the distributions $\{g_\beta\}$ in [5]:

$$(2.3) \qquad g_\beta(t) = \begin{cases} \frac{1}{\Gamma(1+\beta)} D\left(\theta(t)t^\beta\right), & \beta \in (-1,0), \\ \frac{1}{\Gamma(\beta)} t_+^{\beta-1}, & \beta > 0. \end{cases}$$

Here $\theta(t)$ is the standard Heaviside step function, $\Gamma(\cdot)$ is the gamma function, $t_+ = \theta(t)t = \max(t,0)$, and D means the distributional derivative on $\mathbb{R}$. Indeed, g_β can be defined for $\beta \in \mathbb{R}$ (see [5]) so that $\{g_\beta : \beta \in \mathbb{R}\}$ forms a convolution group. In particular, we have

$$(2.4) \qquad g_{\beta_1} * g_{\beta_2} = g_{\beta_1+\beta_2}.$$

Note that the support of g_{β_i} $(i = 1, 2)$ is bounded from the left, so the convolution is well defined.

DEFINITION 2.1 (see [5, 40]). *Let $0 < \alpha < 1$. Consider $X \in L^1_{\mathrm{loc}}([0,T), B)$. Given $X_0 \in B$, we define the αth order generalized Caputo derivative of X, associated with initial value X_0, to be a distribution as $D_c^\alpha X : C_c^\infty(-\infty, T; \mathbb{R}) \to B$ with support in $[0, T)$, given by*

$$(2.5) \qquad D_c^\alpha X = g_{-\alpha} * \left((X - X_0)\theta(t)\right).$$

If $\lim_{t \to 0+} \frac{1}{t} \int_0^t \|X(s) - X_0\|_B ds = 0$, we call $D_c^\alpha X$ the Caputo derivative of X.

The weak Caputo derivatives in [40] for mappings in general Banach spaces was defined through a dual equality using right derivatives. One can verify easily that the one in [40] agrees with Definition 2.1. This generalized definition appears complicated. However, it is theoretically more convenient, since it allows us to take advantage of the underlying group structure. In fact, making use of the convolutional group structure (2.4) (see [5] for more details), it is straightforward to convert (2.1) with (2.5) into the Volterra type equation

$$(2.6) \qquad X(t) = X_0 + \frac{1}{\Gamma(\alpha)} \int_0^t (t-s)^{\alpha-1} f(s)\, ds.$$

Indeed (2.6) is well known for regular enough f; see [41, Lemma 2.3]. The theory in [5, 40] tells us that it still holds for f to be distributions.

For absolutely continuous functions, Definition 2.1 reduces to (2.2). In this paper, we sometimes need the generalized definition, Definition 2.1, and its equivalence to (2.6) since we need to consider the Caputo derivative of a continuous function later.

For numerical setup, we fix the terminal time T and consider time step

$$(2.7) \qquad k = T/N.$$

Define $t_n = nk$. We will use X_n to represent the numerical solution at t_n.

2.1. Discretization of the fractional derivatives: Two options. Depending on whether we discretize (2.1) or (2.6), we can possibly have different schemes (see [37, section 6] for some relevant discussions). Discretization of (2.6) and deconvolution yields a discretization of the Caputo derivative, whose implicit scheme turns out to be very useful for studying two important time fractional dissipative problems. In particular, we can conclude the asymptotic behavior of the FSDEs and study the time fractional gradient flows in separable Hilbert spaces (see sections 4 and 5, respectively).

2.1.1. Discretization of the differential form. Discretizing (2.1) directly is well studied in literature (see [42, 29]). The L^1 scheme in [42, 29] is widely used in applications due to the good sign of the coefficients (see [31, 43]). The scheme is given by

$$(2.8) \qquad (\bar{\mathcal{D}}^\alpha X)_n = k^{-\alpha}(\bar{c}_0 X_n - \bar{c}_1 X_{n-1} - \cdots - \bar{c}_n^n X_0).$$

Here, the coefficients are given by

$$(2.9) \quad \begin{aligned} & \Gamma(2-\alpha)\bar{c}_0 = 1, \\ & \Gamma(2-\alpha)\bar{c}_j = -((j+1)^{1-\alpha} - 2j^{1-\alpha} + (j-1)^{1-\alpha}), \quad 1 \le j \le n-1, \\ & \Gamma(2-\alpha)\bar{c}_n^n = (n^{1-\alpha} - (n-1)^{1-\alpha}). \end{aligned}$$

We have the following observations: (i) $\bar{c}_j > 0$, $\bar{c}_n^n > 0$; (ii) $\bar{c}_0 - \sum_{j=1}^{n-1} \bar{c}_j - \bar{c}_n^n = 0$; (iii)

$$(2.10) \quad \bar{c}_j = \frac{-1}{\Gamma(-\alpha)} j^{-1-\alpha} \left(1 + O\left(\frac{1}{j}\right)\right), \ j \to \infty, \quad \bar{c}_n^n = \frac{n^{-\alpha}}{\Gamma(1-\alpha)} \left(1 + O\left(\frac{1}{n}\right)\right).$$

2.1.2. Discretization of the integral form and deconvolution. Alternatively, we can consider the discretization of the integral form (2.6) and then take deconvolution to get the approximation for the Caputo derivative. In fact, discretizing the integral form has been well studied in literature (see, for example, [28, 44]). Slightly different from the discretizations in these works, what we choose to do is to approximate f with piecewise constant functions. Then, we take deconvolution and get the approximation to the differential form.

To start, we approximate $f(t)$ by

$$(2.11) \qquad \tilde{f}(t) = F_n, \quad t \in (t_{n-1}, t_n].$$

Then, (2.6) gives the scheme

$$(2.12) \qquad X_n - X_0 = k^\alpha \sum_{m=1}^{n} a_{n-m} F_m =: (J_k F)_n,$$

where the right-hand side is a discrete integral and the sequence a is given by

$$(2.13) \qquad a = (a_0, a_1, \ldots, a_n, \ldots) = \frac{1}{\Gamma(1+\alpha)}(1, 2^\alpha - 1, 3^\alpha - 2^\alpha, \ldots).$$

For convenience, we define $F_0 = 0$ and introduce the sequence $F \in B^{\mathbb{N}}$ by

$$(2.14) \qquad F = (F_0, F_1, \ldots, F_n, \ldots),$$

so that for $n \ge 0$, $X_n - X_0 = k^\alpha(a * F)_n$. The convolution between u and v is given by

$$(2.15) \qquad (u * v)_n = \sum_{m=0}^{n} u_m v_{n-m} = (v * u)_n.$$

Let $a^{(-1)}$ be the convolution inverse of a such that

$$(2.16) \qquad a * a^{(-1)} = a^{(-1)} * a = \delta_d := (1, 0, 0, \ldots).$$

Then, we obtain for $n \geq 0$ that $k^{-\alpha}(a^{(-1)} * (X - X_0))_n = F_n$. We therefore obtain a new scheme for discretizing the Caputo derivative

$$(2.17) \qquad\qquad (\mathcal{D}^\alpha X)_n = k^{-\alpha}(a^{(-1)} * (X - X_0))_n.$$

Though equivalent to the discretization of Volterra integral, we regard this as a new scheme because some important properties (e.g., Theorem 3.3(2)–(3) and (5.15)) will be proved based on this differential form (2.17), which will be hard using the integral form (2.12).

3. Properties of the discretization. In this section, we discuss in detail the properties of discretization (2.17). One can refer to [38] for some discussion of using deconvolution to define discrete fractional calculus.

We first introduce some definitions for the discussion. We say a sequence $v = (v_0, v_1, \ldots)$ is completely monotone if $((I - S)^j v)_k \geq 0$ for any $j \geq 0, k \geq 0$, where $(Sv)_j = v_{j+1}$. A sequence is completely monotone if and only if it is the moment sequence of a Hausdorff measure (a finite nonnegative measure on $[0, 1]$) [45]. Another description is given below in Lemma 3.1. The generating function of a sequence $v = (v_0, v_1, \ldots)$ is defined by

$$(3.1) \qquad\qquad F_v(z) = \sum_{n=0}^{\infty} v_n z^n.$$

Another concept we introduce is the Pick function. A function $f : \mathbb{C}_+ \to \mathbb{C}$ (where $\mathbb{C}_+$ denotes the upper half plane, not including the real line) is Pick if it is analytic such that $\text{Im}(z) > 0 \Rightarrow \text{Im}(f(z)) \geq 0$. Now, we state some properties of sequences in terms of the generating functions, for which we omit the proofs.

LEMMA 3.1.
(1) *For convolution, $F_{u*v}(z) = F_u(z)F_v(z)$, and $F_{v^{(-1)}}(z) = (F_v(z))^{-1}$.*
(2) *[46, Corollary VI.1] Assume $F_v(z)$ is analytic on $\Delta := \{z : |z| < R, z \neq 1, |\arg(z - 1)| > \theta\}$, for some $R > 1, \theta \in (0, \frac{\pi}{2})$. If $F_v(z) \sim (1 - z)^{-\beta}$ as $z \to 1, z \in \Delta$ for $\beta \neq 0, -1, -2, -3, \ldots$, then $v_n \sim \frac{1}{\Gamma(\beta)} n^{\beta-1}, \ n \to \infty$.*
(3) $\lim_{n \to \infty} v_n = \lim_{z \to 1^-} (1 - z)F_v(z)$.
(4) *[47] A sequence v is completely monotone if and only if the generating function $F_v(z) = \sum_{j=0}^{\infty} v_j z^j$ is a Pick function that is analytic and nonnegative on $(-\infty, 1)$.*

For convenience, define a sequence $c = (c_0, c_1, \ldots, c_n, \ldots)$ as (see (2.13) for a)

$$(3.2) \qquad\qquad c_0 = a_0^{(-1)}, \ \ c_i = -a_i^{(-1)}, \ \forall i \geq 1.$$

Moreover, it is convenient to introduce

$$(3.3) \qquad\qquad c_n^n = c_0 - \sum_{i=1}^{n-1} c_i.$$

Then, (2.17) can be reformulated as

$$(3.4) \qquad \begin{aligned} (\mathcal{D}^\alpha X)_n &= k^{-\alpha}\Big(c_0(X_n - X_0) - \sum_{i=1}^{n-1} c_i(X_{n-i} - X_0)\Big) \\ &= k^{-\alpha}\Big(c_0 X_n - \sum_{i=1}^{n-1} c_i X_{n-i} - c_n^n X_0\Big). \end{aligned}$$

Using the result in [38], we have the following claims.

PROPOSITION 3.2. *Consider scheme* (3.4). *The following claims hold:*
(1) $c_i > 0$ *for* $i \geq 0$ *and* $c_0 = \sum_{i=1}^{\infty} c_i = \Gamma(1+\alpha)$. *Consequently,* $c_n^n = \sum_{i=n}^{\infty} c_i > 0$
   and $c_0 = \sum_{i=1}^{n-1} c_i + c_n^n$.
(2) *We have the following asymptotics for the coefficients:*

$$(3.5) \quad c_j = \frac{-1}{\Gamma(-\alpha)} j^{-1-\alpha} \left(1 + o(1)\right), \; j \to \infty, \quad c_n^n = \frac{n^{-\alpha}}{\Gamma(1-\alpha)} \left(1 + o(1)\right).$$

Proof. (1) First of all, we recall that (see (2.3) for the definition of g_α)

$$a_n = \int_n^{n+1} g_\alpha(t) \, dt.$$

Since $g_\alpha(\cdot)$ is completely monotone (which means $(-1)^m \frac{d^m}{dt^m} g_\alpha(t) \geq 0$ for any $t \in (0, \infty)$), then the sequence $a = \{a_n\}$ is completely monotone. Since

$$a^{(-1)} = (c_0, -c_1, -c_2, \ldots),$$

by [38, Theorem 2.3], one has that $c_0 > 0$ while $(c_1, c_2, \ldots)$ is a completely monotone sequence. The sign of c_i is thus proved.

Using the explicit formula for a_n, $F_a(z) \to \infty$ as $z \to 1^-$. Hence, we find that $F_{a^{(-1)}}(z) \to 0$, $z \to 1^-$. Noting the sign of elements for $a^{(-1)}$, the monotone convergence theorem holds and thus $\sum_{i=0}^{\infty} a_i^{(-1)} = 0$. The other claims then follow accordingly.

(2) We consider the function

$$H(z) = F_a(z) - (1-z)^{-\alpha} =: \sum_{n=0}^{\infty} d_n z^n.$$

By [46, Theorem VI.1], we have $[z^n](1-z)^{-\alpha} = \frac{1}{\Gamma(\alpha)} n^{\alpha-1}(1 + O(\frac{1}{n}))$, $n \to \infty$, where $[z^n]F(z)$ means the coefficient of z^n in the series expansion of $F(z)$ about 0. Hence, $|d_n| \leq C \frac{1}{n^{2-\alpha}}$, and $H(z)$ is a locally bounded function (bounded on any compact set). Hence,

$$F_{a^{(-1)}}(z) = \frac{(1-z)^\alpha}{1 + (1-z)^\alpha H(z)}.$$

Clearly, this function is analytic in $\{z : |z| < 1 + \varepsilon, z \neq 1, |\arg(z-1)| > \frac{\pi}{4}\}$ for some $\varepsilon > 0$. Applying the second claim in Lemma 3.1 gives the asymptotics for $c_j = -a_j^{(-1)}$ ($j \geq 1$). Using the fact $c_n^n = \sum_{i=n}^{\infty} c_i$, the asymptotics for c_n^n then follows. □

The comparison principles are important for stability of numerical schemes. Below, we prove several important comparison criteria that are helpful for the stability of the implicit schemes. For the stability of some explicit schemes, one may refer to [37].

THEOREM 3.3. *Consider discretization* (3.4). *Let* $u = \{u_n\}$, $v = \{v_n\}$ *and* $w = \{w_n\}$ *be three sequences in* $\mathbb{R}^\mathbb{N}$, *with* $u_0 \leq v_0 \leq w_0$.
(1) *(Convex functional) Suppose* $E(\cdot) : \mathbb{R}^d \to \mathbb{R}, X \mapsto E(X)$ *is convex. Then,*

$$(\mathcal{D}^\alpha E(X))_n \leq (\mathcal{D}^\alpha X)_n \cdot \nabla E(X_n).$$

(2) *(Comparison principle for nonincreasing* f*) Suppose* $f(s, \cdot)$ *is nonincreasing. Assume* u, v, w *satisfy the discrete implicit relations*

$$(\mathcal{D}^\alpha u)_n \leq f(t_n, u_n), \quad (\mathcal{D}^\alpha v)_n = f(t_n, v_n), \quad (\mathcal{D}^\alpha w)_n \geq f(t_n, w_n).$$

Then, $u_n \leq v_n \leq w_n$.

(3) *(Comparison principle for Lipschitz f)* *Assume f is Lipschitz continuous in the second variable with Lipschitz constant L. If*

$$(\mathcal{D}^\alpha u)_n \leq f(t_n, u_n), \quad (\mathcal{D}^\alpha v)_n = f(t_n, v_n), \quad (\mathcal{D}^\alpha w)_n \geq f(t_n, w_n),$$

then for step size k with $c_0 > k^\alpha L$, $u_n \leq v_n \leq w_n$.

(4) *(Comparison principle for integral form)* *Assume $f(s, \cdot)$ is nondecreasing and Lipschitz continuous in the second variable with Lipschitz constant L. Introduce f_u, f_v, f_w by, for example, $f_u = (0, f(t_1, u_1), f(t_2, u_2), \ldots)$. If*

$$u_n \leq u_0 + (J_k f_u)_n, \quad v_n = v_0 + (J_k f_v)_n, \quad w_n \geq w_0 + (J_k f_w)_n,$$

then for step size k with $k^\alpha a_0 L = k^\alpha L/c_0 < 1$, $u_n \leq v_n \leq w_n$.

Proof.

(1) By (3.4), Proposition 3.2, and the convexity of $E(\cdot)$, we have

$$(\mathcal{D}^\alpha X)_n \cdot \nabla E(X_n) = \nabla E(X_n) \cdot k^{-\alpha} \left(\sum_{i=1}^{n-1} (X_n - X_{n-1}) + c_n^n (X_n - X_0) \right)$$

$$\geq k^{-\alpha} \left(\sum_{i=1}^{n-1} (E(X_n) - E(X_{n-1})) + c_n^n (E(X_n) - E(X_0)) \right)$$

$$= (\mathcal{D}^\alpha E(X))_n.$$

(2) Let $\xi_n = u_n - v_n$. Then,

$$\mathcal{D}^\alpha \xi \leq f(t_n, u_n) - f(t_n, v_n).$$

Multiplying $1(\xi_n \geq 0)$ on both sides, and defining $\eta_n = \xi_n \vee 0 = \max(\xi_n, 0)$, we have

$$1(\xi_n \geq 0) k^{-\alpha} \left(c_0 \xi_n - \sum_{i=1}^{n-1} c_i \xi_{n-i} - c_n^n \xi_0 \right) \leq 1(\xi_n \geq 0)(f(t_n, u_n) - f(t_n, v_n)) \leq 0.$$

Since $\xi_n 1(\xi_n \geq 0) = \xi_n \vee 0 = \eta_n$, $\xi_i 1(\xi_n \geq 0) \leq \xi_i \vee 0 = \eta_i$, one easily finds that

$$(\mathcal{D}^\alpha \eta)_n \leq 1(\xi_n \geq 0)(\mathcal{D}^\alpha \xi)_n \leq 0.$$

Since $\eta_0 = 0$, one easily finds $\eta_n \leq 0$, and hence $u_n \leq v_n$. It is similar to compare v_n and w_n.

(3) We compare u_n with v_n. We know already $u_0 \leq v_0$. Now, suppose $n \geq 1$ and assume for all $m \leq n-1$, we have proved $u_m \leq v_m$ already. We now consider $m = n$.

$$k^{-\alpha} c_0 (u_n - v_n) \leq (\mathcal{D}^\alpha (u - v))_n \leq f(t_n, u_n) - f(t_n, v_n) \leq L|u_n - v_n|.$$

If $u_n > v_n$, we then have $(k^{-\alpha} c_0 - L)(u_n - v_n) \leq 0$, which is clearly not true. Hence, induction shows that the claim is true for all n. Comparing the sequence v with w is similar and we omit.

(4) Direct computation shows that

$$u_n - v_n \leq u_0 - v_0 + k^\alpha a_0 (f(t_n, u_n) - f(t_n, v_n))$$

$$+ k^\alpha \sum_{m=1}^{n-1} a_{n-m} (f(t_m, u_m) - f(t_m, v_m)).$$

If we have proved that $u_m \leq v_m$ for $m \leq n-1$, then $u_n - v_n \leq k^\alpha a_0 L |u_n - v_n|$. The proof then follows by induction similarly as in 3. Comparing the sequence v with w is similar and we omit. $\qquad\square$

We now consider the stability of the implicit scheme applied to the simple fractioanl ODEs (FODEs)

$$(3.6) \qquad D_c^\alpha X = \lambda X, \quad X(0) = x_0 > 0,$$

whose solution is given by $X(t) = x_0 E_\alpha(\lambda t^\alpha)$, where

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

is the Mittag–Leffler function [48].

THEOREM 3.4. *Consider the implicit scheme applied on the FODE* (3.6) *with* $X_0 = X(0)$. *Then, for* $n \geq 1$:

$$(3.7) \qquad (\mathcal{D}^\alpha X)_n = \lambda X_n \Leftrightarrow X_n = X_0 + \lambda k^\alpha \sum_{m=1}^n a_{n-m} X_m.$$

(1) *If* $\lambda > 0$ *and* $k^\alpha \lambda < c_0$, *then* $X(t_n) \leq X_n \leq X_{n+1}$. *If otherwise* $\lambda < 0$, $\lim_{n\to\infty} X_n = 0$.

(2) *Consider* $\lambda > 0$. *Suppose* $k_i, i = 1, 2$, *satisfy* $k_i^\alpha \lambda < c_0$ *and* $k_1 = 2^{m_1} k_2$ *for some* $m_1 \in \mathbb{N}$. *Let* $X_n^{(i)}$ *be the numerical solutions. Define the piecewise constant functions* $\bar{X}_i(t)$ *by* $\bar{X}_i(t) = X_n^{(i)}$, $t \in (t_{n-1}^{(i)}, t_n^{(i)}]$ *for* $t_n^{(i)} = nk_i$ ($i = 1, 2$). *Then,* $\bar{X}_1(t) \geq \bar{X}_2(t)$. *Consequently, there exists a constant* $C(\alpha, T) > 0$ *such that for any* k *with* $k^\alpha \lambda \leq \frac{1}{2} c_0$,

$$(3.8) \qquad \sup_{n:nk\leq T} X_n \leq C(\alpha, T) X_0.$$

(3) *When* k *is sufficiently small,* $\sup_{n:nk\leq T} |X_n - X(t_n)| \leq C(\alpha, T) k^\alpha$.

Proof. (1) Consider $\lambda > 0$. The induction formula from the differential form reads

$$(c_0 - k^\alpha \lambda) X_n = \sum_{i=1}^{n-1} c_i X_{n-i} + c_n^n X_0.$$

If $n = 1$, $c_1^1 = c_0$. Then, we clearly have $X_1 = \frac{c_0}{c_0 - k^\alpha \lambda} X_0 > X_0$. Suppose we have proved $X_m \geq X_{m-1}$ for $m \leq n$ with $n \geq 1$. For $n + 1$,

$$(c_0 - k^\alpha \lambda) X_{n+1} = \sum_{i=1}^{n-1} c_i X_{n+1-i} + (c_n X_1 + c_{n+1}^{n+1} X_0) \geq \sum_{i=1}^{n-1} c_i X_{n-i} + (c_n X_0 + c_{n+1}^{n+1} X_0).$$

Since $c_n + c_{n+1}^{n+1} = c_n^n$, the claim then follows.

Now, consider the equivalent integral form (second in (3.7))

$$X_n = X_0 + \frac{\lambda}{\Gamma(\alpha)} \sum_{j=1}^n \int_{t_{j-1}}^{t_j} (t_n - s)^{\alpha-1} X_j \, ds.$$

The accurate solution satisfies

$$X(t_n) = X_0 + \frac{\lambda}{\Gamma(\alpha)} \sum_{j=1}^{n} \int_{t_{j-1}}^{t_j} (t_n - s)^{\alpha-1} X(s) \, ds$$

$$\leq X_0 + \frac{\lambda}{\Gamma(\alpha)} \sum_{j=1}^{n} \int_{t_{j-1}}^{t_j} (t_n - s)^{\alpha-1} X(t_j) \, ds.$$

By the third claim in Theorem 3.3, $X(t_n) \leq X_n$.

Now, we consider $\lambda < 0$. Recall that $F_0 = 0$ and

$$X_n - X_0 = k^\alpha \lambda a * (X - X_0 \delta_{n0}).$$

The generating function of X^n is thus given by

$$F_X(z) = X_0 \frac{(1-z)^{-1} - k^\alpha \lambda F_a(z)}{1 - k^\alpha \lambda F_a(z)}.$$

As $z \to 1^-$, $F_a(z) \to \infty$, $(1-z)F_a(z) \to 0$, and hence

$$\lim_{n \to \infty} X_n = \lim_{z \to 1^-} (1-z) F_X(z) = 0.$$

(2) We only need to consider $m_1 = 1$ (or $k_1 = 2k_2$). By (1), the piecewise constant functions $\bar{X}_i(t)$'s are nondecreasing. Suppose that for $n \geq 1$, one has $\bar{X}_1(t) \geq \bar{X}_2(t)$, $t \in [0, (n-1)k_1]$. Then, for $t \in ((n-1)k_1, nk_1]$, one only needs $\bar{X}_1(nk_1) \geq \bar{X}_2(nk_1) = \bar{X}_2(2nk_2)$ since $\bar{X}_2$ is nondecreasing. By the integral formulation (second in (3.7)),

$$\bar{X}_1(nk_1)$$

$$= X_0 + \frac{\lambda}{\Gamma(\alpha)} \int_0^{t_{n-1}^{(1)}} (nk_1 - s)^{\alpha-1} \bar{X}_1(s) \, ds + \frac{\lambda}{\Gamma(\alpha)} \int_{t_{n-1}^{(1)}}^{t_n^{(1)}} (nk_1 - s)^{\alpha-1} \bar{X}_1(nk_1) \, ds$$

$$\geq X_0 + \frac{\lambda}{\Gamma(\alpha)} \int_0^{(n-1)k_1} (nk_1 - s)^{\alpha-1} \bar{X}_2(s) \, ds + \frac{\lambda k_1^\alpha}{\Gamma(1+\alpha)} \bar{X}_1(nk_1).$$

On the other hand,

$$\bar{X}_2(nk_1) = X_0 + \frac{\lambda}{\Gamma(\alpha)} \int_0^{t_{n-1}^{(1)}} (nk_1 - s)^{\alpha-1} \bar{X}_2(s) \, ds$$

$$+ \frac{\lambda}{\Gamma(\alpha)} \int_{t_{n-1}^{(1)}}^{t_n^{(1)}} (nk_1 - s)^{\alpha-1} \bar{X}_2(s) \, ds.$$

The last term is simply controlled by $\frac{\lambda k_1^\alpha}{\Gamma(1+\alpha)} \bar{X}_2(2nk_2)$ due to monotonicity of $\bar{X}_2$. Since $c_0 = \Gamma(1+\alpha)$ and $\bar{X}_2(2nk_2) = \bar{X}_2(nk_1)$, we then find

$$\bar{X}_1(nk_1) \geq \frac{1}{1 - k_1^\alpha \lambda/c_0} \left(X^0 + \frac{\lambda}{\Gamma(\alpha)} \int_0^{(n-1)k_1} (nk_1 - s)^{\alpha-1} \bar{X}_2(s) \, ds \right) \geq \bar{X}_2(2nk_2).$$

We now prove the stability. For any step size k, we choose $k_0 = 2^m k$ such that $k_0^\alpha \lambda \in (\frac{c_0}{2^{1+\alpha}}, \frac{c_0}{2}]$. Then, with time step k_0, there are

$$N_0 = \frac{T}{k_0} \leq T(2^{1+\alpha} \lambda/c_0)^{1/\alpha}$$

steps. Then, consider the induction using the differential form

$$(c_0 - k_0^\alpha \lambda)X_n \le \sum_{i=1}^{n-1} c_i X_{n-m} + c_n^n X_0 \le c_0 X_{n-1}.$$

Hence, $X_n \le 2X_{n-1} \le 2^{N_0} X_0$. The claim then follows.

(3) By the explicit formula of the solution for (3.6), we know X is α-Hölder continuous, and for $t > 0$, it is smooth. Inserting $X(\cdot)$ into the integral form, we have

$$X(t_n) = X_0 + \frac{\lambda}{\Gamma(\alpha)} \sum_{m=1}^{n} \int_{t_{m-1}}^{t_m} (t_n - s)^{\alpha-1} X(s)\, ds$$

$$= X_0 + \lambda k^\alpha \sum_{m=1}^{n} a_{n-m} X(t_m) + R_n,$$

where

$$R_n = \frac{\lambda}{\Gamma(\alpha)} \sum_{m=1}^{n} \int_{t_{m-1}}^{t_m} (t_n - s)^{\alpha-1}(X(s) - X(t_m))\, ds,$$

and thus

$$|R_n| \le C(T)k^\alpha \frac{\lambda}{\Gamma(\alpha)} \sum_{m=1}^{n} \int_{t_{m-1}}^{t_m} (t_n - s)^{\alpha-1}\, ds = C(\alpha, T)k^\alpha.$$

Hence, the error $E_n := |X_n - X(t_n)|$ satisfies

$$E_n \le k^\alpha |\lambda| \sum_{m=1}^{n} a_{n-m} E_m + C k^\alpha.$$

Using the comparison principle for integral formulation in Theorem 3.3 and the stability result (3.8), we have for k sufficiently small that $E_n \le C(\alpha, T)k^\alpha$. $\qquad\square$

By Theorem 3.4, the following claims hold when we compare the numerical solution with the exact solutions of some FODEs.

COROLLARY 3.5. *Suppose $f(\cdot) \in C^2[A, \infty)$ is nondecreasing and globally Lipschitz for some $A \in \mathbb{R}$. Let $u(\cdot)$ be the solution to the FODE $D_c^\alpha u = f(u)$ with $u(0) = U > A$ and $f(U) > 0$. Let $\{u_n\}$ be the numerical solution of the implicit scheme $(\mathcal{D}^\alpha u)_n = f(u_n)$, $u_0 = u(0)$. Then, for k sufficiently small, $u(t_n) \le u_n \le u_{n+1}$. Moreover, for any T such that u exists on $[0, T]$, we have for some $C(T) > 0$ that*

$$(3.9) \qquad \sup_{n:nk\le T} |u(t_n) - u_n| \le C(T)k^\alpha.$$

Proof. The solution to the FODE satisfies the following [37]:

$$(3.10) \qquad u(t_n) = u_0 + \frac{1}{\Gamma(\alpha)} \sum_{j=1}^{n} \int_{t_{j-1}}^{t_j} (t_n - s)^{\alpha-1} f(u(t_j))\, ds + R_n,$$

where

$$R_n = \frac{1}{\Gamma(\alpha)} \sum_{j=1}^{n} \int_{t_{j-1}}^{t_j} (t_n - s)^{\alpha-1} \Big(f(u(s)) - f(u(t_j)) \Big)\, ds.$$

By the theory in [37], $u(\cdot)$ is nondecreasing and thus $R_n \leq 0$. Consequently, applying Theorem 3.3(4), we have $u(t_n) \leq u_n$.

To prove that $\{u_n\}$ is nondecreasing, we use induction. It is clear that $u_0 \leq u_1$. Now, assume we have proved $u_0 \leq u_1 \leq \cdots \leq u_n$ for $n \geq 1$. We now prove $u_n \leq u_{n+1}$. Using the equivalent integral form, we find

(3.11)
$$
\begin{aligned}
u_n - k^\alpha a_0 f(u_n) &= u_0 + \frac{1}{\Gamma(\alpha)} \sum_{j=1}^{n-1} \int_{t_{j-1}}^{t_j} (t_n - s)^{\alpha-1} f(u_j)\, ds \\
&\leq u_0 + \frac{1}{\Gamma(\alpha)} \sum_{j=2}^{n} \int_{t_{j-1}}^{t_j} (t_{n+1} - s)^{\alpha-1} f(u_j)\, ds \\
&\leq u_{n+1} - k^\alpha a_0 f(u_{n+1}).
\end{aligned}
$$

This implies that $u_n \leq u_{n+1}$ when k is sufficiently small.

Finally, by (3.10) and (3.11), we have

$$
|u(t_n) - u_n| \leq \frac{1}{\Gamma(\alpha)} \sum_{j=1}^{n} \int_{t_{j-1}}^{t_j} (t_n - s)^{\alpha-1} |f(u(t_j)) - f(u_j)|\, ds + |R_n|.
$$

It is well known that u is α-Hölder continuous on $[0, T]$ (one can, for example, combine [37, Lemma 3.1] and [49, Theorem 3.1]). Consequently, $\sup_{n:nk\leq T} |R_n| \leq C(T)k^\alpha$. Hence,

$$
|u(t_n) - u_n| \leq \frac{L}{\Gamma(\alpha)} \sum_{j=1}^{n} |u(t_j) - u_j| \int_{t_{j-1}}^{t_j} (t_n - s)^{\alpha-1}\, ds + C(T)k^\alpha.
$$

Applying Theorem 3.4(2), we thus find $\sup_{n:nk\leq T} |u(t_n) - u_n| \leq C(T)k^\alpha$. $\qquad\square$

4. Limiting behavior of the FSDE. In this section, we use the implicit scheme corresponding to the discretization (3.4) to study the FSDE as advertised in the introduction. In particular, we first of all provide some details for the derivation of the FSDE, and then prove that when the potential is strongly convex, there is a unique limiting measure for the FSDE.

4.1. A formal derivation of the FSDE. We first of all derive the autocorrelation function for the fractional noise. Recall that the fractional Brownian motion has the following:

(4.1) $$ \mathbb{E}(B_t^H B_s^H) = R_H(s, t) := \frac{1}{2}\left(s^{2H} + t^{2H} - |t - s|^{2H}\right). $$

Fix $\tau \neq 0$. Formally, for $t > 0$ with $t + \tau > 0$, it holds that

(4.2)
$$
\begin{aligned}
&\mathbb{E}(\dot{B}_H(t)\dot{B}_H(\tau + t)) \\
&= \lim_{h\to 0, h_1\to 0} \mathbb{E}\left(\frac{B_H(t + h_1) - B_H(t)}{h_1} \frac{B_H(t + \tau + h) - B_h(t + \tau)}{h}\right) \\
&= \lim_{h\to 0, h_1\to 0} \frac{1}{2hh_1}\left(|\tau + h|^{2H} - |\tau + h - h_1|^{2H} - |\tau|^{2H} + |\tau - h_1|^{2H}\right) \\
&= H(2H - 1)|\tau|^{2H-2}.
\end{aligned}
$$

Assume there is no extra singularity for $\tau = 0$; we check formally for $s < t$:

(4.3)

$$\mathbb{E}(B_t^H B_s^H) = \int_0^t \int_0^s \mathbb{E}\left(\dot{B}_H(z)\dot{B}_H(w)\right) dzdw$$

$$= \int_0^s \int_0^t H(2H-1)|z-w|^{2H-2}dzdw = \frac{1}{2}(s^{2H} + t^{2H} - (t-s)^{2H}).$$

Now (4.3) agrees with (4.1). Hence, the assumption for no extra singularity at $\tau = 0$ is reasonable. According to FDT (1.2), the GLE (1.1) is then reduced to the following dimensionless equation:

$$\varepsilon \dot{v} = -\nabla V - \frac{1}{\Gamma(2H-1)} \int_0^t (t-s)^{2H-2} v(s)\, ds + \frac{\sqrt{2}}{\sqrt{\Gamma(2H+1)}} \dot{B}_H.$$

Here, $\varepsilon = \frac{mT^{2H}}{\gamma_0}$ with T being the scale for time and γ_0 being the typical scale for the friction (see [3] for details). In the overdamped regime, $\varepsilon \ll 1$, the GLE with fractional Gaussian noise formally corresponds to the fractional SDE

(4.4)
$$D_c^\alpha X = -\nabla V(X) + \sigma \dot{B}_H$$

with $\alpha = 2-2H$, $\sigma = \frac{\sqrt{2}}{\sqrt{\Gamma(2H+1)}}$. This overdamped GLE is rigorously defined through the following integral formulation:

(4.5)
$$X(t) = X_0 + \frac{1}{\Gamma(\alpha)} \int_0^t (t-s)^{\alpha-1} b(X(s))\, ds + G(t),$$

where

(4.6)
$$G(t) := \frac{\sigma}{\Gamma(\alpha)} \int_0^t (t-s)^{\alpha-1} dB_H(s).$$

We can consider generally $\alpha \in (0,1)$ and $\sigma > 0$ for FSDEs. Of course, only the one with $\alpha = 2-2H$ has physical significance, which is the overdamped GLE. It has been shown in [3] that when $\alpha = 2-2H$ and $\sigma = \frac{\sqrt{2}}{\sqrt{\Gamma(2H+1)}}$, $G(t)$ is another fractional Brownian motion with Hurst parameter $1-H$ up to some multiplicative constant: $G(t) \sim \beta_H B_{1-H}$ with $\beta_H = \frac{\sqrt{2}}{\sqrt{\Gamma(3-2H)}}$.

When the force $-\nabla V(x)$ is linear, the distribution of X converges algebraically to the Gibbs measure [3]. For general cases, whether it converges to the Gibbs measure is unknown. Recently, in the case of overdamped GLE, some numerical experiments indicate that the law of X still converges algebraically to the corresponding Gibbs measure for general potential [23].

4.2. Convergence to equilibrium for strongly convex potentials. In this subsection, we will try to use our discretization to study the limit behaviors of the FSDE for the strongly convex potential V. In particular, we show that there is a unique limiting measure as $t \to \infty$. Letting

(4.7)
$$b(x) = -\nabla V(x),$$

we will assume the following.

Assumption 4.1. There exists some $\mu > 0$ such that

$$(4.8) \qquad (x - y) \cdot (b(x) - b(y)) \leq -\mu|x - y|^2 \quad \forall x, y \in \mathbb{R}^d.$$

Moreover, $b(\cdot)$ is Lipschitz continuous so that for some $L > 0$,

$$(4.9) \qquad |b(x) - b(y)| \leq L|x - y|.$$

The Lipschitz condition of b may be relaxed by proving that the probability density of the process decays fast at infinity. Since this is not our focus, we assume the Lipschitz condition for simplicity. As proved in [3] and [23], with assumption (4.9), the FSDE (4.5) has a unique strong solution and for any $T > 0$ there exists $C(T) > 0$ such that the following hold:

$$(4.10) \qquad \sup_{t \geq T} \mathbb{E}|X(t)|^2 \leq C(T), \quad \sup_{t \geq T} \sqrt{\mathbb{E}|X(t + \delta) - X(t)|^2} \leq C(T)\delta^{H + \alpha - 1}.$$

Now, we consider the limiting behavior of the law for $X(t)$. Given two different initial data $X^{(i)}(0)$, $i = 1, 2$, we consider the strong solutions of (4.5). We will use the synchronization coupling to compare the distributions of the two processes. Taking the difference between two solutions, we have

$$X^{(1)}(t) - X^{(2)}(t) = X^{(1)}(0) - X^{(2)}(0) + \frac{1}{\Gamma(\alpha)} \int_0^t (t - s)^{\alpha - 1}(b(X^{(1)}(s)) - b(X^{(2)}(s))) \, ds.$$

To get the idea of a proof, we apply the theory in [5] so that the Caputo derivatives can be defined pathwise for $X^{(1)} - X^{(2)}$ (see also section 2 for a brief introduction). In the distributional sense, it holds that

$$(4.11) \qquad D_c^\alpha(X^{(1)} - X^{(2)})(t) = b(X^{(1)}(t)) - b(X^{(2)}(t)) \text{ almost surely.}$$

If $X^{(1)} - X^{(2)}$ is regular enough, applying [5, Proposition 3.11], we have

$$(4.12)$$
$$\frac{1}{2}D_c^\alpha|X^{(1)} - X^{(2)}|^2(t) \leq (X^{(1)} - X^{(2)}) \cdot (b(X^{(1)}) - b(X^{(2)})) \leq -\mu|X^{(1)} - X^{(2)}|^2.$$

If we define $u(t) := \mathbb{E}(|X^{(1)}(t) - X^{(2)}(t)|^2)$, it then holds that

$$D_c^\alpha u(t) \leq -2\mu u(t).$$

Applying the comparison principle for $D_c^\alpha u = f(t, u)$ with nonincreasing $f(t, \cdot)$ (see, for example, [50, Theorem 2.1]) yields

$$(4.13) \qquad u(t) \leq u(0)E_\alpha(-2\mu t^\alpha).$$

If this is true, we then are able to compare the laws of the two strong solutions of (4.5) under Wasserstein-2 distance. Recall that the Wasserstein-2 distance is given by [51, pp. 99–102]

$$(4.14) \qquad W_2(\mu, \nu) = \left(\inf_{\gamma \in \Pi(\mu, \nu)} \int_{\mathbb{R}^d \times \mathbb{R}^d} |x - y|^2 d\gamma \right)^{1/2},$$

where $\Pi(\mu, \nu)$ is the set of joint distributions whose marginal distributions are μ and ν, respectively. Equation (4.13) will imply the convergence of the law of the process to the unique limiting measure.

The issue in the above argument is that (4.12) is not justified rigorously. In the following, we shall utilize an implicit scheme based on discretization (3.4) to prove the convergence of the law. In fact, we have the following claims.

THEOREM 4.2. *Suppose assumption* (4.9) *holds and* $X^{(i)}(t)$ *are the two strong solutions to the FSDE* (4.5) *with initial data* $X^i(0) = X_0^{(i)} \sim \mu_0^{(i)}$ *(i = 1, 2), where* $\mu_0^{(i)}$ *are some given probability measures. Then, the laws of* $X^{(i)}(t)$ *satisfy in Wasserstein-2 distance that*

$$(4.15) \qquad W_2(\mu^{(1)}(t), \mu^{(2)}(t)) \leq W_2(\mu_0^{(1)}, \mu_0^{(2)}) \sqrt{E_\alpha(-2\mu t^\alpha)}.$$

Consequently, the FSDE model has a unique limiting measure π.

We will apply the following backward Euler scheme based on (3.4) to FSDE (4.5):

$$(4.16) \qquad X_n = X(0) + k^\alpha \sum_{m=1}^{n} a_{n-m} b(X_m) + G(t_n).$$

We need some preparation for the complete proof. The first is the following convergence result of the scheme (4.16).

LEMMA 4.3. *Suppose assumption* (4.9) *holds and* $X(t)$ *is the unique strong solution to* (4.5). *Let* X_n *be the numerical solution to* (4.16). *Then, for k with* $k^\alpha L < c_0/2$,

$$(4.17) \qquad \sup_{n:nk \leq T} \sqrt{\mathbb{E}(|X_n - X(nk)|^2)} \leq C(\alpha, T) k^{\alpha+H-1}.$$

Proof. The proof is very similar to that for the third claim in Theorem 3.4. In fact, the strong solution of (4.5) satisfies

$$(4.18) \qquad X(t_n) = X(0) + k^\alpha \sum_{m=1}^{n} a_{n-m} b(X(t_m)) + R_n,$$

where

$$R_n := \frac{1}{\Gamma(\alpha)} \sum_{m=1}^{n} \int_{t_{m-1}}^{t_m} (b(X(s)) - b(X(t_m))).$$

Using (4.9) and (4.10), one finds

$$(\mathbb{E}|R_n|^2)^{1/2} \leq C k^{\alpha+H-1}.$$

Taking the difference between (4.16) and (4.18) and defining $E_n := (\mathbb{E}|X_n - X(t_n)|^2)^{1/2}$, one then has

$$(4.19) \qquad E_n \leq k^\alpha L \sum_{m=1}^{n} a_{n-m} E_m + C k^{\alpha+H-1}.$$

Finally, using the comparison principle for integral formulation in Theorem 3.3 and the stability result (3.8), the claim follows. $\qquad\square$

Consider two numerical solutions $\{X_n^{(1)}\}$ and $\{X_n^{(2)}\}$ with initial data $X_0^{(i)}$ ($i = 1, 2$), with synchronization coupling. The variable $Z_n := X_n^{(1)} - X_n^{(2)}$ satisfies the following relation:

$$(4.20) \qquad Z_n = Z_0 + k^\alpha \sum_{m=1}^{n} a_{n-m} (b(X_m^{(1)}) - b(X_m^{(2)})).$$

Equivalently, one has almost surely that

$$k^{-\alpha} \left(\sum_{j=1}^{n-1} c_j (Z_n - Z_{n-j}) + c_n^n (Z_n - Z_0) \right) = b(X_n^{(1)}) - b(X_n^{(2)}).$$

Applying the first claim in Theorem 3.3 for $E(u) = \frac{1}{2}u^2$, one has almost surely that

$$(4.21) \qquad (\mathcal{D}^\alpha |Z|^2)_n \leq -2 Z_n \cdot (b(X_n^{(1)}) - b(X_n^{(2)})) \leq -2\mu |Z_n|^2.$$

The point is that one may pass this inequality somehow to the strong solutions of (4.5) by taking $k \to 0$.

Proof of Theorem 4.2. Define

$$u_n := \mathbb{E}(|Z_n|^2),$$

and correspondingly

$$Z(t) := X^{(1)}(t) - X^{(2)}(t), \quad u(t) := \mathbb{E}(|Z(t)|^2).$$

A direct consequence of inequality (4.21) is

$$\sup_{n \geq 0} u_n \leq u_0.$$

Applying Theorem 3.3(2), $u_n \leq v_n$, where v_n solves the following induction formula:

$$(\mathcal{D}^\alpha v)_n = -2\mu v_n, \quad v_0 = u_0 = \mathbb{E}(|Z_0|^2).$$

Let $v(t)$ solve the FODE, $D_c^\alpha v = -2\mu v$, $v(0) = u_0$. By Theorem 3.4, $|v_n - v(t_n)| \leq Ck^\alpha$.

Applying (4.10) and Lemma 4.3, one has

$$|u(t_n) - u_n| \leq (\sqrt{u_0} + C_1(T))\sqrt{\mathbb{E}(|Z_n - Z(t_n)|^2)} \leq Ck^{\alpha + H - 1}.$$

Hence, for all $n, nk \leq T$, it holds that

$$u(t_n) \leq v(t_n) + C(k^{\alpha + H - 1} + k^\alpha).$$

Taking $k \to 0$ then gives

$$(4.22) \qquad u(t) \leq v(t) \leq u_0 E_\alpha(-2\mu t^\alpha).$$

This inequality clearly implies the claim about the Wasserstein distance using (4.14). □

As in section 4.1, the overdamped GLE with fractional noise corresponds to

$$(4.23) \qquad \alpha = 2 - 2H, \ \sigma = \frac{\sqrt{2}}{\sqrt{\Gamma(2H+1)}}.$$

We guess that the limiting measure is the Gibbs measure $\pi(x) \propto \exp(-V(x))$. Rigorously justifying this seems challenging, and we leave it for the future.
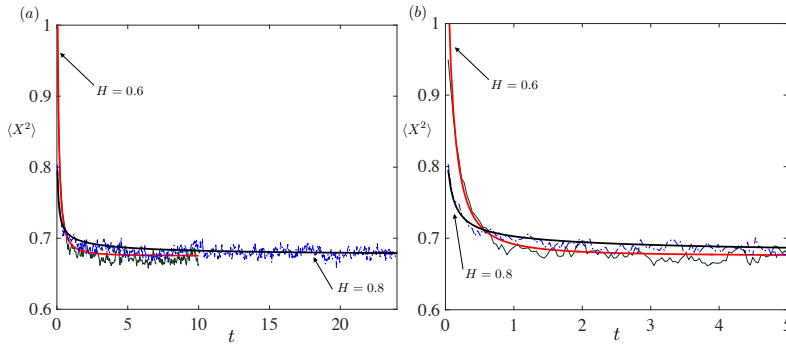
FIG. 1. *Means square distance of the FSDE with potential* (4.24) *for* $H = 0.6$ *and* $H = 0.8$.

4.3. A numerical simulation. In this section, we apply the implicit numerical scheme (4.16) to a one-dimensional FSDE example with

$$(4.24) \qquad\qquad V(x) = \frac{1}{4}x^4.$$

We choose $\alpha = 2 - 2H$ so that $G \sim \beta_H B_{1-H}$ as we have mentioned. This potential is convex but not strongly convex, and the corresponding force $-\nabla V(x) = -x^3$ is nonlinear. Justification of convergence to a limiting measure is by no means easy, not to mention whether the limiting measure is the Gibbs distribution

$$\pi(x) \propto \exp\left(-\frac{1}{4}x^4\right).$$

Figure 1 shows the trend of mean square distance $\mathbb{E}X^2 =: \langle X^2 \rangle$ with $X(0) = X_0 = 1$. Figure 1(b) enlarges the portion $t \in [0, 5]$ of Figure 1(a). If the distribution of the FSDE converges to the Gibbs measure, then $\mathbb{E}X^2 \to 0.675$. In the figures, the green curve is the numerical simulation for $H = 0.6$ with $k = 5/2^7 \approx 0.391$ and $N_s := 10^4$ samples, while the red solid curve is the fitting curve $0.675 + 0.02(0.12 + t)^{4H-4}$. The blue curve is the numerical simulation for $H = 0.8$ with the same k and number of samples, while the black curve is the fitting curve $0.675 + 0.015(0.05 + t)^{4H-4} + 0.04(3 + t)^{4H-4}$. We use $4H - 4$ power to fit because the variance of X in the linear forcing case has been shown to converge with rate t^{4H-4} in [3]. In this sense, the rate in Theorem 4.2 might not be optimal.

When $H = 0.6$ and $\alpha = 0.8$, the mean square distance already converges after $t = 5$ or so. However, for $H = 0.8$ or $\alpha = 0.4$, the mean square distance has a rapid drop at the early stage, but then the memory lingers for long time so that the convergence is very slow. This is also the case for normal FODE [37]. In fact, in Figure 2, we plot the empirical density versus the Gibbs measure (black line). When $\alpha = 2 - 2H = 0.8$, the distribution is already close to the desired Gibbs distribution at $t = 7.5$. However, when $\alpha = 2 - 2H = 0.4$, the distribution is roughly like the Gibbs distribution but still has some difference even at $t = 20$. In fact, this kind of distribution stays for long time ($t \sim 10^2$). We expect that when t is very large, it can be close to the Gibbs distribution as in [23]. We choose not to do the simulation for $H = 0.8$ and $t \sim 10^3$ since the complexity for obtaining a sample path is $O(N^2)$ and the simulation is expensive (we need 10^4 samples). For long time simulation when H is close to 1 (like $H = 0.8$ for $T \gtrsim 100$), it is good to adopt the fast scheme
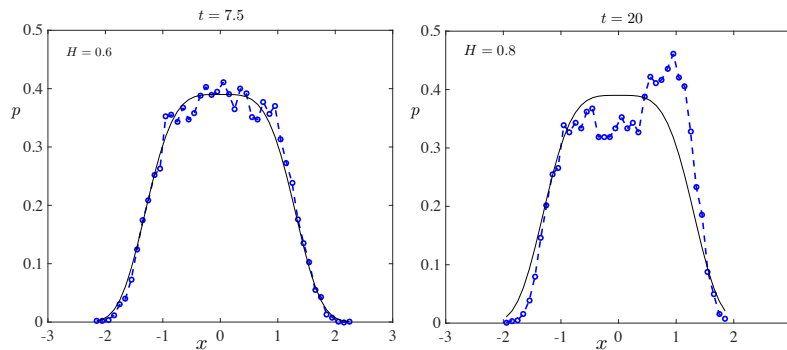
FIG. 2. *Empirical density. When $\alpha = 2 - 2H = 0.8$, the convergence is obtained after a reasonable time; when $\alpha = 2 - 2H = 0.4$, an intermediate distribution lingers for long time before converging to the final equilibrium.*

in [23]. However, our scheme here is appropriate for dissipative problems due to its good stability properties and can be used to analyze the time continuous problems compared with the one in [23].

5. Time fractional gradient flows. In this section, we investigate the time fractional gradient flows using the implicit scheme based on our discretization (3.4) and establish the error estimates of the numerical scheme. We will use $\langle \cdot, \cdot \rangle$ to denote the inner product in H and $\| \cdot \|$ to denote the norm on H. We will focus on convex functionals ϕ.

Assumption 5.1. Suppose the functional ϕ is lower semicontinuous and convex and $\inf_{u \in \mathbb{R}^d} \phi(u) > -\infty$.

Remark 5.2. All the claims in section 5 regarding convex functionals have analogies for λ-convex functionals (i.e., $\exists \lambda \geq 0$, $u \mapsto \phi(u) + \frac{\lambda}{2}|u|^2$ is convex; of course, the proof is more involved). Considering clarity of presentation, we only focus on convex functionals.

The Frechét subdifferential of convex ϕ satisfies

(5.1) $\qquad \xi \in \partial\phi(u) \Leftrightarrow \partial\phi(v) \neq \emptyset \quad \forall w \in H, \phi(w) - \phi(v) - \langle \xi, w - v \rangle \geq 0.$

The following strong-weak closure property is a straightforward consequence of this characterization.

LEMMA 5.3. *Suppose Assumption* 5.1 *holds. Assume sequences $\{\xi_n\}$ and $\{u_n\}$ satisfy $\xi_n \in \partial\phi(u_n)$ for all n, $u_n \to u$ strongly, and that $\xi_n \rightharpoonup \xi$ weakly. Then $\xi \in \partial\phi(u)$.*

Fix time $T > 0$. Similarly as in [52, Definition 2.2], we define as follows.

DEFINITION 5.4. *$u \in L^1_{\mathrm{loc}}([0,T), H)$ is called a strong solution to (1.5) if (i) $D^\alpha_c u$ is locally integrable on $[0, T)$; (ii) $\lim_{t \to 0^+} \frac{1}{t} \int_0^t \|u(s) - u_0\| \, ds = 0$; (iii) for almost every $t \in [0, T)$, we have $D^\alpha_c u \in -\partial\phi(u)$.*

Remark 5.5. If $\alpha = 1$, the local integrability of the distributional derivative Du clearly implies that u is absolutely continuous on $[0, T_1]$ for any $T_1 \in (0, T)$. The conditions (i)–(iii) in [52, Definition 2.2] are automatically satisfied. For $\alpha \in (0, 1)$, imposing $D^\alpha_c u \in L^1_{\mathrm{loc}}[0, T)$ does not ensure the uniqueness of u_0 (see [40]). To kill this ambiguity, we impose $\lim_{t \to 0^+} \frac{1}{t} \int_0^t \|u(s) - u_0\| \, ds = 0$.

We aim to approximate the solutions of (1.5) (though the existence is unclear at this point), following the method of De Giorgi [53, 54].

$$(5.2) \qquad U_n = \operatorname{argmin}\left(\frac{1}{2k^\alpha}\left(\sum_{j=1}^{n-1} c_j\|u - U_{n-j}\|^2 + c_n^n\|u - U_0\|^2\right) + \phi(u)\right).$$

Note that the functional on the right-hand side of (5.2) is the sum of a convex function and some quadratic functions. Then

$$(5.3) \qquad \partial\left(\frac{1}{2k^\alpha}\left(\sum_{j=1}^{n-1} c_j\|u - U_{n-j}\|^2 + c_n^n\|u - U_0\|^2\right) + \phi(u)\right)$$
$$= k^{-\alpha}\left(c_0 u - \left(\sum_{j=1}^{n-1} c_j U_{n-j} + c_n^n U_0\right)\right) + \partial\phi(u),$$

and the numerical solution satisfies

$$(5.4) \qquad -\xi_n := (\mathcal{D}^\alpha U)_n \in -\partial\phi(U_n).$$

Motivated by the proof of Theorem 3.4, we consider the set of time steps

$$(5.5) \qquad E_T = \{k > 0 : k = 2^{-m}T, \ m \in \mathbb{N}\}.$$

The following results from (5.3) and the strong convexity of the functional in (5.2) (proof omitted).

LEMMA 5.6. *Suppose Assumption* 5.1 *holds. Then, for sufficiently small* $k \in E_T$, *the discrete schemes* (5.2) *and* (5.4) *are equivalent and they have a unique solution* $\{U^n\}$.

5.1. Properties of the discrete solutions. Consider the solution given by (5.4). Define the function $V(t)$ such that

$$(5.6) \qquad V(t) := -\xi_n, \ t \in (t_{n-1}, t_n].$$

Using the function V, define a natural continuous version interpolation of U_n by

$$(5.7) \qquad U(t) = U_0 + \frac{1}{\Gamma(\alpha)}\int_0^t (t - s)^{\alpha-1}V(s)\,ds$$

with $U(t_m) = U_m$. This continuous interplocation justifies why the discretization (3.4) is suitable for (1.5). By (2.4) and the generalized definition (Definition 2.1), one has

$$(5.8) \qquad D_c^\alpha U(t) = V(t).$$

LEMMA 5.7. *Assume Assumption* 5.1. *For* $k \in E_T$ *small enough, it holds that*

$$(5.9) \qquad \sup_{n:nk\leq T} |\phi(U_n)| \leq C(U_0, T)$$

and that

$$(5.10) \qquad \sup_{t\leq T} \frac{1}{\Gamma(\alpha)}\int_0^t (t - s)^{\alpha-1}\|V(s)\|^2\,ds \leq C(U_0, T).$$

Proof. Paring with $\xi_n = -(\mathcal{D}^\alpha U)_n \in \partial\phi(U_n)$, and noting $\langle \xi_n, U_n - U_j \rangle \geq \phi(U_n) - \phi(U_j)$, one has

$$-\|\xi_n\|^2 = k^{-\alpha}\left(\sum_{j=1}^{n-1} c_j\langle\xi_n, U_n - U_{n-j}\rangle + c_n^n\langle\xi_n, U_n - U_0\rangle\right) \geq (\mathcal{D}^\alpha\phi(U))_n.$$

Using the equivalence between (2.17) and (2.12), and nonnegativity of $\{a_m\}$, one has

$$\phi(U_n) - \phi(U_0) \leq -k^\alpha \sum_{m=1}^{n} a_{n-m}\|\xi_m\|^2 = -\frac{1}{\Gamma(\alpha)}\int_0^{t_n}(t_n - s)^{\alpha-1}\|V(s)\|^2\,ds.$$

The first claim and the second claim with $t = t_n$ hold. For general $t \in (t_{n-1}, t_n)$, the following trivial observation with the result just proved yields the claim in the statement of the lemma.

$$\int_0^t (t-s)^{\alpha-1}\|V(s)\|^2\,ds$$

$$= \int_0^{t_{n-1}}(t-s)^{\alpha-1}\|V(s)\|^2\,ds + \|\xi_n\|^2\int_{t_{n-1}}^t(t-s)^{\alpha-1}\,ds$$

$$\leq \int_0^{t_{n-1}}(t_{n-1}-s)^{\alpha-1}\|V(s)\|^2\,ds + \|\xi_n\|^2\int_{t_{n-1}}^{t_n}(t_n-s)^{\alpha-1}\,ds$$

$$\leq \int_0^{t_{n-1}}(t_{n-1}-s)^{\alpha-1}\|V(s)\|^2\,ds + \int_0^{t_n}(t_n-s)^{\alpha-1}\|V(s)\|^2\,ds. \qquad \square$$

Now, we compare the numerical solutions with different time steps.

LEMMA 5.8. *There exists $C(U_0, T)$ independent of k such that when k is small enough,*

$$(5.11) \qquad \|U(t) - U(t+\delta)\| \leq C(U_0, T)|\delta|^{\alpha/2} \quad \text{if } \max(t, t+\delta) \leq T.$$

Let $U_i(t)$ be two functions given by (5.7) for step sizes $k_i \in E_T$ ($i = 1, 2$). Then,

$$(5.12) \qquad \sup_{0 \leq t \leq T}\|U_1(t) - U_2(t)\|^2 \leq C(U_0, T)(k_1^{\alpha/2} + k_2^{\alpha/2}).$$

Proof. Without loss of generality, we assume $\delta > 0$. Then, by (5.7),

$$\|U(t) - U(t+\delta)\| \leq \frac{1}{\Gamma(\alpha)}\left(\int_0^t[(t-s)^{\alpha-1} - (t+\delta-s)^{\alpha-1}]\|V(s)\|\,ds\right.$$

$$\left. + \int_t^{t+\delta}(t+\delta-s)^{\alpha-1}\|V(s)\|\,ds\right) =: I_1 + I_2.$$

The second term is estimated easily by

$$I_2 \leq \frac{1}{\Gamma(\alpha)}\left(\int_t^{t+\delta}\|V(s)\|^2(t+\delta-s)^{\alpha-1}\,ds\right)^{1/2}\left(\int_t^{t+\delta}(t+\delta-s)^{\alpha-1}\,ds\right)^{1/2}$$

$$\leq C\delta^{\alpha/2}.$$

For the first term I_1, we have by the Hölder inequality

$$(5.13) \qquad I_1 \leq \frac{1}{\Gamma(\alpha)} \left(\int_0^t (t-s)^{\alpha-1} \left(1 - \left(\frac{t+\delta-s}{t-s} \right)^{\alpha-1} \right)^2 ds \right)^{1/2}$$

$$\times \left(\int_0^t (t-s)^{\alpha-1} \|V(s)\|^2 \, ds \right)^{1/2}.$$

Clearly,

(5.14)
$$\int_0^t (t-s)^{\alpha-1} \left(1 - \left(\frac{t+\delta-s}{t-s} \right)^{\alpha-1} \right)^2 ds \leq \int_0^t ((t-s)^{\alpha-1} - (t+\delta-s)^{\alpha-1}) \, ds$$
$$\leq C\delta^\alpha.$$

The claim follows.

To compare the numerical solutions with steps $k_i, i = 1, 2$ (V_i, $\xi_n^{(i)}$, $i = 1, 2$, similarly defined), we fix $t \in [0, T]$. Then, there exist n_1 and n_2 such that $t \in ((n_1 - 1)k_1, n_1 k_1] \cap ((n_2 - 2)k_2, n_2 k_2]$ and such that $V_i(t) = -\xi_{n_i}^{(i)}$. Denote

$$\Delta_i(t) := U_{n_i}^{(i)} - U_i(t), \quad i = 1, 2.$$

By the definition of $U_i(t)$ and convexity of ϕ, one has

$$(5.15) \qquad \begin{aligned} \langle D_c^\gamma U_1 - D_c^\gamma U_2, U_1 - U_2 \rangle &= \langle V_1(t) - V_2(t), U_1(t) - U_2(t) \rangle \\ &= -\langle \xi_{n_1}^{(1)} - \xi_{n_2}^{(2)}, U_{n_1}^{(1)} - U_{n_2}^{(2)} \rangle + R(t) \leq R(t), \end{aligned}$$

where $\langle \xi_{n_1}^{(1)} - \xi_{n_2}^{(2)}, U_{n_1}^{(1)} - U_{n_2}^{(2)} \rangle \geq 0$ by convexity of ϕ and

$$R(t) = -\langle V_1(t) - V_2(t), \Delta_1(t) \rangle + \langle V_1(t) - V_2(t), \Delta_2(t) \rangle.$$

It suffices to estimate $\Delta_i(t)$. We take $i = 1$ as the example. By the definition of $U_1(t)$,

$$\Gamma(\alpha)\Delta_1(t) = -\int_t^{n_1 k_1} (n_1 k_1 - s)^{\alpha-1} \xi_{n_1}^{(1)} \, ds + \int_0^t [(n_1 k_1 - s)^{\alpha-1} - (t-s)^{\alpha-1}] V_1(s) \, ds$$
$$=: I_1^1 + I_1^2.$$

The terms corresponding to I_1^1 are controlled by (noting $0 < n_1 k_1 - t \leq k_1$)

$$\begin{aligned} |\langle V_1(t) - V_2(t), I_1^1 \rangle| &= |\langle \xi_{n_1}^{(1)} - \xi_{n_2}^{(2)}, I_1^1 \rangle| \\ &= \frac{(n_1 k_1 - t)^\alpha}{\Gamma(1+\alpha)} |\langle \xi_{n_1}^{(1)} - \xi_{n_2}^{(2)}, \xi_{n_1}^{(1)} \rangle| \\ &\leq C k_1^\alpha (\|V_1(t)\|^2 + \|V_2(t)\|^2). \end{aligned}$$

The terms corresponding to I_1^2 can be estimated similarly as in (5.13)–(5.14),

$$\begin{aligned} |\langle \xi_{n_1}^{(1)} - \xi_{n_2}^{(2)}, I_1^2 \rangle| &\leq \frac{1}{\Gamma(\alpha)} \|V_1(t) - V_2(t)\| \int_0^t \left[(t-s)^{\alpha-1} - (n_1 k_1 - s)^{\alpha-1} \right] \|V_1(s)\| \, ds \\ &\leq C \|V_1(t) - V_2(t)\| k_1^{\alpha/2}. \end{aligned}$$

By the explicit formula (5.7), $U_i(t)$ is absolutely continuous. Proposition 3.11 in [5] can be easily generalized to show that

$$\frac{1}{2} D_c^\alpha \|U_1(t) - U_2(t)\|^2 \leq \langle D_c^\alpha(U_1 - U_2), U_1(t) - U_2(t) \rangle.$$

Overall,

$$D_c^\alpha(\|U_1(t) - U_2(t)\|^2) \leq C(k_1^{\alpha/2} + k_2^{\alpha/2})(\|V_1(t)\|^2 + \|V_2(t)\|^2).$$

Lemma 5.7 then yields the result. □

5.2. Well-posedness and numerical error estimates. In this subsection, we establish the existence and uniqueness of the time fractional gradient flow under some assumptions and give the error estimate of the numerical scheme. Besides Assumption 5.1, we also need a certain regularity property of the subdifferential mapping $\partial\phi$.

Assumption 5.9. Let Assumption 5.1 hold. Moreover, $v_n \to v$ strongly implies any sequence $\{\xi_n\}$ with $\xi_n \in \partial\phi(v_n)$ converges weakly to some $\xi \in H$.

THEOREM 5.10. *Suppose Assumption* 5.9 *holds. For any $T > 0$, the fractional gradient flow* (1.5) *has a unique strong solution u on $[0, T)$ in the sense of Definition* (5.4). *The strong solution is Hölder continuous on $[0, T)$:*

$$(5.16) \qquad \|u(t + \delta) - u(t)\| \leq C|\delta|^{\alpha/2}.$$

Besides, we have the following error estimates for the numerical solution (5.2):

$$(5.17) \qquad \sup_{n:nk \leq T} \|U_n - u(t_n)\| \leq C(U_0, T)k^{\alpha/4}.$$

Proof of Theorem 5.10. By Lemma 5.8, the family $\{U_i(\cdot) : i \in E_T\}$ is a Cauchy sequence in $C([0, T], H)$. Therefore, there exists $u(\cdot) \in C([0, T]; H)$ such that $U_i(\cdot)$ converges to u in $C([0, T]; H)$.

Consider the piecewise constant interpolation of the numerical fractional derivative, $\{V_i(t)\}$. Lemma 5.7 implies that $V_i(t) \in L^2(0, T)$ with the $L^2(0, T)$ norm uniformly bounded. Then, there is a further subsequence so that we have the weak convergent sequence $V_i(t) \rightharpoonup v(t)$ in $L^2(0, T; H)$, with the estimate

$$\sup_{t \leq T} \frac{1}{\Gamma(\alpha)} \int_0^t (t - s)^{\alpha-1} \|v(s)\|^2 \, ds \leq C(U_0, T).$$

With this convergence in hand, $u(t) = u_0 + \frac{1}{\Gamma(\alpha)} \int_0^t (t - s)^{\alpha-1} v(s) \, ds$. Hence, in the distributional sense,

$$(5.18) \qquad D_c^\alpha u = v(t).$$

Similarly as in the proof of Lemma 5.8, $u(\cdot)$ is Hölder continuous with order at least $\alpha/2$.

By Lemma 5.7, we have found that $\phi(U_n)$ is uniformly bounded. Now we consider piecewise linear interpolation of $\{U_n\}$, denoted by $\bar{U}_i(\cdot)$,

$$\bar{U}_i(t) = U_n^{(i)}, \quad t \in (t_{n-1}, t_n].$$

Then, $V_i(t) \in -\partial\phi(\bar{U}_i(t))$ for all $t \in [0, T)$, for all i. By the uniform Hölder continuity of $U_i(\cdot)$ in Lemma 5.8 and $\bar{U}_i(t) = U_i(t_n)$ for $t \in (t_{n-1}, t_n]$, we find that for all t, $\bar{U}_i(t)$ converges strongly to $u(t)$. With Assumption 5.9, we find that for all t,

$$(5.19) \qquad V_i(t) \rightharpoonup \bar{v}(t), \ H,$$

so that

$$(5.20) \qquad \bar{v}(t) \in -\partial\phi(u(t)),$$

by the weak-strong closure property.

Hence, for any $w \in H$ so that $w 1_{[0,T]} \in L^2(0,T;H)$, we find that $\langle V_i(t), w \rangle$ converges in $L^2(0,T;\mathbb{R})$ to $\langle v(t), w \rangle$. By (5.19), we also have $\langle V_i(t), w \rangle \to \langle \bar{v}(t), w \rangle$ for all $t \in [0,T]$. Consequently, for a.e. $t \in [0,T]$, we have

$$\langle v(t), w \rangle = \langle \bar{v}(t), w \rangle.$$

Since H is separable, we can find a basis consisting of countable elements $\{w_n\}$. Consequently, we have

$$\langle v(t) - \bar{v}(t), w_n \rangle = 0 \quad \forall n, a.e. t \in [0,T]$$

Moreover, $v - \bar{v} \in L^2(0,T;H')$ (where H' means H equipped with the weak star topology), we have then

$$\int_0^T \langle v(t) - \bar{v}(t), w(t) \rangle \, dt = 0$$

for $w(t)$ to be simple functions, and then $L^2(0,T;H)$ functions. Hence, $v(t) = \bar{v}(t)$ for a.e. $t \in [0,T]$. By (5.20) and (5.18), we find that $u(t)$ is a strong solution under Definition 5.4. Taking limit $k_2 \to 0$ in $\|U_1(t) - U_2(t)\|^2 \leq C(U_0,T)(k_1^{\alpha/2} + k_2^{\alpha/2})$ yields $\|U_1(t) - u(t)\|^2 \leq C(U_0,T)k_1^{\alpha/2}$.

For the uniqueness, suppose we have two strong solutions $u_i(t), i = 1,2$, such that $-\xi_i(t) := D_c^\alpha u_i(t) \in -\partial\phi(u_i(t))$. Then, $\langle u_1 - u_2, D_c^\alpha(u_1 - u_2) \rangle \leq 0$. Some regularization procedure can yield $\frac{1}{2}D_c^\alpha\|u_1 - u_2\|^2 \leq \langle u_1 - u_2, D_c^\alpha(u_1 - u_2) \rangle \leq 0$. This then implies uniqueness. □

As a concluding remark, the orders of estimates in Theorem 5.10 are not optimal. If one can show that ξ_n is bounded, then one can improve the orders. Last, we give a quick glimpse of the case $H = \mathbb{R}^d$, $\phi \in C^1(\mathbb{R}^d)$ (instead of requiring $\nabla\phi$ to be Lipschitz as in [5, 37]) so that (1.5) becomes the FODE:

$$(5.21) \qquad\qquad\qquad D_c^\alpha u = -\nabla\phi(u).$$

The following asymptotic behavior holds when ϕ is strongly convex.

PROPOSITION 5.11. *Assume that $\phi \in C^1(\mathbb{R}^d)$ and $\phi - \frac{\mu}{2}|u|^2$ is convex for some $\mu > 0$. Let u^* be the global minimizer of ϕ. Then, for some C depending on u_0, μ,*

$$\phi(u(t)) - \phi(u^*) \leq (\phi(u_0) - \phi(u^*))E_\alpha\left(-Ct^\alpha\right).$$

Moreover, $|u(t) - u^*| \leq C(1 + t)^{-\alpha}$.

Proof. Consider the implicit scheme

$$(\mathcal{D}^\alpha U)_n = -\nabla\phi(U_n).$$

Using the first claim in Theorem 3.3, one has

$$(\mathcal{D}^\alpha\phi(U))_n \leq -|\nabla\phi(U_n)|^2 \leq 0.$$

Hence $\phi(U_n)$ is bounded. Since ϕ is strongly convex and thus $\lim_{R\to\infty} \inf_{|u|\geq R} \phi(u) = +\infty$, $\{U_n\}$ is in a compact domain K that only depends on u_0. By Theorem 3.3 and the Polyak–Lojasiewicz inequality ($|\nabla\phi(x)|^2 \geq 2\mu(\phi(x) - \phi(u^*))$), one has

$$(\mathcal{D}^\alpha(\phi(U) - \phi(u^*)))_n = (\mathcal{D}^\alpha\phi(U))_n \leq -|\nabla\phi(U_n)|^2 \leq -2\mu(\phi(U_n) - \phi(u^*)).$$

The Polyak–Lojasiewicz inequality is obtained by

$$\phi(y) \geq \phi(u) + \nabla\phi(u) \cdot (y - u) + \frac{\mu}{2}|y - u|^2 \geq \phi(u) - \frac{1}{2\mu}|\nabla\phi(u)|^2.$$

By the second claim in Theorem 3.3 and the third claim in Theorem 3.4, it holds for any $nk \leq T$ that

$$\phi(U_n) - \phi(u^*) \leq (\phi(u_0) - \phi(u^*))E_\alpha(-2\mu(nk)^\alpha) + o(k).$$

Taking $k \to 0$ and by Theorem 5.10 (convergence and continuity of $u(t)$), one thus has for any $t \leq T$:

$$\phi(u(t)) - \phi(u^*) \leq (\phi(u_0) - \phi(u^*))E_\alpha(-2\mu t^\alpha).$$

Since T is arbitrary, the first claim is true for all t.

Similarly, $(\mathcal{D}^\alpha|U - u^*|)_n \leq \frac{1}{|U_n - u^*|}\langle U_n - u^*, -\nabla\phi(U_n)\rangle \leq -\mu|U_n - u^*|$. Theorem 5.10 allows us to take $k \to 0$ to obtain $|u(t) - u^*| \leq |u_0 - u^*|E_\alpha(-\mu t^\alpha)$. Since $E_\alpha(-s) \sim C_1 s^{-1}$ as $s \to \infty$, the second claim follows. □

REFERENCES

[1] A. A. KILBAS, H. M. SRIVASTAVA, AND J. J. TRUJILLO, *Theory and Applications of Fractional Differential Equations*, North-Holland Math. Stud. 204, Elsevier, Amsterdam, 2006.

[2] K. DIETHELM, *The Analysis of Fractional Differential Equations: An Application-Oriented Exposition Using Differential Operators of Caputo Type*, Springer, New York, 2010.

[3] L. LI, J.-G. LIU, AND J. LU, *Fractional stochastic differential equations satisfying fluctuation-dissipation theorem*, J. Stat. Phys., 169 (2017), pp. 316–339.

[4] M. CAPUTO, *Linear models of dissipation whose Q is almost frequency independent–II*, Geophys. J. Internat., 13 (1967), pp. 529–539.

[5] L. LI AND J.-G. LIU, *A generalized definition of Caputo derivatives and its application to fractional ODEs*, SIAM J. Math. Anal., 50 (2018), pp. 2867–2900.

[6] S. C. KOU AND X S. XIE, *Generalized Langevin equation with fractional Gaussian noise: Subdiffusion within a single protein molecule*, Phys. Rev. Lett., 93 (2004), 180603.

[7] S BUROV AND E BARKAI, *Critical exponent of the fractional Langevin equation*, Phys. Rev. Lett., 100 (2008), 070601.

[8] A. TALONI, A. CHECHKIN, AND J. KLAFTER, *Generalized elastic model yields a fractional Langevin equation description*, Phys. Rev. Lett., 104 (2010), 160602.

[9] V. VERGARA AND R. ZACHER, *Lyapunov functions and convergence to steady state for differential equations of fractional order*, Math. Z., 259 (2008), pp. 287–309.

[10] T. SANDEV, *Generalized Langevin equation and the Prabhakar derivative*, Mathematics, 5 (2017), 66.

[11] A. LIEMERT, T. SANDEV, AND H. KANTZ, *Generalized Langevin equation with tempered memory kernel*, Phys. A, 466 (2017), pp. 356–369.

[12] Y. CHEN, X. WANG, AND W. DENG, *Tempered fractional Langevin–Brownian motion with inverse β-stable subordinator*, J. Phys. A, 51 (2018), 495001.

[13] D. MOLINA-GARCIA, T. SANDEV, H. SAFDARI, G. PAGNINI, A. CHECHKIN, AND R. METZLER, *Crossover from anomalous to normal diffusion: Truncated power-law noise correlations and applications to dynamics in lipid bilayers*, New J. Phys., 20 (2018), 103027.

[14] R. KUBO, *The fluctuation-dissipation theorem*, Rep. Prog. Phys., 29 (1966), pp. 255–284.

[15] H. MORI, *A continued-fraction representation of the time-correlation functions*, Prog. Theor. Phys., 34 (1965), pp. 399–416.

[16] R. ZWANZIG, *Nonlinear generalized Langevin equations*, J. Stat. Phys., 9 (1973), pp. 215–220.

[17] C. HIJÓN, P. ESPAÑOL, E. VANDEN-EIJNDEN, AND R. DELGADO-BUSCALIONI, *Mori–Zwanzig formalism as a practical computational tool*, Faraday Discussions, 144 (2010), pp. 301–322.

[18] Z. LI, H. S. LEE, E. DARVE, AND G. E. KARNIADAKIS, *Computing the non-Markovian coarse-grained interactions derived from the Mori–Zwanzig formalism in molecular systems: Application to polymer melts*, J. Chem. Phys., 146 (2017), 014104.

[19] B. LEIMKUHLER AND M. SACHS, *Ergodic Properties of Quasi-Markovian Generalized Langevin Equations with Configuration Dependent Noise*, preprint, arXiv:1804.04029, 2018.

[20] S. C. KOU, *Stochastic modeling in nanoscale biophysics: Subdiffusion within proteins*, Ann. Appl. Stat., (2008), pp. 501–535.

[21] W. DENG AND E. BARKAI, *Ergodic properties of fractional Brownian-Langevin motion*, Phys. Rev. E, 79 (2009), 011112.

[22] D. NUALART, *Fractional Brownian motion: Stochastic calculus and applications*, in Proceedings of the International Congress of Mathematicians, Vol. 3, 2006, pp. 1541–1562.

[23] D. FANG AND L. LI, *Numerical Study of Ergodicity for the Overdamped Generalized Langevin Equation with Fractional Noise*, preprint, arXiv:1810.03810, 2018.

[24] L.-Q. CHEN, *Phase-field models for microstructure evolution*, Annu. Rev. Materials Research, 32 (2002), pp. 113–140.

[25] C. LIU AND J. SHEN, *A phase field model for the mixture of two incompressible fluids and its approximation by a Fourier-spectral method*, Phys. D, 179 (2003), pp. 211–228.

[26] M. G. CRANDALL AND J. A. NOHEL, *An abstract functional differential equation and a related nonlinear Volterra equation*, Israel J. Math., 29 (1978), pp. 313–328.

[27] PH CLÉMENT AND J. A. NOHEL, *Asymptotic behavior of solutions of nonlinear Volterra equations with completely positive kernels*, SIAM J. Math. Anal., 12 (1981), pp. 514–535.

[28] C. H. LUBICH, *Discretized fractional calculus*, SIAM J. Math. Anal., 17 (1986), pp. 704–719.

[29] Y. LIN AND C. XU, *Finite difference/spectral approximations for the time-fractional diffusion equation*, J. Comput. Phys., 225 (2007), pp. 1533–1552.

[30] X. LI AND C. XU, *A space-time spectral method for the time fractional diffusion equation*, SIAM J. Numer. Anal., 47 (2009), pp. 2108–2131.

[31] R. H. NOCHETTO, E. OTAROLA, AND A. J. SALGADO, *A PDE approach to space-time fractional parabolic problems*, SIAM J. Numer. Anal., 54 (2016), pp. 848–873.

[32] B. JIN, B. LI, AND Z. ZHOU, *Correction of high-order BDF convolution quadrature for fractional evolution equations*, SIAM J. Sci. Comput., 39 (2017), pp. A3129–A3152.

[33] B. JIN, B. LI, AND Z. ZHOU, *Discrete maximal regularity of time-stepping schemes for fractional evolution equations*, Numer. Math., 138 (2018), pp. 101–131.

[34] A. ATANGANA AND K. M. OWOLABI, *New numerical approach for fractional differential equations*, Math. Model. Natural Phenomena, 13 (2018).

[35] C.-T. SHENG AND J. SHEN, *A space-time spectral Petrov-Galerkin spectral method for time fractional diffusion equation*, Int. J. Numer. Anal. Model., to appear.

[36] J. SHEN AND C.-T. SHENG, *Spectral Methods for Fractional Differential Equations using Generalized Jacobi Functions*, Handbook of Fractional PDEs, to appear.

[37] Y. FENG, L. LI, J.-G. LIU, AND X. XU, *Continuous and discrete one dimensional autonomous fractional ODEs*, Discrete Cont. Dyn. Syst. Ser. B, 23 (2018), pp. 3109–3135.

[38] L. LI AND J.-G. LIU, *A note on deconvolution with completely monotone sequences and discrete fractional calculus*, Quart. Appl. Math., 76 (2018), pp. 189–198.

[39] K. M. OWOLABI, *Mathematical modelling and analysis of two-component system with Caputo fractional derivative order*, Chaos Solitons Fractals, 103 (2017), pp. 544–554.

[40] L. LI AND J.-G. LIU, *Some compactness criteria for weak solutions of time fractional PDEs*, SIAM J. Math. Anal., 50 (2018), pp. 3963–3995.

[41] K. DIETHELM AND N. J. FORD, *Analysis of fractional differential equations*, J. Math. Anal. Appl., 265 (2002), pp. 229–248.

[42] K. DIETHELM, *An algorithm for the numerical solution of differential equations of fractional order*, Electron. Trans. Numer. Anal, 5 (1997), pp. 1–6.

[43] J.-G. LIU, Z. MA, AND Z. ZHOU, *Explicit and implicit TVD schemes for conservation laws with Caputo derivatives*, J. Sci. Comput., 72 (2017), pp. 291–313.

[44] L. GALEONE AND R. GARRAPPA, *Explicit methods for fractional differential equations and their stability properties*, J. Comput. Appl. Math., 228 (2009), pp. 548–560.

[45] D.V. WIDDER, *The Laplace Transform*, Princeton University Press, Princeton, NJ, 1941.

[46] P. FLAJOLET AND R. SEDGEWICK, *Analytic Combinatorics*, Cambridge University Press, Cambridge, UK, 2009.

[47] J. LIU AND R. PEGO, *On generating functions of Hausdorff moment sequences*, Trans. Amer. Math. Soc., 368 (2016), pp. 8499–8518.

[48] H. J. HAUBOLD, A. M. MATHAI, AND R. K. SAXENA, *Mittag-Leffler functions and their applications*, J. Appl. Math., 2011 (2011).

[49] S. G. SAMKO, A. A. KILBAS, AND O. I. MARICHEV, *Fractional Integrals and Derivatives: Theory and Applications*, Gordon and Breach, London, 1993.

[50] Y. FENG, L. LI, J.-G. LIU, AND X. XU, *A note on one-dimensional time fractional ODEs*, Appl. Math. Lett., 83 (2018), pp. 87–94.

[51] F. Santambrogio, *Optimal Transport for Applied Mathematicians*, Progr. Nonlinear Differential Equations Anal. 87, Birkhäuser, Basel, 2015.

[52] M. G. Crandall and T. M. Liggett, *Generation of semi-groups of nonlinear transformations on general Banach spaces*, Amer. J. Math., 93 (1971), pp. 265–298.

[53] E. De Giorgi, *New problems on minimizing movements*, in Boundary Value Problems for Partial Differential Equations and Applications, RMA Res. Notes Appl. Math. 29, Masson, Paris, 1993, pp. 81–98.

[54] R. Rossi and G. Savaré, *Gradient flows of non convex functionals in Hilbert spaces and applications*, ESAIM Control Optim. Calc. Var., 12 (2006), pp. 564–614.