Mathematical Modelling and Numerical Analysis
Modélisation Mathématique et Analyse Numérique

Will be set by the publisher

NUMERICAL APPROXIMATION OF STOCHASTIC TIME-FRACTIONAL DIFFUSION

Bangti Jin[1], Yubin Yan[2] and Zhi Zhou[3]

Abstract. We develop and analyze a numerical method for stochastic time-fractional diffusion driven by additive fractionally integrated Gaussian noise. The model involves two nonlocal terms in time, i.e., a Caputo fractional derivative of order $\alpha \in (0,1)$, and fractionally integrated Gaussian noise (with a Riemann-Liouville fractional integral of order $\gamma \in [0,1]$ in the front). The numerical scheme approximates the model in space by the standard Galerkin method with continuous piecewise linear finite elements and in time by the classical Grünwald-Letnikov method (for both Caputo fractional derivative and Riemann-Liouville fractional integral), and the noise by the L^2-projection. Sharp strong and weak convergence rates are established, using suitable nonsmooth data error estimates for the discrete solution operators for the deterministic inhomogeneous problem. One- and two-dimensional numerical results are presented to support the theoretical findings.

1991 Mathematics Subject Classification. 60H15, 60H35, 65M12.

March 9, 2019.

1. Introduction

In this work, we consider numerical methods for solving the following time-fractional diffusion equation driven by fractionally integrated additive Gaussian noise, with $0 < \alpha < 1$, $0 \leq \gamma \leq 1$:

$$\partial_t^\alpha u(t) + Au(t) = {}_0I_t^\gamma \dot{W}(t), \quad \forall 0 < t \leq T, \qquad \text{with } u(0) = u_0, \tag{1.1}$$

where the notation ${}_0I_t^\gamma v(t)$ denotes the Riemann-Liouville fractional integral of order $\gamma > 0$ of a function $v : [0,T] \to \mathbb{R}$ defined by

$${}_0I_t^\gamma v(t) = \frac{1}{\Gamma(\gamma)} \int_0^t (t-s)^{\gamma-1} v(s)\,\mathrm{d}s,$$

where $\Gamma(\cdot)$ denotes the Gamma function defined by $\Gamma(z) = \int_0^\infty s^{z-1}e^{-s}\mathrm{d}s$ (for $\Re z > 0$), with the convention ${}_0I_t^0 v(t) = v(t)$. For $\gamma \in (-1,0)$, ${}_0I_t^\gamma v$ denotes the Riemann-Liouville fractional derivative of order $-\gamma \in (0,1)$,

Keywords and phrases: stochastic time-fractional diffusion, Galerkin finite element method, Grünwald-Letnikov method, strong convergence, weak convergence

[1] Department of Computer Science, University College London, Gower Street, London WC1E 6BT, UK (`bangti.jin@gmail.com`, `b.jin@ucl.ac.uk`)

[2] Department of Mathematics, University of Chester, Chester CHI 4BJ, UK (`y.yan@chester.ac.uk`)

[3] Department of Applied Mathematics, The Hong Kong Polytechnic University, Kowloon, Hong Kong (`zhizhou@polyu.edu.hk`)

defined by ${}_0I_t^\gamma v := ({}_0I_t^{1+\gamma} v(t))'$. The notation $\partial_t^\alpha v(t)$, $0<\alpha<1$, denotes the Caputo fractional derivative of order α defined by [27, p. 91]

$$\partial_t^\alpha v(t) = {}_0I_t^{1-\alpha} v'(t).$$

In the model (1.1), the operator A denotes the negative Laplacian $-\Delta$ with a zero Dirichlet boundary condition in a convex polygonal domain $D \subset \mathbb{R}^d$ $(d = 1, 2, 3)$, with its domain $\mathcal{D}(A) = H_0^1(D) \cap H^2(D)$. The noise $W(t)$ is given by a Wiener process with a covariance operator Q on a filtered probability space $(\Omega, \mathcal{F}, \mathbb{P}, \{\mathcal{F}_t\}_{t\geq 0})$, with $\{\mathcal{F}_t\}_{t\geq 0}$ being an increasing filtration of σ-fields $\mathcal{F}_t \subset \mathcal{F}$, each of which contains all $(\mathcal{F}, \mathbb{P})$-null sets. Let $\mathbb{E}$ denote the expectation (with respect to $\mathbb{P}$). The function u_0 is an $\mathcal{F}_0$-measurable random variable, and belongs to $L^2(D)$ or its subspace. In order to ensure the well-posedness of problem (1.1) [12, pp. 1473–1474], we assume the following condition:

$$\alpha \in (0, 1), \quad \gamma \in [0, 1] \quad \text{and} \quad \alpha + \gamma > 1/2. \tag{1.2}$$

The deterministic counterpart of the model (1.1), commonly known as subdiffusion, has been extensively studied in the literature over the last few decades [25], due to its numerous applications in engineering, physics and biology [37]. The term ${}_0I_t^\gamma \dot{W}(t)$ in the model (1.1) is to describe random effects on transport of particles in medium with memory or particles subject to sticking and trapping [12]. The fractionally integrated noise ${}_0I_t^\gamma \dot{W}(t)$ reflects the fact that the internal energy depends also on the past random effects. In recent years, stochastic fractional diffusion, e.g., the model (1.1), has been very actively researched [5, 10–12, 32]. Chen et al [12] studied the L^2 theory of (1.1) in both divergence and non-divergence forms. Anh et al [5] discussed sufficient conditions for a Gaussian solution (in the mean-square sense) and derived temporal, spatial and spatiotemporal Hölder continuity of the solution. Chen [10] analyzed moments, Hölder continuity and intermittency of the solution for 1D nonlinear stochastic subdiffusion. Liu et al [32] analyzed the existence and uniqueness of the solution to (1.1) with fairly general quasi-linear elliptic operators; see also [17, 33] for further analytic results.

To the best of our knowledge, there seems no work on the numerical analysis of the stochastic time-fractional PDEs driven by fractionally integrated Gaussian noise, except in a few special cases (to be described below). It is precisely this gap that we aim at filling in the present work. Specifically, we develop a numerical scheme for problem (1.1), based on the standard Galerkin finite element method (FEM) with continuous linear finite elements in space, the classical Grünwald-Letnikov method (i.e., backward Euler convolution quadrature [24, 34, 35]) in time (for both Caputo fractional derivative and Riemann-Liouville fractional integral) and the L^2-projection of the noise, cf. (3.3). The scheme combines discretization techniques for subdiffusion [24] and stochastic heat equation [39], and it is easy to implement. We prove nearly sharp strong and weak convergence rates for the fully discrete approximation in Theorems 5.1 and 5.2, respectively, which represent the main theoretical contributions of the work.

The proofs uses well established stochastic error analysis tools from the SPDE community. The strong convergence analysis employs an operator theoretic approach, which was first developed in the work [39] for the stochastic heat equation and subsequently used in many works. In the analysis, one crucial ingredient is certain nonsmooth data error estimates for the solution operators associated with the deterministic inhomogeneous problem, i.e., the semi-discrete and fully solution operators; see $\bar{E}_h(t)$ in (4.4) and B_j in (4.5) in Section 4.1. Due to the presence of the fractional integral operator ${}_0I_t^\gamma$, such estimates differ greatly from that for standard subdiffusion, and are still unavailable. We employ Laplace transform and generating function [35] to derive the requisite nonsmooth data estimates, which represents the main technical novelty of this work. We refer interested readers to [13, 23, 24, 26, 36] for related works on nonsmooth data estimates for deterministic subdiffusion; see also the survey [25] and the references therein. For the weak convergence, we employ a powerful tool, Malliavin calculus, recently developed in [3]. This approach is more involved than the traditional approach based Kolmogorov's equation, but admits extensions to more complex problems, e.g., semilinear equations (see Remark 5.4 for further details). The technique in [3] relies on a new family of Sobolev-Malliavin spaces that capture the temporal integrability of the Malliavin derivative, and a new Burkholder type inequality in the dual norm of these Sobolev-Malliavin spaces. The challenge lies in deriving the error estimate in the dual norm of refined Sobolev-Malliavin spaces.

Table 1. Strong and weak convergence rates for the numerical scheme (3.3) with $u_0 = 0$ and trace class noise, and compared with the standard stochastic heat equation.

(α,γ)	strong
$(1,0)$	$O(h+\tau^{\frac{1}{2}-\epsilon})$ [39]
$\gamma<1/2$	$O(h^{2-\frac{1-2\gamma}{\alpha}-\epsilon}+\tau^{\min(1,\alpha+\gamma-\frac{1}{2}-\epsilon)})$
$\gamma>1/2$	$O(h^2+\tau^{\min(1,\alpha+\gamma-\frac{1}{2}-\epsilon)})$

(α,γ)	weak
$(1,0)$	$O(h^2+\tau^{1-\epsilon})$
$\gamma<\frac{1-\alpha}{2}$	$O(h^{4-\frac{2(1-2\gamma)}{\alpha}-\epsilon}+\tau^{\min(1,\alpha+\gamma-\epsilon)})$
$\gamma>\frac{1-\alpha}{2}$	$O(h^2+\tau^{\min(1,\alpha+\gamma-\epsilon)})$

Theorems 5.1 and 5.2 indicate that the fractionally integrated Gaussian noise ${}_0I_t^\gamma \dot W(t)$ induces convergence behaviors substantially different from that of stochastic diffusion. In particular, the fractional order γ can exert strong influence on both strong and weak convergence rates: dependent of the γ value, with h and τ being the mesh size and time step size, respectively, the spatial convergence rate may reach $O(h^2)$ and thus the temporal convergence rate $O(\tau)$ in both strong and weak sense, and the usual dichotomy of the weak temporal rate being twice the strong one is generally not valid; See Table 1 for convergence rates when the Wiener process $W(t)$ belongs to trace class, where the results for stochastic diffusion (i.e., $(\alpha,\gamma) = (1,0)$) are also given for the purpose of comparison. In the table, ϵ is an arbitrarily small positive constant. Further, the results for stochastic diffusion are recovered upon letting $\alpha \to 1^-$ and $\gamma \to 0^+$. These theoretical findings are fully supported by the extensive numerical experiments in one- and two-dimensions in Section 6.

To the best of our knowledge, there is no work on the numerical analysis of the general model (1.1), except some special cases, which we describe next. First, the model (1.1) represents the fractional analogue of the classical heat equation (but with a nonstandard noise term), and recovers the latter model for the special choice ($\alpha = 1, \gamma = 0$). Thus, naturally our results generalize that for the stochastic heat equation; see Table 1 for the case of trace class noise. The literature on the stochastic heat equation is vast. See, e.g., [1, 16, 39] for strong convergence, and, e.g., [4, 9, 15], for weak convergence, and interested readers are also referred to the surveys [22, 30] and references therein for further pointers to the vast literature. Our overall proof strategies are inspired by the works [39] and [3], respectively for the strong and weak convergence analysis. Second, the stochastic fractional model (1.1) was studied earlier for the case $\gamma = 1-\alpha$ [31], where the strong convergence of a discontinuous Galerkin method in time was analyzed; see also [21] for a related fractional-order model with white noise. We also refer interested readers to [2, 28, 29] for strong and weak convergence rates for numerical approximations of (linear/semilinear) stochastic Volterra equations, which involve more general kernels, but their smoothing properties differ from that for the model (1.1).

The rest of the paper is organized as follows. In Section 2, we give preliminaries on Wiener process and Malliavin calculus. In Section 3, we describe the numerical scheme, and in Section 4, we derive crucial nonsmooth data error estimates for deterministic subdiffusion. The strong and weak error estimates for approximations are given in Section 5. In Section 6, we discuss the implementation of the noise, and present numerical results to support the theoretical analysis. In Appendix A, we present some regularity results. Throughout, the notation c, with or without a subscript, denotes a generic constant, which may differ at each occurrence, but it is always independent of the mesh size h and the time step size τ. Further, $\epsilon > 0$ is always a small positive constant.

2. Preliminaries

In this section, we collect preliminary facts on Wiener process and Malliavin calculus.

2.1. Wiener process

Let $(U, \|\cdot\|_U, \langle\cdot,\cdot\rangle_U)$ and $(V, \|\cdot\|_V, \langle\cdot,\cdot\rangle_V)$ be separable Hilbert spaces. Let $L(U;V)$ be the Banach space of all bounded linear operators $U \to V$, and we denote $L(U) = L(U;U)$. $\mathcal{L}_2(U;V) \subset L(U;V)$ denotes the subspace of all Hilbert-Schmidt operators, with norms and inner products respectively given by

$$\|T\|^2_{\mathcal{L}_2(U;V)} = \sum_{j\in\mathbb{N}} \|Tu_j\|^2_V, \quad \langle S,T\rangle = \sum_{j\in\mathbb{N}} \langle Su_j, Tu_j\rangle_V.$$

Both are independent of the specific choice of orthonormal basis $\{u_j\}_{j\in\mathbb{N}}$.

Let $H = L^2(D)$ with the norm $\|\cdot\|$ and the inner product $(\cdot,\cdot)$. A Wiener process $W(t)$ with covariance Q may be characterized as follows. Let Q be a bounded, linear, selfadjoint, positive definite operator on H, with the pairs of eigenvalue and eigenfunction denoted by $\{(\gamma_\ell, e_\ell)\}_{\ell=1}^\infty$. Let $H_0 = Q^{\frac12}H$ be the Hilbert space endowed with the inner product $\langle u,v\rangle_{H_0} = (Q^{-\frac12}u, Q^{-\frac12}v)$. Let $\{\beta_\ell(t)\}_{\ell=1}^\infty$ be a sequence of real-valued independently and identically distributed (i.i.d.) Brownian motions. Then the series

$$W(t) = \sum_{\ell=1}^\infty \gamma_\ell^{\frac12} e_\ell \beta_\ell(t), \tag{2.1}$$

is a Wiener process with covariance operator Q. If Q is of trace class, i.e., $\sum_{\ell=1}^\infty \gamma_\ell < \infty$, then $W(t)$ is an H-valued process. If Q is not in trace class, e.g., $Q = I$, then $W(t)$ does not belong to H, in which case $W(t)$ is called a cylindrical Wiener process [14, Chapter 4].

The notation $\mathcal{L}_2^0 = \mathcal{L}_2(H_0;H)$ denotes the space of Hilbert-Schmidt operators from H_0 to H, i.e.,

$$\mathcal{L}_2^0 = \Big\{\Phi \in L(H) : \sum_{\ell=1}^\infty \|\Phi Q^{\frac12} e_\ell\|^2 < \infty\Big\},$$

with the norm $\|\cdot\|_{\mathcal{L}_2^0}$ defined by

$$\|\Phi\|_{\mathcal{L}_2^0} = \Big(\sum_{\ell=1}^\infty \|\Phi Q^{\frac12} e_\ell\|^2\Big)^{\frac12},$$

where $\{e_\ell\}_{\ell=1}^\infty$ is an orthonormal basis of H. This definition is independent of the choice of the basis. For any $\Phi \in L^2(0,T;\mathcal{L}_2^0)$, $\int_0^t \Phi(s)\mathrm{d}W(s)$ is well defined in the sense of stochastic integral [14, p. 95].

For any $p \ge 1$, we define the space of H-valued p-integrable random variables by

$$L^p(\Omega;H) = \Big\{v : \mathbb{E}\|v\|^p = \int_\Omega \|v(\omega)\|^p \,\mathrm{d}\mathbb{P}(\omega) < \infty\Big\},$$

with norm $\|v\|_{L^p(\Omega;H)} = (\mathbb{E}\|v\|^p)^{\frac1p}$. Similarly, we define the space $L^p(\Omega;V)$, for a Banach space V.

2.2. **Malliavin calculus**

In this part, we recall some concepts related to Malliavin derivatives of H-valued random variables, which will be employed for deriving weak convergence rates. More details can be found in [3,30]. Let $\mathcal{G}_p^\infty(\mathbb{R}^n;\mathbb{R})$ be the space of all infinitely many times Gâteaux differentiable mappings $\phi : \mathbb{R}^n \to \mathbb{R}$ such that ϕ and all its derivatives satisfy a polynomial bound. Let $\mathcal{B}(H;\mathbb{R})$ denote the Banach space of all bilinear mappings $b : H \times H \to \mathbb{R}$ equipped with the norm

$$\|b\|_{\mathcal{B}(H;\mathbb{R})} = \sup_{0\neq u_1,u_2\in H} \frac{|b\cdot(u_1,u_2)|}{\|u_1\|\|u_2\|}.$$

For any $\ell \ge 2$, let $\Phi \in \mathcal{G}_p^{2,\ell}(H;\mathbb{R})$, with

$$\mathcal{G}_p^{2,\ell}(H;\mathbb{R}) = \Big\{\Phi : H \to \mathbb{R},\ |\Phi|_{\mathcal{G}_p^{2,\ell}(H;\mathbb{R})} = \sup_{u\in H} \frac{\|\Phi^{(2)}(u)\|_{\mathcal{B}(H;\mathbb{R})}}{(1+\|u\|_H^{\ell-2})} < \infty\Big\}, \tag{2.2}$$

where $\Phi^{(2)}(u) \in \mathcal{B}(H;\mathbb{R})$ denotes the second-order Gâteaux derivative of $\Phi \in \mathcal{G}_p^{2,\ell}(H;\mathbb{R})$ at $u \in H$.

For $q \in [2,\infty]$, $S^q(\mathbb{R})$ denotes the class of smooth cylindrical random variables of the form

$$F = f\Big(\int_0^T \phi_1(s)\,\mathrm{d}W(s),\ldots,\int_0^T \phi_n(s)\,\mathrm{d}W(s)\Big),$$

where $f \in \mathcal{G}_p^\infty(\mathbb{R}^n;\mathbb{R})$ and $\{\phi_k\}_{k=1}^n \subset L^q(0,T;\mathcal{L}_2(H;\mathbb{R}))$, $n \in \mathbb{N}$. (Recall that the space $\mathcal{L}_2(H;\mathbb{R})$ is defined by $\mathcal{L}_2(H;\mathbb{R}) = \big\{\Phi \in L(H;\mathbb{R}) : \sum_{\ell=1}^\infty |\Phi Q^{\frac{1}{2}} e_\ell|_{\mathbb{R}}^2 < \infty\big\}$, where $L(H;\mathbb{R})$ denotes the Hilbert space of all bounded operators from H to $\mathbb{R}$ and $|\cdot|_{\mathbb{R}}$ denotes the Euclidean norm in $\mathbb{R}$.) For $F \in S^q(\mathbb{R})$, we define the Malliavin derivative by

$$DF(\sigma) = \sum_{j=1}^n \partial_j f\Big(\int_0^T \phi_1(s)\,\mathrm{d}W(s),\ldots,\int_0^T \phi_n(s)\,\mathrm{d}W(s)\Big) \otimes \phi_j(\sigma), \quad \sigma \in [0,T].$$

Note that $DF(\sigma)$ is an H_0-valued stochastic process.

Next, we recall the Malliavin derivative for H-valued random variables. Let $S^q(H)$ be the space of all H-valued random variables of the form $Y = \sum_{i=1}^m v_i \otimes F_i$ with $\{v_i\}_{i=1}^m \subset H$, $\{F_i\}_{i=1}^m \subset S^q(\mathbb{R})$, $m \in \mathbb{N}$, where $\otimes$ denotes the tensor product. Then the Malliavin derivative of $Y \in S^q(H)$ is defined by

$$DY(\sigma) = \sum_{i=1}^m v_i \otimes DF_i(\sigma).$$

Since $DF_i(\sigma)$ is an H_0-valued stochastic process, $DY(\sigma)$ is an $H \otimes H_0 = \mathcal{L}_2^0$-valued process.

For $p \in [2,\infty), q \in [2,\infty]$, $S^q(H) \subset L^p(\Omega;H)$ is dense [3, Lemma 3.1]. Further, the operator $D : S^q(H) \to L^p(\Omega, L^q(0,T;\mathcal{L}_2^0))$ is closable [3, Lemma 3.2]. Let $M^{1,p,q}(H)$ be the closure of $S^q(H)$ with respect to the norm

$$\|Y\|_{M^{1,p,q}(H)} = \big(\|Y\|^p_{L^p(\Omega;H)} + \|DY\|^p_{L^p(\Omega;L^q(0,T;\mathcal{L}_2^0))}\big)^{\frac{1}{p}}.$$

Then $M^{1,p,q}(H)$ are Banach spaces, densely embedded into $L^2(\Omega;H)$, and $M^{1,p,q}(H) \subset L^2(\Omega;H) \subset M^{1,p,q}(H)^*$ is a Gel'fand triple. Further, we denote $M^{1,p}(H) = M^{1,p,p}(H)$ and $M^{1,p}(H)^* = M^{1,p,p}(H)^*$

We shall use frequently Burkholder's inequality ([14, Lemma 7.2] and [3, Lemma 3.5]). For any exponent $p \geq 1$, $p' \geq 1$ denotes its conjugate exponent, i.e., $p^{-1} + {p'}^{-1} = 1$.

Lemma 2.1. *For $p \geq 2$, let $(\Phi(t))_{t\in[0,T]}$ be a predictable and $\mathcal{L}_2^0$-valued stochastic process such that $\|\Phi\|_{L^p(\Omega;L^2(0,T;\mathcal{L}_2^0))} < \infty$. Then there hold*

$$\|\int_0^T \Phi(t)\mathrm{d}W(t)\|_{L^p(\Omega;H)} \leq c\|\Phi\|_{L^p(\Omega;L^2(0,T;\mathcal{L}_2^0))}, \tag{2.3}$$

$$\|\int_0^t \Phi(t)\mathrm{d}W(t)\|_{M^{1,p}(H)^*} \leq c\|\Phi\|_{L^{p'}(\Omega;L^{p'}(0,T;\mathcal{L}_2^0))}. \tag{2.4}$$

Last, we recall one result on the chain rule for Malliavin derivative [3, Lemma 3.3].

Lemma 2.2. *Let U,V be two separable Hilbert spaces and $\gamma \in C^1(U;V)$ satisfy for $u \in U$*

$$\|\gamma(u)\|_V \leq c(1+\|u\|_U^{1+r}) \quad \text{and} \quad \|\gamma'(u)\|_{L(U;V)} \leq c(1+\|u\|_U^r),$$

for some $r \geq 0$. Then for $u \in M^{1,(1+r)p,q}(U)$ with $p > 1$ and $q \geq 2$, $\gamma(u) \in M^{1,p,q}(V)$ and

$$\|\gamma(u)\|_{M^{1,p,q}(V)} \leq c(1+\|u\|^{1+r}_{M^{1,(1+r)p,q}(U)}) \quad \text{and} \quad D[\gamma(u)](\sigma) = \gamma'(u)D[u](\sigma),\ \sigma \in [0,T].$$

3. Numerical scheme

Now we develop a numerical scheme for problem (1.1) based on the Galerkin FEM with conforming piecewise linear FEM in space, Grünwald-Letnikov formula in time, and L^2-projection of the noise $W(t)$. Let $\mathcal{T}_h$ be a shape regular quasi-uniform triangulation of the domain D, and $X_h \subset H_0^1(D)$ be the space of continuous piecewise linear functions on $\mathcal{T}_h$. On the FEM space X_h, we define the $L^2(D)$-projection $P_h : H \to X_h$ by

$$(P_h v, \chi) = (v, \chi), \quad \forall v \in H, \chi \in X_h.$$

Further, let $A_h : X_h \to X_h$ be the discrete analogue of the negative Laplacian A, defined by

$$(A_h v, \chi) = a(v, \chi), \quad \forall v, \chi \in X_h,$$

where $a(v, \chi) = (\nabla v, \nabla \chi)$ is the bilinear form associated with A. Then the semidiscrete Galerkin FEM scheme reads: Given $u_h(0) = P_h u_0$, find $u_h(t) \in X_h$ such that

$$\partial_t^\alpha u_h(t) + A_h u_h(t) = {}_0I_t^\gamma P_h \dot{W}(t), \quad \forall 0 < t \leq T. \tag{3.1}$$

For the time discretization, let $t_n = n\tau$, $n = 0, \ldots, N$, be a uniform partition of the interval $[0, T]$ and $\tau = T/N$ the time step size. We approximate the Riemann-Liouville fractional integral / derivative ${}_0I_t^\gamma v(t_n)$ by Grünwald-Letnikov formula (with $v^k = v(t_k)$)

$${}_0I_t^\gamma v(t_n) \approx \tau^\gamma \sum_{k=0}^n b_{n-k}^{(-\gamma)} v^k, \tag{3.2}$$

where the weights $b_j^{(-\gamma)}$ are generated by power series expansion (with $\delta(\zeta) = 1 - \zeta$):

$$\delta(\zeta)^{-\gamma} = \sum_{j=0}^\infty b_j^{(-\gamma)} \zeta^j.$$

The coefficients $b_j^{(-\gamma)}$ can be computed efficiently via a recursion formula. Since $\partial_t^\alpha u = {}_0I_t^{-\alpha}(u - u(0))$ [27, p. 91], upon letting $f^0 = 0$ and

$$f^k = \tau^{-1} P_h \Delta W^k, \quad \text{with } \Delta W^k = W(t_k) - W(t_{k-1}), \quad k = 1, 2, \ldots, N,$$

the numerical scheme for problem (1.1) reads: find $U^n \in X_h$ such that

$$\tau^{-\alpha} \sum_{k=0}^n b_{n-k}^{(\alpha)} (U^k - U^0) + A_h U^n = \tau^\gamma \sum_{k=0}^n b_{n-k}^{(-\gamma)} f^k, \quad n = 1, 2, \ldots, N, \tag{3.3}$$

with the initial data $U^0 = u_h(0)$. We refer to Section 6.1 below for further implementation details of the term f^k. Note that at the time step n, the scheme employs all preceding numerical solutions U^k and source terms f^k, and thus the computational complexity per time step grows linearly, resulting a quadratic growth of the overall complexity.

4. Nonsmooth data estimates

In this part we prove certain nonsmooth data error estimates.

4.1. **Solution representations**

First we give the solution representations, which are useful for nonsmooth error analysis below. Problem (1.1) admits a unique mild solution of the form

$$u(t) = E(t)u_0 + \int_0^t \bar{E}(t-s)\,\mathrm{d}W(s), \tag{4.1}$$

where the solution operators E and $\bar{E}$ are respectively defined by

$$E(t) = \frac{1}{2\pi\mathrm{i}}\int_\Gamma e^{zt} z^{\alpha-1}(z^\alpha + A)^{-1}\mathrm{d}z \quad \text{and} \quad \bar{E}(t) = \frac{1}{2\pi\mathrm{i}}\int_\Gamma e^{zt} z^{-\gamma}(z^\alpha + A)^{-1}\mathrm{d}z. \tag{4.2}$$

Here the contour Γ is a line in the complex plane $\mathbb{C}$ with $\Re z = a > 0$ for some $a > 0$. One can deform Γ to $\Gamma_{\theta,\delta} := \{z \in \mathbb{C} : z = re^{\pm\mathrm{i}\theta},\ r \geq \delta\} \cup \{z \in \mathbb{C} : z = \delta e^{\mathrm{i}\varphi},\ |\varphi| \leq \theta\}$ for some $\theta > \pi/2$, oriented with an increasing imaginary part. The representation (4.1) can be derived from Laplace transform as follows. Let $g : \mathbb{R}_+ \mapsto H$ be subexponential, i.e., for any $\epsilon > 0$, the function $t \to g(t)e^{-\epsilon t}$ belongs to $L^1(\mathbb{R}_+, H)$. We define Laplace transform $\widehat{g} : \mathbb{C}_+ \mapsto H$ by $\widehat{g}(z) = \int_0^\infty g(t)e^{-zt}\mathrm{d}t$, where $\mathbb{C}_+ = \{z \in \mathbb{C}, \Re z > 0\}$. Then by applying Laplace transform to the following deterministic problem

$$\partial_t^\alpha u(t) + Au(t) = {}_0I_t^\gamma f(t),$$

with $u(0) = u_0$, and using the identities $\widehat{\partial_t^\alpha u}(z) = z^\alpha \widehat{u} - z^{\alpha-1}u_0$ [27, p. 98, Lemma 2.24] and $\widehat{{}_0I_t^\gamma f}(z) = z^{-\gamma}\widehat{f}(z)$ (for $\gamma > 0$) [27, p. 84, Lemma 2.14], we obtain

$$z^\alpha \widehat{u}(z) - z^{\alpha-1}u_0 + A\widehat{u}(z) = z^{-\gamma}\widehat{f}(z),$$

i.e., $\widehat{u}(z) = (z^\alpha + A)^{-1}z^{\alpha-1}u_0 + (z^\alpha + A)^{-1}z^{-\gamma}\widehat{f}(z)$. Then by inverse Laplace transform, we obtain (4.1). It is worth noting that the operator $\bar{E}(t)$ differs from that for standard subdiffusion, due to the presence of the fractional integral ${}_0I_t^\gamma$.

The analysis below relies on smoothing properties of $E(t)$ and $\bar{E}(t)$ in the space $\dot{H}^r(D)$. For any $r \in \mathbb{R}$, let the space $\dot{H}^r(D) = \mathcal{D}(A^{\frac{r}{2}})$ with the norm given by $|v|_r = \|A^{\frac{r}{2}}v\|$. We use extensively the following estimates on $E(t)$ and $\bar{E}(t)$ below.

Lemma 4.1. *For $p, q \in \mathbb{R}$ with $0 \leq p - q \leq 2$, there hold*

$$|E(t)v|_p \leq ct^{-\alpha\frac{p-q}{2}}|v|_q \quad \text{and} \quad |\bar{E}(t)v|_p + t|\bar{E}'(t)v|_p \leq ct^{-\alpha\frac{p-q}{2} + (\alpha+\gamma-1)}|v|_q.$$

Proof. Recall the resolvent estimate [6, Example 3.7.5 and Theorem 3.7.11]

$$\|(z + A)^{-1}\| \leq c_\phi |z|^{-1}, \quad \forall z \in \Sigma_\phi \equiv \{0 \neq z \in \mathbb{C} : |\arg(z)| \leq \phi\},\ \forall\, \phi \in (0, \pi). \tag{4.3}$$

Then simple computation gives $\|A^r(z^\alpha + A)^{-1}\| \leq c|z|^{r\alpha-\alpha}$ for $z \in \Sigma_\theta$ and $r \in [0, 1]$. Thus, taking $\delta = t^{-1}$ in the contour $\Gamma_{\theta,\delta}$ leads to

$$\begin{aligned}|E(t)v|_p &\leq \int_{\Gamma_{\theta,\delta}} e^{\Re z t}|z^{\alpha-1}|\|A^{\frac{p-q}{2}}(z^\alpha + A)\||\mathrm{d}z|\|A^{\frac{q}{2}}v\| \\ &\leq c|v|_q\Big(\int_{t^{-1}}^\infty e^{t\rho\cos\theta}\rho^{\frac{p-q}{2}\alpha-1}\mathrm{d}\rho + ct^{-\frac{p-q}{2}\alpha}\int_{-\theta}^\theta \mathrm{d}\phi\Big) \leq ct^{-\frac{p-q}{2}\alpha}|v|_q.\end{aligned}$$

This shows the desired estimate on $E(t)$, and the other follows similarly. $\square$

Likewise, the semidiscrete solution $u_h(t) \in X_h$ to problem (3.1) is represented by

$$u_h(t) = E_h(t)P_h u_0 + \int_0^t \bar{E}_h(t-s)P_h \mathrm{d}W(s),$$

with the discrete analogues of $E(t)$ and $\bar{E}(t)$, defined by

$$E_h(t) = \frac{1}{2\pi\mathrm{i}}\int_{\Gamma_{\theta,\delta}} e^{zt}z^{\alpha-1}(z^\alpha + A_h)^{-1}\mathrm{d}z \quad \text{and} \quad \bar{E}_h(t) = \frac{1}{2\pi\mathrm{i}}\int_{\Gamma_{\theta,\delta}} e^{zt}z^{-\gamma}(z^\alpha + A_h)^{-1}\mathrm{d}z. \tag{4.4}$$

Next, we give a representation of the fully discrete solution U^n to the scheme (3.3). For a given sequence $\{f^n\}_{n=0}^\infty$, the generating function is given by $\widetilde{f}(\zeta)$, i.e., $\widetilde{f}(\zeta) = \sum_{n=0}^\infty f^n\zeta^n$. Next we introduce operators B_j by

$$\widetilde{B}(\zeta) = \sum_{j=0}^\infty B_j\zeta^j \quad \text{with } \widetilde{B}(\zeta) = 1 + \zeta\big(\tau^{-\alpha}\delta(\zeta)^\alpha + A_h\big)^{-1}\tau^{\gamma-1}\delta(\zeta)^{-\gamma}. \tag{4.5}$$

Proposition 4.1. *The solution U^n to the scheme* (3.3) *is given by*

$$U^n = U_h^n + \tau\sum_{k=1}^n B_{n-(k-1)}f^k, \quad \textit{with } n = 1, 2, \ldots, \tag{4.6}$$

where U_h^n is the fully discrete solution to the homogeneous problem of (3.3).

Proof. We split U^n into $U^n = U_h^n + U_i^n$, where U_h^n and U_i^n are the solutions to the homogeneous and inhomogeneous problems of (3.3), respectively, where U_i^n satisfies

$$\tau^{-\alpha}\sum_{k=0}^n b_{n-k}^{(\alpha)}U_i^k + A_hU_i^n = \tau^\gamma\sum_{k=0}^n b_{n-k}^{(-\gamma)}f^k, \quad n = 1, 2, \ldots,$$

with $U_i^0 = 0$. Multiplying both sides with ξ^n and summing over n from 1 to ∞ yield

$$\tau^{-\alpha}\sum_{n=1}^\infty\Big(\sum_{k=0}^n b_{n-k}^{(\alpha)}U_i^k\Big)\zeta^n + \sum_{k=1}^\infty(A_hU_i^n)\zeta^n = \tau^\gamma\sum_{n=1}^\infty\Big(\sum_{k=0}^n b_{n-k}^{(-\gamma)}f^k\Big)\zeta^n.$$

Since $U_i^0 = 0$ and $f^0 = 0$, by discrete convolution rule and the definitions of $\widetilde{U}_i(\zeta)$ and $\widetilde{f}(\zeta)$,

$$\sum_{n=1}^\infty\Big(\sum_{k=0}^n b_{n-k}^{(\alpha)}U_i^k\Big)\zeta^n = \delta(\zeta)^\alpha\widetilde{U}_i(\zeta) \quad \text{and} \quad \sum_{n=1}^\infty\Big(\sum_{k=0}^n b_{n-k}^{(-\gamma)}f^k\Big)\zeta^n = \delta(\zeta)^{-\gamma}\widetilde{f}(\zeta),$$

from which it directly follows

$$\widetilde{U}_i(\zeta) = \big(\tau^{-\alpha}\delta(\zeta)^\alpha + A_h\big)^{-1}\tau^\gamma\delta(\zeta)^{-\gamma}\widetilde{f}(\zeta).$$

By the defining relation (4.5) of $\widetilde{B}$ and noting $f^0 = 0$, we have

$$\widetilde{U}_i(\zeta) = \tau\frac{\widetilde{B}(\zeta) - 1}{\zeta}\widetilde{f}(\zeta) = \tau\sum_{n=1}^\infty\Big(\sum_{k=0}^{n-1} B_{n-k}f^{k+1}\Big)\zeta^n = \tau\sum_{n=1}^\infty\Big(\sum_{k=1}^n B_{n-(k-1)}f^k\Big)\zeta^n,$$

which implies directly the desired relation. □

The next result holds for the solution U_h^n to the homogeneous problem [24, Theorem 3.5].

Lemma 4.2. *Let $u(t_n)$ and U_h^n be the solution of homogeneous problem and its fully discrete approximation by the scheme* (3.3)*, respectively. Then there holds for $0 \le q \le 2$*

$$\|U_h^n - u(t_n)\| \le c(\tau t_n^{-1+\frac{q}{2}\alpha} + h^2 t_n^{\frac{q-2}{2}\alpha})|u_0|_q.$$

4.2. **Nonsmooth data estimates**

Now we derive some important error estimates for $\bar{E}_h$ and B_j, which are crucial for the error analysis of the scheme (3.3). First, we give spatial discretization errors. On the space X_h, for any $r \in \mathbb{R}$, we define the norm $|||\chi|||_r = \|A_h^{\frac{r}{2}}\chi\|_{L^2(D)}$, which is the discrete analogue of the norm $|\cdot|_r$. Clearly, $|||\cdot|||_0$ coincides the usual $L^2(D)$-norm. Further, on quasiuniform triangulations $\mathcal{T}_h$, for $g \in \dot{H}^r(D)$ with $0 \le r \le 1$, there holds

$$|||P_h g|||_r \le c|g|_r.$$

In fact, the case $r = 0$ follows by the $L^2(D)$-stability of P_h, and the case $r = 1$ by the $H^1(D)$-stability of P_h. The case $r \in (0,1)$ follows by interpolation. Further, the following bound holds

$$\|A_h^{-\frac{s}{2}} P_h A^{\frac{s}{2}}\| \le c, \quad 0 \le s \le 1. \tag{4.7}$$

Indeed, the case $s = 0$ is trivial. Meanwhile, by [38, (3.15)], $\|A_h^{-\frac{1}{2}} P_h v\| \le |v|_{-1}$ for all $v \in \dot{H}^{-1}(D)$. Hence, $\|A_h^{-1/2} P_h A^{1/2}\| \le 1$, and by interpolation, the bound (4.7) follows.

The operator $\bar{E}_h(t)$ satisfies a smoothing property similar to Lemma 4.1. The proof follows from the resolvent estimate for A_h [38, p. 93]:

$$\|(z + A_h)^{-1}\| \le c_\phi |z|^{-1}, \quad \forall z \in \Sigma_\phi, \ \forall \phi \in (0, \pi).$$

Lemma 4.3. *For $p, q \in \mathbb{R}$ with $0 \le p - q \le 2$, there holds*

$$|||\bar{E}_h(t)\chi|||_p \le ct^{-\alpha\frac{p-q}{2}+(\alpha+\gamma-1)}|||\chi|||_q \quad \forall \chi \in S_h.$$

The next lemma gives an error estimate on $\bar{E}_h$.

Lemma 4.4. *Let $0 \le s \le 1$ and $0 \le r \le 2$ with $r + s \le 2$. Then there holds*

$$\|A^{\frac{s}{2}}(\bar{E}(t) - \bar{E}_h(t)P_h)\| \le ct^{\frac{r}{2}\alpha+\gamma-1}h^{2-s-r}.$$

Proof. Fix $g \in L^2(D)$. In the case $s = 0$, by (4.2) and (4.4), there holds

$$\|\big(\bar{E}(t) - \bar{E}_h(t)P_h\big)g\| \le \int_{\Gamma_{\theta,\delta}} e^{\Re zt}\|((z^\alpha + A)^{-1} - (z^\alpha + A_h)^{-1}P_h)g\||z|^{-\gamma}|\mathrm{d}z|.$$

Since $\|((z^\alpha + A)^{-1} - (z^\alpha + A_h)^{-1}P_h)g\| \le ch^2\|g\|$ for all $z \in \Sigma_\theta$ (cf. [19, p. 820] or [8, Lemma 3.4]), we have

$$\|(\bar{E}(t) - \bar{E}_h(t)P_h)g\| \le ct^{\gamma-1}h^2\|g\|.$$

Meanwhile, by Lemmas 4.1 and 4.3 and the triangle inequality,

$$\|(\bar{E}(t) - \bar{E}_h(t)P_h)g\| \le ct^{\alpha+\gamma-1}\|g\|.$$

Similarly, for $s = 1$, there hold

$$\|A^{\frac{1}{2}}(\bar{E}(t) - \bar{E}_h(t)P_h)g\| \le ct^{\gamma-1}h\|g\| \quad \text{and} \quad \|A^{\frac{1}{2}}(\bar{E}(t) - \bar{E}_h(t)P_h)g\| \le ct^{\frac{\alpha}{2}+\gamma-1}\|g\|.$$

Now the desired assertion follows by interpolation. □

Now we analyze the temporal error of the approximation B_nP_h. We begin with an integral representation of B_nP_h.

Lemma 4.5. *For $g \in H$, the function $V^n = B_nP_hg$ is given by (with $\Gamma^\tau_{\theta,\delta} = \{z \in \Gamma_{\theta,\delta} : |\Im z| \le \frac{\pi}{\tau}\}$)*

$$V^n = \frac{1}{2\pi \mathrm{i}} \int_{\Gamma^\tau_{\theta,\delta}} e^{zt_n} (\tau^{-\alpha}\delta(e^{-z\tau})^\alpha + A_h)^{-1} \tau^\gamma \delta(e^{-z\tau})^{-\gamma} P_h g \, \mathrm{d}z.$$

Proof. Direct computation gives (with $V^0 = 0$)

$$\widetilde{V}(\zeta) = \sum_{n=0}^{\infty} V^n \zeta^n = \sum_{n=1}^{\infty} B_n P_h g \zeta^n = \Big(\sum_{n=1}^{\infty} B_n \zeta^n \Big) P_h g.$$

The defining relation (4.5) for $\widetilde{B}(\zeta)$ and B_n leads to

$$\widetilde{V}(\zeta) = (\widetilde{B}(\zeta) - 1)P_h g = (\tau^{-\alpha}\delta(\zeta)^\alpha + A_h)^{-1}\tau^{\gamma-1}\delta(\zeta)^{-\gamma}\zeta P_h g.$$

By Cauchy integral formula, we have, for small $\rho > 0$:

$$V^n = B_n P_h g = \frac{1}{2\pi \mathrm{i}} \int_{|\zeta|=\rho} \zeta^{-n-1} \widetilde{V}(\zeta) \, \mathrm{d}\zeta.$$

The assertion follows by changing variables $\zeta = e^{-z\tau}$ and then deforming $|\zeta| = \rho$ into $\Gamma^\tau_{\theta,\delta}$. □

With Lemma 4.5 and the resolvent estimate on A_h, the following smoothing property and error estimate on B_n follow easily (see, e.g., [24]).

Lemma 4.6. *For any $s \in [0, 1]$, there hold*

$$\|A_h^{\frac{s}{2}} B_n\| \le t_{n+1}^{(1-\frac{s}{2})\alpha+\gamma-1} \quad \text{and} \quad \|A_h^{\frac{s}{2}} (\bar{E}_h(t_n) - B_n)P_h\| \le c\tau \, t_{n+1}^{(1-\frac{s}{2})\alpha+\gamma-2}.$$

For any $s \in [0, 1]$, we define an index $s^* \equiv s^*(\alpha, \gamma)$ by

$$s^* = \begin{cases} \infty, & \text{if } (1 - \frac{s}{2})\alpha + \gamma - 1 \ge 0, \\ \frac{2}{2-2(\alpha+\gamma)+s\alpha}, & \text{otherwise.} \end{cases}$$

For $s^* \ge 2$, s should satisfy the condition $s \le 2 - \frac{1-2\gamma}{\alpha}$. Then the following property holds for B_n.

Lemma 4.7. *For any $s \in [0, 1]$ with $s < 2 - \frac{1-2\gamma}{\alpha}$ and $p \in [2, s^*)$, there hold*

$$\tau \sum_{j=1}^{n} \|B_{n-j}P_h\|^p_{\mathcal{L}_2^0} \le c\|A^{-\frac{s}{2}}\|^p_{\mathcal{L}_2^0}.$$

Proof. By Lemma 4.6, we deduce

$$\begin{aligned} \tau \sum_{j=1}^{n} \|B_{n-j}P_h\|^p_{\mathcal{L}_2^0} &\le \tau \sum_{j=1}^{n} \|A_h^{\frac{s}{2}} B_{n-j}\|^p \|A_h^{-\frac{s}{2}} P_h A^{\frac{s}{2}}\|^p \|A^{-\frac{s}{2}}\|^p_{\mathcal{L}_2^0} \\ &\le c\tau \sum_{j=1}^{n} t_{n-j+1}^{((1-\frac{s}{2})\alpha+\gamma-1)p} \|A^{-\frac{s}{2}}\|^p_{\mathcal{L}_2^0} < \infty. \end{aligned}$$

where the second line follows from (4.7) and the choice of the exponent p. □

Last, we give an important error estimate. It is the main result of this section, and crucial to both strong and weak convergence. Recall that p' is the conjugate exponent of $p \geq 1$.

Theorem 4.1. *For any $0 \leq s \leq 1$, $p \in [1, s^*)$, there holds*

$$\Big(\sum_{j=0}^{n-1}\int_{t_j}^{t_{j+1}} \|A^{\frac{s}{2}}\big(\bar{E}(t_n-t) - B_{n-j}P_h\big)\|^p \,\mathrm{d}s\Big)^{1/p} \leq c(t_n^{\frac{r\alpha}{2}+\gamma-\frac{1}{p'}} h^{2-s-r} + t_n^{\max(\eta-1,0)}\tau^{\mu}),$$

with $\eta = (1-\frac{s}{2})\alpha + \gamma - \frac{1}{p'}$ and the exponents r and μ given respectively by

$$r \in \begin{cases} (\frac{2}{\alpha}(p'^{-1}-\gamma), 2-s], & p'\gamma < 1, \\ (0, 2-s], & p'\gamma = 1, \\ [0, 2-s], & p'\gamma > 1, \end{cases} \quad \text{and} \quad \mu = \begin{cases} \eta, & \eta < 1, \\ 1-\epsilon, & \eta = 1, \\ 1, & \eta > 1. \end{cases}$$

Proof. By the triangle inequality, we split the left hand side (LHS) into

$$\begin{aligned} \text{LHS} \leq & \Big(\sum_{j=0}^{n-1}\int_{t_j}^{t_{j+1}} \|A^{\frac{s}{2}}(\bar{E}(t_n-t) - \bar{E}_h(t_n-t)P_h)\|^p \,\mathrm{d}t\Big)^{1/p} \\ & + \Big(\sum_{j=0}^{n-1}\int_{t_j}^{t_{j+1}} \|A^{\frac{s}{2}}(\bar{E}_h(t_n-t)P_h - \bar{E}_h(t_n-t_j)P_h)\|^p \mathrm{d}t\Big)^{1/p} \\ & + \Big(\sum_{j=0}^{n-1}\int_{t_j}^{t_{j+1}} \|A^{\frac{s}{2}}(\bar{E}(t_n-t_j) - B_{n-j}P_h)\|^p \mathrm{d}t\Big)^{1/p} := \sum_{i=1}^{3} \mathrm{I}_i^{1/p}. \end{aligned}$$

It suffices to bound the three terms I_i. By the choice of the exponent r, $(\frac{r\alpha}{2}+\gamma-1)p > -1$, and thus, by Lemma 4.4,

$$\begin{aligned} \mathrm{I}_1 & \leq ch^{(2-s-r)p}\int_0^{t_n}(t_n-t)^{(\frac{r\alpha}{2}+\gamma-1)p}\mathrm{d}t \\ & \leq ct_n^{(\frac{r\alpha}{2}+\gamma-1)p+1}h^{(2-s-r)p}. \end{aligned}$$

For the second term I_2, simple interpolation between $s = 0, 1$ allows replacing A with A_h, and thus

$$\begin{aligned} \mathrm{I}_2 \leq & \sum_{j=0}^{n-2}\int_{t_j}^{t_{j+1}} \|A_h^{\frac{s}{2}}(\bar{E}_h(t_n-t) - \bar{E}_h(t_n-t_j))P_h\|^p \mathrm{d}t \\ & + \int_{t_{n-1}}^{t_n} \|A_h^{\frac{s}{2}}(\bar{E}_h(t_n-t) - \bar{E}_h(\tau))P_h\|^p \mathrm{d}t := \mathrm{I}_{2,1} + \mathrm{I}_{2,2}. \end{aligned}$$

For the summation $\mathrm{I}_{2,1}$, by Hölder inequality and the smoothing property of $\bar{E}_h'(s)$,

$$\begin{aligned} \mathrm{I}_{2,1} = & \sum_{j=0}^{n-2}\int_{t_j}^{t_{j+1}} \|\int_{t_j}^{s} A_h^{\frac{s}{2}}\bar{E}_h'(t_n-t)P_h\mathrm{d}t\|^p\mathrm{d}s \\ \leq & \sum_{j=0}^{n-2}\int_{t_j}^{t_{j+1}} \tau^{\frac{p}{p'}}\int_{t_j}^{s} \|A_h^{\frac{s}{2}}\bar{E}_h'(t_n-t)P_h\|^p\mathrm{d}t\mathrm{d}s \end{aligned}$$

$$\leq c\tau^p \int_\tau^{t_n} \|A_h^{\frac{s}{2}} \bar{E}_h'(t)\|^p \mathrm{d}t \leq c\tau^p \int_\tau^{t_n} t^{((1-\frac{s}{2})\alpha+\gamma-2)p} \mathrm{d}t.$$

By the definition of η, $p((1-\frac{s}{2})\alpha+\gamma-2) = p(\eta-1)-1$, and then direct computation leads to

$$\mathrm{I}_{2,1} \leq c \begin{cases} \tau^{p\eta}, & \eta < 1, \\ \tau^p \ell_n, & \eta = 1, \\ \tau^p t_n^{p(\eta-1)}, & \eta > 1, \end{cases}$$

with $\ell_n = \ln(1+t_n/\tau)$. For the term $\mathrm{I}_{2,2}$, by the triangle inequality and Lemmas 4.3 and 4.6, we deduce

$$\begin{aligned} \mathrm{I}_{2,2} &\leq c\int_0^\tau \|A_h^{\frac{s}{2}} \bar{E}_h(t)\|^p \mathrm{d}t + c\int_0^\tau \|A_h^{\frac{s}{2}} \bar{E}_h(\tau)\|^p \mathrm{d}t \\ &\leq c\int_0^\tau t^{((1-\frac{s}{2})\alpha+\gamma-1)p} \mathrm{d}t + c\tau^{((1-\frac{s}{2})\alpha+\gamma-1)p+1} \leq c\tau^{p\eta}, \end{aligned}$$

where the last step is due to the choice of the exponent $p \in [1, s^*)$. For the last term I_3, by Lemma 4.6,

$$\begin{aligned} \mathrm{I}_3 &= \sum_{j=0}^{n-1} \int_{t_j}^{t_{j+1}} \|A^{\frac{s}{2}}(\bar{E}_h(t_n-t_j) - B_{n-j})\|^p \, \mathrm{d}t \\ &\leq c\tau^{p+1} \sum_{j=0}^{n-1} (t_{n+1}-t_j)^{((1-\frac{s}{2})\alpha+\gamma-2)p} \leq c \begin{cases} \tau^{p\eta}, & \eta < 1, \\ \tau^p \ell_n, & \eta = 1, \\ \tau^p t_n^{p(\eta-1)}, & \eta > 1. \end{cases} \end{aligned}$$

Combining the preceding estimates on I_is completes the proof of the theorem. □

Remark 4.1. *Note that for $p \in [1, s^*)$, $\eta > 0$ and $\frac{2}{\alpha}(p'^{-1}-\gamma) < 2-s$, and thus the condition on r makes sense. The fractional orders α, γ, the noise regularity index s, and the integrability index p all enter into the final error estimate, and their properly balancing gives the best possible rate.*

5. Strong and weak convergence

This part gives the strong and weak error estimates of the numerical approximation by the scheme (3.3).

5.1. Strong convergence

Now we can state a strong convergence result in $L^p(\Omega; H)$ with $p \geq 2$.

Theorem 5.1. *Let $u(t_n)$ and U^n be the solutions of problems (1.1) and (3.3), respectively. If $\|A^{-\frac{s}{2}}\|_{\mathcal{L}_0^2} < \infty$ for some $s \in [0,1]$ with $s < 2 - \frac{1-2\gamma}{\alpha}$, then for any $p \in [2, s^*)$ and $u_0 \in L^p(\Omega; \dot{H}^q(\Omega))$, $0 \leq q \leq 2$, there holds*

$$\|u(t_n) - U^n\|_{L^p(\Omega;H)} \leq c(\tau t_n^{-1+\frac{q}{2}\alpha} + h^2 t_n^{\frac{q-2}{2}\alpha})\|u_0\|_{L^p(\Omega;\dot{H}^q(D))} + c(t_n^{\frac{r\alpha}{2}+\gamma-\frac{1}{2}} h^{2-s-r} + t_n^{\max(\eta-1,0)}\tau^\mu).$$

with $\eta = (1-\frac{s}{2})\alpha + \gamma - \frac{1}{2}$ and the exponents r and μ given respectively by

$$r \in \begin{cases} (\frac{2}{\alpha}(\frac{1}{2}-\gamma), 2-s], & \gamma < \frac{1}{2}, \\ (0, 2-s], & \gamma = \frac{1}{2}, \\ [0, 2-s], & \gamma > \frac{1}{2}, \end{cases} \quad \text{and} \quad \mu = \begin{cases} \eta, & \eta < 1, \\ 1-\epsilon, & \eta = 1, \\ 1, & \eta > 1. \end{cases}$$

Proof. By the triangle inequality, we have

$$\begin{aligned}\|u(t_n)-U^n\|_{L^p(\Omega;H)} &\le \|(\bar E(t_n)-B_nP_h)u_0\|_{L^p(\Omega;H)}\\ &+\|\sum_{j=0}^{n-1}\int_{t_j}^{t_{j+1}}(\bar E(t_n-t)-B_{n-j}P_h)\mathrm{d}W(t)\|_{L^p(\Omega;H)} := \mathrm{I}+\mathrm{II}.\end{aligned}$$

By Lemma 4.2, it suffices to bound II. By Burkholder's inequality (2.3), since $\|A^{-\frac{s}{2}}\|_{\mathcal{L}_0^2}<\infty$, there holds

$$\begin{aligned}\mathrm{II}^2 &\le c\sum_{j=0}^{n-1}\int_{t_j}^{t_{j+1}}\|\bar E(t_n-t)-B_{n-j}P_h\|^2_{\mathcal{L}_2^0}\mathrm{d}t\\ &\le c\|A^{-\frac{s}{2}}\|^2_{\mathcal{L}_2^0}\sum_{j=0}^{n-1}\int_{t_j}^{t_{j+1}}\|A^{\frac{s}{2}}(\bar E(t_n-t)-B_{n-j}P_h)\|^2\mathrm{d}t.\end{aligned}$$

Then the desired assertion follows from Theorem 4.1 with $p=2$. □

Remark 5.1. *The condition $s<2-\frac{1-2\gamma}{\alpha}$ requests that the noise $W(t)$ should not be too rough, and the condition always holds for trace class noise, since $\alpha+\gamma>1/2$, cf. (1.2). This restriction stems from the limited smoothing property of the solution operator $\bar E(t)$, cf. Lemma A.1. For $u_0=0$ and trace class noise, i.e., $s=0$, the following statements hold:*

(i) *The spatial convergence rate is $O(h^{2-\frac{1-2\gamma}{\alpha}-\epsilon})$ for $\gamma<1/2$, and $O(h^2)$ for $\gamma>1/2$. The former is due to the limited smoothing property of $\bar E(t)$, and it may be enhanced to $O(h^2)$ for smoother noise.*
(ii) *The temporal convergence rate is $O(\tau^{\min(1,\alpha+\gamma-\frac12-\epsilon)})$. When $\gamma=1-\alpha$, it is $O(\tau^{\frac12-\epsilon})$, which coincides with that for the stochastic heat equation [39], but the spatial convergence rate is $O(h^2)$ only if $\alpha<1/2$ or the noise has extra regularity.*

These convergence rates agree with the regularity results in Theorems A.1 and A.2 in Appendix A.

5.2. **Weak convergence**

For the weak convergence, first we give a Malliavin regularity of the solution to problem (1.1).

Proposition 5.1. *If $\|A^{-\frac{s}{2}}\|_{\mathcal{L}_0^2}<\infty$ for some $s\in[0,1]$ with $s\le 2-\frac{1-2\gamma}{\alpha}$, then for any $p\ge 2$ and $q\in[2,s^*)$, and for any $u_0\in L^p(\Omega;\dot H^q(\Omega))$, $0\le q\le 2$, up to modification, there exists a unique stochastic process $u:[0,T]\times\Omega\to H$ satisfying (4.1) such that $u\in C([0,T];M^{1,p,q}(H))$.*

Proof. The proof is similar to [2, Proposition 4.4], and thus we only give a sketch. First, we show $u\in L^2(0,T;M^{1,2}(H))$. This can be done by first proving $\|U^n\|_{M^{1,2}(H)}+\|U^n\|_{M^{1,p,2}}<\infty$, by straightforward calculation of the term $D[U^n](\sigma)$ (see [2, Proposition 4.3]), and then proving the error estimate of $\|u(t_n)-U^n\|_{L^2(\Omega;H)}$, by the argument of [2, Theorem 4.2]. Then a limiting procedure gives $u\in L^2(0,T;M^{1,2}(H))$.

Since $u\in L^2(0,T;M^{1,2}(H))$, we may apply [18, Proposition 3.5 (ii)] or [3, (3.8)] to obtain the Malliavin derivative of the solution u: for any $\sigma\in[0,T]$,

$$\begin{aligned}D[u(t)](\sigma) &= \begin{cases} D[E(t)u_0](\sigma)+D[\int_0^t\bar E(t-s)\,\mathrm{d}W(s)](\sigma), & \sigma\le t\le T,\\ 0, & 0<t<\sigma,\end{cases}\\ &= \begin{cases}\bar E(t-\sigma), & \sigma\le t\le T,\\ 0, & 0<t<\sigma.\end{cases}\end{aligned}$$

Then the smoothing property of $\bar E(t)$ in Lemma 4.1 implies

$$\|D[u(t)]\|^q_{L^p(\Omega;L^q(0,T;\mathcal{L}_2^0))} = \|D[u(t)]\|^q_{L^p(\Omega;L^q(0,t;\mathcal{L}_2^0))}$$

$$
\begin{aligned}
=&\|\bar{E}(t-\cdot)\|^q_{L^q(0,t;\mathcal{L}_2^0)} = \int_0^t \|\bar{E}(t-s)\|^q_{\mathcal{L}_2^0}\mathrm{d}s\\
\le& c\|A^{-\frac{s}{2}}\|^q_{\mathcal{L}_2^0}\int_0^t s^{((1-\frac{s}{2})\alpha+\gamma-1)q}\mathrm{d}s<\infty,
\end{aligned}
$$

where the last inequality is due to the choice of the exponent q. □

The next result gives a similar bound on the discrete solution U^n.

Proposition 5.2. *If $\|A^{-\frac{s}{2}}\|_{\mathcal{L}_0^2}<\infty$ for some $s\in[0,1]$ with $s\le 2-\frac{1-2\gamma}{\alpha}$, then for any $p\ge 2$ and $q\in[2,s^*)$, and $u_0\in L^p(\Omega;\dot H^q(\Omega))$, $0\le q\le 2$, the solution U^n to (3.3) satisfies $\|U^n\|_{M^{1,p,q}(H)}\le c$.*

Proof. By the representation (4.6), we have

$$
\|U^n\|_{L^p(\Omega;H)}\le \|U_h^n\|_{L^p(\Omega;H)}+\Big\|\int_0^T\sum_{j=0}^{n-1}\chi_{[t_j,t_{j+1})}(s)B_{n-j}P_h\mathrm{d}W(s)\Big\|_{L^p(\Omega;H)}:=\mathrm{I}_1+\mathrm{I}_2.
$$

In view of Lemma 4.2, it suffices to bound the second term I_2. By Burkholder's inequality (2.3) and Lemma 4.7 with $p=2$, we get

$$
\mathrm{I}_2\le c\|\sum_{j=0}^{n-1}\chi_{[t_j,t_{j+1})}(s)B_{n-j}P_h\|_{L^2(0,T;\mathcal{L}_2^0)} = c\big(\tau\sum_{j=0}^{n-1}\|B_{n-j}P_h\|^2_{\mathcal{L}_2^0}\big)^{\frac12}<\infty.
$$

This directly implies $\|U^n\|_{L^p(\Omega;H)}\le c$. Next we bound the Malliavin derivative $D[U^n](\sigma)$ of U^n, $\sigma\in[0,T]$. By applying the Malliavin derivative to the representation (4.6) termwise and noting the identity $D[\int_{t_j}^{t_{j+1}}B_{n-j}P_h\mathrm{d}W(s)](\sigma)=\chi_{[t_j,t_{j+1})}(\sigma)B_{n-j}P_h$ [3, Proposition 3.16], we obtain

$$
D[U^n](\sigma)=\sum_{j=0}^{n-1}\chi_{[t_j,t_{j+1})}(\sigma)B_{n-j}P_h,\quad \sigma\in[0,T].
$$

Hence, by Lemma 4.7, there holds

$$
\begin{aligned}
&\|D[U^n]\|^q_{L^p(\Omega;L^q(0,T;\mathcal{L}_2^0))} = \|\sum_{j=0}^{n-1}\chi_{[t_j,t_{j+1})}(s)B_{n-j}P_h\|^q_{L^p(\Omega;L^q(0,T;\mathcal{L}_2^0))}\\
=&\|\sum_{j=0}^{n-1}\chi_{[t_j,t_{j+1})}(s)B_{n-j}P_h\|^q_{L^q(0,T;\mathcal{L}_2^0)}=\tau\sum_{j=0}^{n-1}\|B_{n-j}P_h\|^q_{\mathcal{L}_2^0}<\infty.
\end{aligned}
$$

This completes the proof of the proposition. □

Last, we can give the weak convergence of the approximation U^n.

Theorem 5.2. *Let $u(t_n)$ and U^n be the solutions of (1.1) and (3.3), respectively, and $\Phi\in\mathcal{G}_p^{2,2}(H;\mathbb{R})$. If $\|A^{-\frac{s}{2}}\|_{\mathcal{L}_0^2}<\infty$ for some $s\in[0,1]$ with $s<2-\frac{1-2\gamma}{\alpha}$, then for any $p\in[2,s^*)$ and $u_0\in L^p(\Omega;\dot H^q(D))$, $0\le q\le 2$, there holds*

$$
|\mathbb{E}[\Phi(u(t_n))-\Phi(U^n)]|\le c(\tau t_n^{-1+\frac{q}{2}\alpha}+h^2t_n^{\frac{q-2}{2}\alpha})\|u_0\|_{L^p(\Omega;\dot H^q(D))}+c(t_n^{\frac{r\alpha}{2}+\gamma-\frac1p}h^{2-s-r}+t_n^{\max(\eta-1,0)}\tau^{\mu}),
$$

with $\eta=(1-\frac{s}{2})\alpha+\gamma-\frac{1}{p}$ and the exponents r and μ given respectively by

$$r\in\begin{cases}(\frac{2}{\alpha}(\frac{1}{p}-\gamma),2-s], & \gamma p<1,\\ (0,2-s], & \gamma p=1,\\ [0,2-s], & \gamma p>1,\end{cases}\quad \text{and}\quad \mu=\begin{cases}\eta, & \eta<1,\\ 1-\epsilon, & \eta=1,\\ 1, & \eta>1.\end{cases}$$

Proof. In view of the Gel'fand triple $M^{1,p}(H)\subset L^2(\Omega;H)\subset M^{1,p}(H)^*$ and the definitions of the norms $\|\cdot\|_{M^{1,p}(H)}$ and $\|\cdot\|_{M^{1,p}(H)^*}$, there holds

$$\begin{aligned}|\mathbb{E}[\Phi(u(t_n))-\Phi(U^n)]| &= |\mathbb{E}[(\int_0^1\Phi'(\rho u(t_n)+(1-\rho)U^n)\,\mathrm{d}\rho,u(t_n)-U^n)]|\\ &\leq \|\int_0^1\Phi'(\rho u(t_n)+(1-\rho)U^n)\,\mathrm{d}\rho\|_{M^{1,p}(H)}\|u(t_n)-U^n\|_{M^{1,p}(H)^*}.\end{aligned}$$

Now we claim any $p\in[2,s^*)$, $\|\int_0^1\Phi'(\rho u(t_n)+(1-\rho)U^n)\mathrm{d}\rho\|_{M^{1,p}(H)}<\infty$. Actually, by Lemma 2.2 with $\gamma=\Phi'$ and $r=1$ and $q=p$, $p\in[2,s^*)$, we get

$$\begin{aligned}&\|\Phi'(\rho u(t_n)+(1-\rho)U^n)\|_{M^{1,p}(H)}\\ \leq& c(1+\|\rho u(t_n)+(1-\rho)U^n\|_{M^{1,p}(H)})\\ \leq& c(1+\|u(t_n)\|_{M^{1,2p,p}(H)}+\|U^n\|_{M^{1,2p,p}(H)}).\end{aligned}$$

Thus the claim follows from Propositions 5.1 and 5.2. It remains to bound $\|u(t_n)-U^n\|_{M^{1,p}(H)^*}$. By the triangle inequality,

$$\begin{aligned}\|u(t_n)-U^n\|_{M^{1,p}(H)^*}\leq&\|(E(t_n)-B_nP_h)u_0\|_{M^{1,p}(H)^*}\\ &+\|\int_0^{t_n}\bar{E}(t_n-t)\mathrm{d}W(t)-\sum_{j=0}^{n-1}\int_{t_j}^{t_{j+1}}B_{n-j}P_h\mathrm{d}W(t)\|_{M^{1,p}(H)^*}:=\mathrm{I}+\mathrm{II}.\end{aligned}$$

In view of Lemma 4.2, it suffices to bound the term II. By Burkholder inequality (2.4), we have

$$\begin{aligned}\mathrm{II}^{p'}&\leq c\sum_{j=0}^{n-1}\int_{t_j}^{t_{j+1}}\|\bar{E}(t_n-t)-B_{n-j}P_h\|_{\mathcal{L}_2^0}^{p'}\mathrm{d}t\\ &\leq c\|A^{-\frac{s}{2}}\|_{\mathcal{L}_2^0}^{p'}\sum_{j=0}^{n-1}\int_{t_j}^{t_{j+1}}\|A^{\frac{s}{2}}(\bar{E}(t_n-t)-B_{n-j}P_h)\|^{p'}\,\mathrm{d}s.\end{aligned}$$

Then Theorem 4.1 with $p'\in(1,2)$ completes the proof. □

Remark 5.2. *The condition $s<2-\frac{1-2\gamma}{\alpha}$ ensures that $s^*>2$ so that the choice $p\in[2,s^*)$ is valid. We specialize Theorem 5.2 to $u_0=0$ and trace class noise $W(t)$, i.e., $s=0$, and distinguish two cases for the weak error estimates:* (a) $\alpha+\gamma\geq 1$ *and* (b) $\alpha+\gamma<1$*:*

(a) *The exponent p can be arbitrarily large. Thus, the spatial convergence rate is $O(h^2)$ for any $\gamma\geq 1-\alpha$, and the temporal one $O(\tau^{\min(1,\alpha+\gamma-\epsilon)})$. When $\gamma=1-\alpha$, the temporal rate is $O(\tau^{1-\epsilon})$, which coincides with that for the stochastic heat equation, but the spatial rate is $O(h^2)$ only if $\alpha<1/2$ or $W(t)$ has extra regularity.*

(b) *The largest possible exponent p is $p=\frac{1}{1-\alpha-\gamma}-\epsilon>2$. Hence, the spatial rate is $O(h^2)$ for $\gamma>\frac{1-\alpha}{2}$, and $O(h^{4-\frac{2(1-2\gamma)}{\alpha}-\epsilon})$ for $\gamma\leq\frac{1-\alpha}{2}$ (note that $4-\frac{2}{\alpha}(1-2\gamma)\in(0,2]$ under the designated conditions* (1.2) *and $\gamma\leq\frac{1-\alpha}{2}$). The temporal rate is always $O(\tau^{\alpha+\gamma-\epsilon})$.*

Remark 5.3. *Note that our analysis relies only on Laplace transform and resolvent estimate. Hence, it applies also to slightly more general positive kernels, for which however we are not aware of any physical modeling with fractionally integrated Gaussian noise.*

Remark 5.4. *Since both analytical and numerical solutions can be represented explicitly via a variation of constants formula (cf. (4.1) and (4.6)), alternatively one may employ a (more classical) approach based on Kolmogorov's equation to derive weak convergence rates. In [29, Theorem 4.3], an error representation formula comparing functionals of prescribed time evaluations of such processes (weak convergence) is given for parabolic problems with a positive memory, and upon suitable modifications, it may be combined with the nonsmooth deterministic error estimates in Section 4 to obtain weak convergence rates. However, this approach works only in the linear case and is nontrivial to extend to semilinear stochastic time-fractional PDEs, to which, in contrast, the duality approach can be extended naturally (see, e.g., [2] for stochastic Volterra equations).*

6. Numerical experiments and discussions

Now we present numerical results for the model (1.1) with $0 < \alpha < 1$ and $0 \le \gamma \le 1$ to support the analysis.

6.1. Implementation details

First, we describe the implementation of the noise term $W(t)$, following [39]. We consider only the case the covariance operator Q shares the eigenfunctions with the operator A. Recall the Fourier expansion of the Wiener process $W(t)$ in (2.1):

$$W(t) = \sum_{\ell=1}^{\infty} \gamma_\ell^{\frac{1}{2}} e_\ell \beta_\ell(t),$$

where β_ℓ, $\ell = 1, 2, \ldots$, are i.i.d. Brownian motions, and γ_ℓ and e_ℓ are the eigenvalues (ordered nondecreasingly, with multiplicity counted) and eigenfunctions of Q. Thus the $L^2(D)$-projection $P_h W(t) \in X_h$ is given by (with L term truncation)

$$(P_h W(t), \chi) = \sum_{\ell=1}^{\infty} \gamma_\ell^{\frac{1}{2}} \beta_\ell(t)(e_\ell, \chi) \approx \sum_{\ell=1}^{L} \gamma_\ell^{\frac{1}{2}} \beta_\ell(t)(e_\ell, \chi), \quad \forall \chi \in X_h.$$

Since $\beta_\ell(t)$s are i.i.d. Brownian motions, the increments $\Delta\beta_\ell^k$ are given by

$$\Delta\beta_\ell^k = \beta_\ell(t_k) - \beta_\ell(t_{k-1}) \sim \sqrt{\tau}\mathcal{N}(0,1), \quad k = 1, 2, \ldots, N,$$

where $\mathcal{N}(0,1)$ denotes the standard Gaussian distribution. Further, the fractionally integrated noise $P_h\dot{W}(t_k)$ is approximated by backward difference

$$P_h\dot{W}(t_k) \approx \frac{P_h W(t_k) - P_h W(t_{k-1})}{\tau},$$

and with $P_h\dot{W}(t_0) = 0$. Using Grünwald-Letnikov formula (3.2), the term ${}_0I_t^\gamma P_h\dot{W}(t_n)$ is approximated by

$${}_0I_t^\gamma P_h\dot{W}(t_n) \approx \tau^\gamma \sum_{k=1}^{n} \beta_{n-k}^{(-\gamma)} \Big[\sum_{\ell=1}^{L} \gamma_\ell^{\frac{1}{2}} P_h e_\ell \frac{\Delta\beta_\ell^k}{\tau} \Big].$$

It is well known that for a quasi-uniform triangulation $\mathcal{T}_h$, it is sufficient to take $L \ge N_h$ in the truncation [39], with N_h being the FEM degree of freedom, in order to preserve the desired convergence. The truncation number $L = N_h$ is employed in our numerical experiments.

Below we present numerical results separately for one- and two-dimensional examples, and with $u_0 = 0$. In the 1D case, the domain D is the unit interval $D = (0, 1)$. The eigenfunctions $e_\ell(x)$ are given by $\sqrt{2}\sin(\ell\pi x)$, $\ell = 1, 2, \ldots$, and let $\gamma_\ell = \ell^{-m}, m \geq 0$ (The borderline for trace class noise is $m = 1$, and $m = 0$ corresponds roughly to $s = -1$). In the 2D case, the domain D is the unit square $(0, 1)^2$. The eigenfunctions $e_{i,j}(x)$ are given by $2\sin(i\pi x)\sin(j\pi y)$, $i, j = 1, 2, \ldots$, let $\gamma_{i,j} = (i^2 + j^2)^{-\frac{m}{2}}, m \geq 0$ (The borderline for trace class noise is $m = 1$). In the 1D case, the domain $D = (0, 1)$ is divided into M subintervals of length $h = 1/M$, and similarly, for the 2D, each side $(0, 1)$ is divided into M subintervals of length $h = 1/M$, and the diagonals are connected to obtain a uniform triangulation. The time step size τ is fixed at $\tau = t/N$, where t is the time of interest. To check the convergence rate, we choose the $L^2(\Omega; H)$ norm for strong convergence, and $\Phi(u(t)) = \int_D u(t)^2 \mathrm{d}x$ for weak convergence. All the expected values are computed with 100 trajectories.

6.2. Numerical results for temporal convergence in 1D

In this set of experiments, we fix the final time t at $t = 0.01$ and $M = 100$. The reference solution is computed with a much finer temporal mesh with $N = 3200$. The numerical results for various fractional orders α and γ and trace class noise (with $m = 2$) are given in Table 2. In the table, the numbers in the bracket in the last column denote the theoretical rates predicted by Theorems 5.1 and 5.2 (and Remarks 5.1 and 5.2), and for each α value, the first and second rows give the strong and weak errors, respectively. When $s = 0$, the theoretical rate is nearly $O(\tau^{\min(\alpha+\gamma-\frac{1}{2},1)})$ and $O(\tau^{\min(\alpha+\gamma,1)})$ (up to possibly a logarithmic factor) in the strong and weak sense, respectively. Overall, the empirical rates agree well with the theoretical ones. The convergence rate improves steadily as the fractional orders α and γ increase, due to the improved temporal solution regularity. Further, the weak rate is generally not twice the strong one, unlike the case for the stochastic heat equation.

By Theorems 5.1 and 5.2, the regularity of $W(t)$ also affects the temporal convergence via the term $\|A^{-\frac{s}{2}}\|_{\mathcal{L}_0^2}$: it is slower for white noise than trace class noise, cf. Table 3. By the asymptotics $O(\ell^2)$ of the 1D negative Laplacian A, $m = 0$ corresponds to roughly $s = 1$, and thus Theorems 5.1 and 5.2 yield the theoretical rates $O(\tau^{\min(\frac{\alpha}{2}+\gamma-\frac{1}{2},1)})$ and $O(\tau^{\min(\alpha+2\gamma-1,1)})$ in the strong and weak convergence, respectively; see Table 3. The empirical rates are slightly higher than the theoretical one. Further, noise regularity (indicated by m) beyond trace class affects very little the temporal convergence; see the results for $m = 2, 3$ in Table 3.

6.3. Numerical results for spatial convergence in 1D

Next we examine the spatial convergence. Here, we fix the number M of time steps at $M = 200$ and the final time t at $t = 1$, and compute the reference solution at $N = 480$. The numerical results are given in Table 4 for trace class noise (with $m = 2$) with various α and γ values. A convergence rate $O(h^2)$ is consistently observed for all combinations, concurring Theorems 5.1 and 5.2.

The influence of the noise regularity (indicated by m) on the convergence rates is shown in Table 5. It is observed that for $m = 2$, the weak and strong rates saturate at $O(h^2)$, due to the use of linear finite elements, despite the improved noise regularity. However, it deteriorates when the noise regularity is lowered to the borderline of trace class (i.e., $m = 1$) or white noise (i.e., $m = 0$): for $m = 1$, the strong and weak rates are predicted to be $O(h^{2-\frac{1-2\gamma}{\alpha}})$ and $O(h^2)$, respectively; and for $m = 0$, they are $O(h^{1-\frac{1-2\gamma}{\alpha}})$ and $O(h^{1-\min(1-2\gamma-\frac{\alpha}{2},0)})$, respectively. The empirical rates are much higher than the theoretical ones when $m = 0$, indicating an interesting superconvergence phenomenon, whose precise mechanism remains to be ascertained.

6.4. Numerical results in 2D

Now we present numerical results for the two-dimensional case. Since the computation in the 2D is far more expensive (one convergence rate in the tables below can take about one hour, in `MATLAB` 2018a on a laptop with 2.5 GHz CPU and 8.00 GB RAM), we only present two sets of numerical results, one for temporal convergence and the other for spatial convergence; see Tables 6 and 7.

For temporal convergence, we fix the final time t at $t = 0.01$ and $M = 40$. The reference solution is computed with a finer temporal mesh with $N = 1280$. The numerical results for various (α, γ) pairs are given in Table

Table 2. The $L^2(\Omega; H)$-error for trace class noise ($m = 2$) at $t = 0.01$.

γ	$\alpha\backslash N$	40	80	160	320	640	rate
0.3	0.2	6.68e-1	6.38e-1	6.07e-1	5.55e-1	4.98e-1	0.10 (−−)
		3.82e-1	3.70e-1	3.26e-1	2.79e-1	2.23e-1	0.19 (−−)
	0.4	1.19e-1	9.85e-2	8.14e-2	6.86e-2	5.83e-2	0.25 (0.20)
		1.76e-2	1.37e-2	7.58e-3	6.30e-3	3.37e-3	0.60 (0.70)
	0.6	1.05e-2	7.53e-3	5.13e-3	3.80e-3	2.58e-3	0.50 (0.40)
		2.65e-4	2.22e-4	3.71e-5	1.05e-5	1.49e-5	1.03 (0.90)
	0.8	1.22e-3	8.46e-4	5.61e-4	3.67e-4	2.36e-4	0.59 (0.60)
		2.64e-5	1.56e-5	7.95e-6	3.49e-6	3.31e-7	1.40 (1.00)
0.5	0.2	5.86e-2	5.10e-2	4.16e-2	3.34e-2	2.78e-2	0.26 (0.20)
		4.23e-3	3.36e-3	2.65e-3	1.48e-3	7.26e-4	0.63 (0.70)
	0.4	9.96e-3	7.16e-3	4.98e-3	3.81e-3	2.64e-3	0.47 (0.40)
		3.90e-4	2.33e-4	1.13e-4	5.75e-5	1.64e-5	1.14 (0.90)
	0.6	9.97e-4	6.73e-4	4.43e-4	2.97e-4	1.85e-4	0.60 (0.60)
		2.79e-5	1.33e-5	4.59e-6	3.19e-6	1.67e-6	1.01 (1.00)
	0.8	6.00e-4	3.90e-4	2.30e-4	1.37e-4	8.29e-5	0.71 (0.80)
		2.41e-6	1.64e-6	8.49e-7	2.74e-7	2.16e-7	0.87 (1.00)
0.7	0.2	4.95e-3	4.17e-3	3.21e-3	2.33e-3	1.75e-3	0.37 (0.40)
		5.45e-5	4.27e-5	2.53e-5	1.68e-5	2.11e-5	0.34 (0.90)
	0.4	4.39e-4	2.98e-4	2.15e-4	1.51e-4	1.08e-4	0.50 (0.60)
		5.50e-6	2.81e-6	1.32e-6	7.69e-7	4.36e-7	0.91 (1.00)
	0.6	4.39e-4	3.12e-4	2.06e-4	1.31e-4	7.20e-5	0.65 (0.80)
		1.90e-6	1.56e-6	5.88e-7	3.78e-7	2.03e-7	0.80 (1.00)
	0.8	2.44e-4	1.40e-4	6.75e-5	3.67e-5	1.93e-5	0.91 (1.00)
		4.36e-7	1.37e-7	1.16e-7	5.97e-8	1.83e-8	1.14 (1.00)
0.9	0.2	1.38e-4	7.94e-5	4.83e-5	2.69e-5	1.56e-5	0.78 (0.60)
		1.67e-6	8.28e-7	4.38e-7	2.35e-7	1.28e-7	0.92 (1.00)
	0.4	2.90e-4	2.00e-4	1.31e-4	8.44e-5	5.50e-5	0.59 (0.80)
		8.79e-7	2.83e-7	1.50e-7	3.12e-9	6.37e-8	0.94 (1.00)
	0.6	1.85e-4	1.03e-4	5.35e-5	3.07e-5	1.76e-5	0.84 (1.00)
		3.57e-7	1.53e-7	7.42e-8	4.94e-8	2.46e-8	0.96 (1.00)
	0.8	9.94e-5	5.28e-5	2.50e-5	1.28e-5	5.72e-6	1.02 (1.00)
		1.76e-7	1.12e-7	4.05e-8	1.50e-8	8.36e-9	1.09 (1.00)

6, where the choice $m = 1$ corresponds to the borderline case of trace class noise. The theoretical rate is nearly $O(\tau^{\min(\alpha+\gamma-\frac{1}{2},1)})$ and $O(\tau^{\min(\alpha+\gamma,1)})$ (up to possibly a logarithmic factor) in the strong and weak sense, respectively. Overall, the empirical rates agree excellently with the theoretical ones from Theorems 5.1 and 5.2, although the weak convergence rate suffers from a slight loss when the fractional order α is small, whose precise mechanism is to be ascertained. Like before, the convergence rate improves steadily as the fractional orders α and γ increase, due to the improved temporal solution regularity. For spatial convergence, we fix M at $M = 200$ and the final time t at $t = 1$, and compute the reference solution at $N = 201$. The numerical results are given in Table 4 for trace class noise (with $m = 2$) with various α and γ values. Due to the good spatial regularity of the noise, a convergence rate $O(h^2)$ is consistently observed in both strong and weak norms for all pairs (α, γ), agreeing well with the theoretical predictions from Theorems 5.1 and 5.2.

In summary, all the numerical results in one- and two-dimensional domains indicate that the convergence rates in Theorems 5.1 and 5.2 are nearly sharp. However, the rate in either strong or weak norm is limited to $O(h^2)$ and $O(\tau)$, due to the use of low-order FEM and backward Euler convolution quadrature.

TABLE 3. The $L^2(\Omega; H)$-error at $t = 0.01$ with $\gamma = 0.4$ and noise regularity index m.

m	$\alpha\backslash N$	40	80	160	320	640	rate
0	0.3	1.13e-1	1.02e-1	8.91e-2	7.33e-2	6.06e-2	0.22 (0.05)
		8.12e-3	6.36e-3	6.71e-3	4.74e-3	5.50e-3	0.14 (0.10)
	0.5	2.21e-2	1.83e-2	1.48e-2	1.15e-2	8.41e-3	0.34 (0.15)
		1.09e-3	7.45e-4	4.78e-4	3.18e-4	1.94e-4	0.62 (0.30)
	0.7	4.47e-3	3.30e-3	2.36e-3	1.67e-3	1.11e-3	0.50 (0.25)
		8.24e-5	4.78e-5	3.19e-5	1.90e-5	8.09e-6	0.83 (0.50)
	0.9	1.66e-3	1.11e-3	7.17e-4	4.53e-4	2.75e-4	0.64 (0.30)
		7.49e-6	6.08e-6	2.11e-6	1.25e-6	6.04e-7	0.90 (0.70)
1	0.3	9.76e-2	8.46e-2	7.08e-2	6.23e-2	5.06e-2	0.23 (0.20)
		9.32e-3	7.89e-3	6.72e-3	4.52e-3	2.74e-3	0.44 (0.70)
	0.5	1.37e-2	1.03e-2	7.82e-3	5.79e-3	4.08e-3	0.43 (0.40)
		5.50e-4	3.66e-4	1.33e-4	7.44e-5	4.19e-5	0.92 (0.90)
	0.7	1.84e-3	1.23e-3	8.28e-4	5.44e-4	3.41e-4	0.60 (0.60)
		4.09e-5	1.90e-5	6.90e-6	2.82e-6	2.68e-6	0.98 (1.00)
	0.9	9.01e-4	5.50e-4	3.29e-4	2.03e-4	1.18e-4	0.73 (0.80)
		5.54e-6	2.37e-6	1.55e-6	5.69e-7	1.95e-7	1.20 (1.00)
2	0.3	9.04e-2	7.83e-2	6.85e-2	5.82e-2	4.48e-2	0.25 (0.20)
		7.91e-3	5.70e-3	2.47e-3	6.43e-4	7.08e-4	0.87 (0.70)
	0.5	1.10e-2	8.10e-3	5.76e-3	4.07e-3	2.82e-3	0.49 (0.40)
		2.70e-4	1.93e-4	1.02e-4	5.61e-5	1.82e-5	0.97 (0.90)
	0.7	1.03e-3	7.44e-4	5.11e-4	3.27e-4	2.01e-4	0.59 (0.60)
		1.49e-5	2.73e-6	1.70e-6	1.25e-6	8.04e-7	1.05 (1.00)
	0.9	7.15e-4	4.22e-4	2.60e-4	1.48e-4	8.82e-5	0.75 (0.80)
		4.34e-6	2.50e-6	1.34e-6	3.97e-7	2.89e-7	0.97 (1.00)
3	0.3	9.22e-2	7.53e-2	6.14e-2	5.44e-2	4.05e-2	0.29 (0.20)
		9.75e-3	6.80e-3	4.67e-3	2.39e-3	1.47e-3	0.68 (0.70)
	0.5	1.00e-2	6.86e-3	5.06e-3	3.26e-3	2.07e-3	0.57 (0.40)
		2.92e-4	1.17e-4	4.47e-5	2.53e-5	2.61e-5	0.87 (0.90)
	0.7	9.06e-4	6.16e-4	4.21e-4	2.85e-4	1.81e-4	0.57 (0.60)
		1.81e-5	1.04e-5	3.52e-6	2.28e-6	6.66e-7	1.19 (1.00)
	0.9	6.59e-4	3.99e-4	2.23e-4	1.38e-4	7.36e-5	0.79 (0.80)
		3.13e-6	1.08e-6	6.53e-7	5.00e-7	1.46e-7	1.10 (1.00)

7. Concluding remarks

In this work, we have developed a numerical scheme for approximating the stochastic time-fractional diffusion problem driven by fractionally integrated Gaussian noise. The scheme employs the Galerkin finite element method in space, Grünwald-Letnikov formula in time (for both Caputo fractional derivative and Riemann-Liouville fractional integral) and L^2-projection for the noise. We have presented a thorough error analysis of the scheme in both strong and weak sense, using novel nonsmooth data error estimates. Our extensive numerical results in one spatial dimensional indicate that the convergence rates are nearly sharp.

There are several avenues for further research. First, the computational complexity of the discrete scheme is fairly high, due to nonlocal nature of the fractional-order derivative / integral and the approximation of the expectation. The former may be alleviated by sum of exponential approximations [7], and the latter partly by multilevel Monte Carlo type techniques [20]. It is important to study these computational techniques both numerically and theoretically for stochastic fractional diffusion. Second, this work focuses exclusively on linear

TABLE 4. The $L^2(\Omega; H)$-error with trace class noise ($m = 2$) at $t = 1$.

γ	$\alpha \backslash M$	10	20	40	80	160	rate
0.2	0.3	7.65e-3	2.01e-3	5.15e-4	1.27e-4	2.96e-5	2.00 (−−)
		4.60e-3	1.17e-3	2.92e-4	7.17e-5	1.64e-5	2.03 (−−)
	0.5	6.07e-3	1.63e-3	4.21e-4	1.05e-4	2.44e-5	1.98 (2.00)
		2.48e-3	6.33e-4	1.58e-4	3.88e-5	8.87e-6	2.03 (2.00)
	0.7	4.82e-3	1.32e-3	3.46e-4	8.74e-5	2.03e-5	1.97 (2.00)
		1.30e-3	3.35e-4	8.40e-5	2.06e-5	4.71e-6	2.02 (2.00)
	0.9	4.05e-3	1.12e-3	2.96e-4	7.52e-5	1.76e-5	1.96 (2.00)
		8.79e-4	2.25e-4	5.65e-5	1.38e-5	3.17e-6	2.02 (2.00)
0.6	0.3	2.39e-3	6.25e-4	1.59e-4	3.93e-5	9.09e-6	2.01 (2.00)
		4.68e-4	1.19e-4	2.97e-5	7.29e-6	1.66e-6	2.03 (2.00)
	0.5	2.30e-3	6.02e-4	1.53e-4	3.80e-5	8.78e-6	2.00 (2.00)
		4.22e-4	1.07e-4	2.68e-5	6.58e-6	1.50e-6	2.03 (2.00)
	0.7	2.26e-3	5.92e-4	1.50e-4	3.73e-5	8.64e-6	2.00 (2.00)
		4.02e-4	1.02e-4	2.56e-5	6.27e-6	1.43e-6	2.03 (2.00)
	0.9	2.27e-3	5.94e-4	1.51e-4	3.75e-5	8.67e-6	2.00 (2.00)
		4.09e-4	1.04e-4	2.60e-5	6.37e-6	1.45e-6	2.03 (2.00)

TABLE 5. The $L^2(\Omega; H)$-error at $t = 1$ with $\gamma = 0.4$, and noise regularity index m.

m	$\alpha \backslash M$	10	20	40	80	160	rate
0	0.3	9.58e-3	3.48e-3	1.25e-3	4.36e-4	1.38e-4	1.52 (0.33)
		1.62e-3	4.44e-4	1.15e-4	2.89e-5	6.69e-6	1.98 (0.95)
	0.5	9.20e-3	3.40e-3	1.23e-3	4.33e-4	1.37e-4	1.51 (0.60)
		1.26e-3	3.51e-4	9.22e-5	2.32e-5	5.37e-6	1.96 (1.00)
	0.7	8.75e-3	3.30e-3	1.21e-3	4.29e-4	1.36e-4	1.49 (0.72)
		1.02e-3	2.87e-4	7.60e-5	1.92e-5	4.46e-6	1.95 (1.00)
	0.9	8.27e-3	3.17e-3	1.19e-3	4.24e-4	1.36e-4	1.48 (0.78)
		9.13e-4	2.56e-4	6.81e-5	1.72e-5	4.01e-6	1.95 (1.00)
1	0.3	5.11e-3	1.48e-3	4.16e-4	1.12e-4	2.79e-5	1.87 (1.33)
		1.20e-3	3.11e-4	7.83e-5	1.92e-5	4.40e-6	2.02 (2.00)
	0.5	4.70e-3	1.38e-3	3.95e-4	1.07e-4	2.69e-5	1.86 (1.60)
		8.84e-4	2.29e-4	5.77e-5	1.41e-5	3.24e-6	2.02 (2.00)
	0.7	4.34e-3	1.30e-3	3.75e-4	1.03e-4	2.59e-5	1.84 (1.72)
		6.90e-4	1.79e-4	4.53e-5	1.11e-5	2.55e-6	2.01 (2.00)
	0.9	4.09e-3	1.23e-3	3.59e-4	9.95e-5	2.51e-5	1.83 (1.78)
		6.27e-4	1.63e-4	4.11e-5	1.01e-5	2.31e-6	2.02 (2.00)
2	0.3	3.62e-3	9.49e-4	2.41e-4	6.00e-5	1.38e-5	2.00 (2.00)
		1.06e-3	2.71e-4	6.79e-5	1.66e-5	3.80e-6	2.03 (2.00)
	0.5	3.17e-3	8.39e-4	2.15e-4	5.35e-5	1.23e-5	2.00 (2.00)
		7.59e-4	1.93e-4	4.83e-5	1.18e-5	2.70e-6	2.03 (2.00)
	0.7	2.87e-3	7.63e-4	1.96e-4	4.89e-5	1.13e-5	1.99 (2.00)
		5.83e-4	1.48e-4	3.72e-5	9.12e-6	2.08e-6	2.03 (2.00)
	0.9	2.74e-3	7.28e-4	1.87e-4	4.67e-5	1.08e-5	1.99 (2.00)
		5.35e-4	1.36e-4	3.41e-5	8.36e-6	1.91e-6	2.03 (2.00)

TABLE 6. The $L^2(\Omega; H)$-error for noise with $(m = 1)$ at $t = 0.01$.

γ	$\alpha\backslash N$	20	40	80	160	320	rate
0.4	0.3	2.96e-2	2.63e-2	2.40e-2	1.95e-2	1.43e-2	0.26 (0.20)
		1.25e-3	1.18e-3	9.32e-4	7.07e-4	5.01e-4	0.33 (0.70)
	0.5	6.13e-3	4.72e-3	3.65e-3	2.61e-3	1.72e-3	0.45 (0.40)
		8.38e-5	6.82e-5	4.12e-5	2.57e-5	1.41e-5	0.64 (0.90)
	0.7	9.12e-4	5.96e-4	3.88e-4	2.39e-4	1.38e-4	0.68 (0.60)
		7.96e-6	3.56e-6	1.98e-6	9.20e-7	3.92e-7	1.08 (1.00)
	0.9	3.73e-4	2.36e-4	1.48e-4	8.81e-5	4.78e-5	0.74 (0.80)
		1.09e-6	3.56e-7	2.24e-7	1.05e-7	5.29e-8	1.09 (1.00)
0.8	0.3	1.55e-4	1.05e-4	7.29e-5	4.70e-5	2.70e-5	0.63 (0.60)
		2.43e-7	2.15e-7	9.49e-8	5.21e-8	2.06e-8	0.89 (1.00)
	0.5	1.12e-4	7.92e-5	5.35e-5	3.31e-5	1.97e-5	0.62 (0.80)
		1.96e-7	1.05e-7	4.43e-8	2.01e-8	1.56e-9	1.74 (1.00)
	0.7	7.97e-5	4.78e-5	2.80e-5	1.58e-5	7.99e-6	0.82 (1.00)
		2.28e-9	1.73e-8	5.20e-9	2.11e-9	1.27e-10	1.04 (1.00)
	0.9	4.57e-5	2.45e-5	1.23e-5	6.16e-6	2.88e-6	0.99 (1.00)
		2.85e-8	1.77e-8	9.14e-9	4.12e-9	1.49e-9	1.06 (1.00)

TABLE 7. The $L^2(\Omega; H)$-error with noise $(m = 2)$ at $t = 1$.

γ	$\alpha\backslash M$	10	20	40	80	160	rate
0.4	0.3	5.50e-3	1.57e-3	4.12e-4	1.03e-4	2.44e-5	1.95 (2.00)
		1.13e-3	3.29e-4	8.57e-5	2.14e-5	4.85e-6	1.96 (2.00)
	0.5	4.09e-3	1.17e-3	3.14e-4	8.03e-5	1.84e-5	1.94 (2.00)
		5.38e-4	1.56e-4	4.04e-5	1.06e-5	2.26e-6	1.97 (2.00)
	0.7	2.95e-3	8.74e-4	2.40e-4	6.00e-5	1.38e-5	1.93 (2.00)
		2.43e-4	7.15e-5	1.95e-5	4.70e-6	1.00e-6	1.98 (2.00)
	0.9	2.58e-3	7.59e-4	2.00e-4	5.16e-5	1.21e-5	1.93 (2.00)
		2.13e-4	6.12e-5	1.55e-5	3.99e-6	8.65e-7	1.98 (2.00)
0.8	0.3	1.20e-3	3.43e-4	8.94e-5	2.16e-5	5.37e-6	1.95 (2.00)
		5.05e-5	1.49e-5	3.86e-6	9.04e-7	2.21e-7	1.95 (2.00)
	0.5	1.03e-3	3.04e-4	7.95e-5	2.01e-5	4.55e-6	1.95 (2.00)
		3.75e-5	1.11e-5	2.90e-6	7.40e-7	1.53e-7	1.98 (2.00)
	0.7	1.02e-3	2.87e-4	7.72e-5	1.95e-5	4.53e-6	1.95 (2.00)
		3.71e-5	1.02e-5	2.76e-6	7.07e-7	1.52e-7	1.98 (2.00)
	0.9	8.34e-4	2.54e-4	6.55e-5	1.65e-5	3.91e-6	1.93 (2.00)
		2.32e-5	7.22e-6	1.83e-6	4.67e-7	1.02e-7	1.95 (2.00)

stochastic fractional PDEs. It is of much interest to extend the analysis to more complex situations, e.g., semilinear or quasi-linear problems. Third and last, it is important to develop temporally high-order schemes with optimal strong and weak convergence rates. This is conjectured for a certain regime of $\alpha + \gamma$, for which the solution has good temporal regularity.

Acknowledgements

The authors are grateful to two anonymous referees for their constructive comments. The research of Z. Zhou is partially supported by a start-up grant from the Hong Kong Polytechnic University and a grant from Research Grants Council of the Hong Kong Special Administrative Region (Project No. 25300818).

References

[1] E. J. Allen, S. J. Novosel, and Z. Zhang. Finite element and difference approximation of some linear stochastic partial differential equations. *Stoc. Stoc. Rep.*, 64(1-2):117–142, 1998.
[2] A. Andersson, M. Kovács, and S. Larsson. Weak error analysis for semilinear stochastic Volterra equations with additive noise. *J. Math. Anal. Appl.*, 437(2):1283–1304, 2016.
[3] A. Andersson, R. Kruse, and S. Larsson. Duality in refined Sobolev-Malliavin spaces and weak approximation of SPDE. *Stoch. Partial Differ. Equ. Anal. Comput.*, 4(1):113–149, 2016.
[4] A. Andersson and S. Larsson. Weak convergence for a spatial approximation of the nonlinear stochastic heat equation. *Math. Comp.*, 85(299):1335–1358, 2016.
[5] V. V. Anh, N. N. Leonenko, and M. Ruiz-Medina. Space-time fractional stochastic equations on regular bounded open domains. *Fract. Calc. Appl. Anal.*, 19(5):1161–1199, 2016.
[6] W. Arendt, C. J. Batty, M. Hieber, and F. Neubrander. *Vector-valued Laplace Transforms and Cauchy Problems*. Birkhäuser, Basel, 2nd edition, 2011.
[7] D. Baffet and J. S. Hesthaven. A kernel compression scheme for fractional differential equations. *SIAM J. Numer. Anal.*, 55(2):496–520, 2017.
[8] E. Bazhlekova, B. Jin, R. Lazarov, and Z. Zhou. An analysis of the Rayleigh-Stokes problem for a generalized second-grade fluid. *Numer. Math.*, 131(1):1–31, 2015.
[9] C.-E. Bréhier, M. Hairer, and A. M. Stuart. Weak error estimates for trajectories of SPDEs under spectral Galerkin discretization. *J. Comput. Math.*, 36(2):159–182, 2018.
[10] L. Chen. Nonlinear stochastic time-fractional diffusion equations on $\mathbb{R}$: moments, Hölder regularity and intermittency. *Trans. Amer. Math. Soc.*, 369(12):8497–8535, 2017.
[11] L. Chen, Y. Hu, and D. Nualart. Nonlinear stochastic time-fractional slow and fast diffusion equations on $\mathbb{R}^d$. preprint, arXiv:1509.07763, 2015.
[12] Z.-Q. Chen, K.-H. Kim, and P. Kim. Fractional time stochastic partial differential equations. *Stochastic Process. Appl.*, 125(4):1470–1499, 2015.
[13] E. Cuesta, C. Lubich, and C. Palencia. Convolution quadrature time discretization of fractional diffusion-wave equations. *Math. Comp.*, 75(254):673–696, 2006.
[14] G. Da Prato and J. Zabczyk. *Stochastic Equations in Infinite Dimensions*. Cambridge University Press, Cambridge, 2nd edition, 2014.
[15] A. Debussche and J. Printems. Weak order for the discretization of the stochastic heat equation. *Math. Comp.*, 78(266):845–863, 2009.
[16] Q. Du and T. Zhang. Numerical approximation of some linear stochastic partial differential equations driven by special additive noises. *SIAM J. Numer. Anal.*, 40(4):1421–1445, 2002.
[17] M. Foondun. Remarks on a fractional-time stochastic equation. Preprint, arXiv:1811.05391, 2018.
[18] M. Fuhrman and G. Tessitore. Nonlinear Kolmogorov equations in infinite dimensional spaces: the backward stochastic differential equations approach and applications to optimal control. *Ann. Probab.*, 30(3):1397–1465, 2002.
[19] H. Fujita and T. Suzuki. Evolution problems. In *Handbook of Numerical Analysis, Vol. II*, pages 789–928. North-Holland, Amsterdam, 1991.
[20] M. B. Giles. Multilevel Monte Carlo methods. *Acta Numer.*, 24:259–328, 2015.
[21] M. Gunzburger, B. Li, and J. Wang. Convergence of finite element solution of stochastic partial integral-differential equations driven by white noise. *arXiv preprint arXiv:1711.01998*, 2017.
[22] A. Jentzen and P. E. Kloeden. The numerical approximation of stochastic partial differential equations. *Milan J. Math.*, 77:205–244, 2009.
[23] B. Jin, R. Lazarov, and Z. Zhou. Error estimates for a semidiscrete finite element method for fractional order parabolic equations. *SIAM J. Numer. Anal.*, 51(1):445–466, 2013.
[24] B. Jin, R. Lazarov, and Z. Zhou. Two fully discrete schemes for fractional diffusion and diffusion-wave equations with nonsmooth data. *SIAM J. Sci. Comput.*, 38(1):A146–A170, 2016.
[25] B. Jin, R. Lazarov, and Z. Zhou. Numerical methods for time-fractional evolution equations with nonsmooth data: A concise overview. *Comput. Methods Appl. Mech. Engrg.*, 346:332–358, 2019.
[26] B. Jin, B. Li, and Z. Zhou. Correction of high-order BDF convolution quadrature for fractional evolution equations. *SIAM J. Sci. Comput.*, 39(6):A3129–A3152, 2017.

[27] A. A. Kilbas, H. M. Srivastava, and J. J. Trujillo. *Theory and Applications of Fractional Differential Equations*. Elsevier Science B.V., Amsterdam, 2006.
[28] M. Kovács and J. Printems. Strong order of convergence of a fully discrete approximation of a linear stochastic Volterra type evolution equation. *Math. Comp.*, 83(289):2325–2346, 2014.
[29] M. Kovács and J. Printems. Weak convergence of a fully discrete approximation of a linear stochastic evolution equation with a positive-type memory term. *J. Math. Anal. Appl.*, 413(2):939–952, 2014.
[30] R. Kruse. *Strong and Weak Approximation of Semilinear Stochastic Evolution Equations*, volume 2093 of *Lecture Notes in Mathematics*. Springer, Heidelberg, 2014.
[31] X. Li and X. Yang. Error estimates of finite element methods for stochastic fractional differential equations. *J. Comput. Math.*, 35(3):346–362, 2017.
[32] W. Liu, M. Röckner, and J. L. da Silva. Quasi-linear (stochastic) partial differential equations with time-fractional derivatives. *SIAM J. Math. Anal.*, 50(3):2588–2607, 2018.
[33] S. V. Lototsky and B. L. Rozovsky. Classical and generalized solutions of fractional stochastic differential equations. Preprint, arXiv:1810.12951, 2018.
[34] C. Lubich. Discretized fractional calculus. *SIAM J. Math. Anal.*, 17(3):704–719, 1986.
[35] C. Lubich, I. H. Sloan, and V. Thomée. Nonsmooth data error estimates for approximations of an evolution equation with a positive-type memory term. *Math. Comp.*, 65(213):1–17, 1996.
[36] W. McLean and K. Mustapha. Time-stepping error bounds for fractional diffusion problems with non-smooth initial data. *J. Comput. Phys.*, 293:201–217, 2015.
[37] R. Metzler, J. H. Jeon, A. G. Cherstvy, and E. Barkai. Anomalous diffusion models and their properties: non-stationarity, non-ergodicity, and ageing at the centenary of single particle tracking. *Phys. Chem. Chem. Phys.*, 16(44):24128–24164, 2014.
[38] V. Thomée. *Galerkin Finite Element Methods for Parabolic Problems*. Springer-Verlag, Berlin, second edition, 2006.
[39] Y. Yan. Galerkin finite element methods for stochastic parabolic partial differential equations. *SIAM J. Numer. Anal.*, 43(4):1363–1384, 2005.

Appendix A. Regularity theory

In this appendix, we describe some regularity results for problem (1.1). First, we state a result on $\bar{E}(t)$.

Lemma A.1. *Let condition* (1.2) *hold, and let* (*with* $\epsilon > 0$ *small*)

$$\kappa = \begin{cases} 2, & \text{if } 1/2 < \gamma \leq 1, \\ 2-\epsilon, & \text{if } \gamma = 1/2, \\ 2 - \frac{1-2\gamma}{\alpha}, & \text{if } 0 \leq \gamma < 1/2. \end{cases} \tag{A.1}$$

Then there holds

$$\|\bar{E}(s)\|_{L^2(0,t;\dot{H}^\kappa(D))} \leq ct^{(2-\kappa)\frac{\alpha}{2}+\gamma-\frac{1}{2}}.$$

Proof. By Lemma 4.1, for any $\kappa \in [0,2]$, there holds $\|A^{\frac{\kappa}{2}}\bar{E}(t)\| \leq ct^{(1-\frac{\kappa}{2})\alpha+\gamma-1}$, and consequently

$$\int_0^t \|A^{\frac{\kappa}{2}}\bar{E}(s)\|^2 \mathrm{d}s \leq c\int_0^t s^{(2-\kappa)\alpha+2\gamma-2}\mathrm{d}s.$$

Under Assumption 1.2 and the choice of the exponent κ in (A.1), $(2-\kappa)\alpha + 2\gamma - 2 > -1$ (except for the case $0 \leq \gamma < 1/2$ and $\kappa = 2 - \frac{1-2\gamma}{\alpha})$), thus we obtain the desired result upon integration. In the exceptional case, $(2-\kappa)\alpha + 2\gamma - 2 = -1$, and the assertion follows from direct computation. ☐

Remark A.1. *For* $\gamma \in (1/2, 1]$, $\bar{E}(t)$ *has a maximum order two smoothing in space. However, for* $\gamma \in [0, 1/2)$, κ *is restricted to* $[0, 2-\frac{1-2\gamma}{\alpha}]$. *This again reflects a certain limited smoothing property of* $\bar{E}(t)$, *and also restricts the best possible strong and weak spatial convergence rates.*

Now we can state the spatial regularity of the mild solution (4.1).

Theorem A.1. *Let condition* (1.2) *hold, and* κ *be defined by* (A.1), *and* $r, q \in \mathbb{R}$ *with* $0 \leq r - q \leq 2$. *For* $u_0 \in L^p(\Omega; \dot{H}^q(D))$, *let* $u(t)$ *be the mild solution of problem* (1.1) *defined in* (4.1). *Then there holds*

$$\|u(t)\|_{L^p(\Omega;\dot{H}^r(D))} \leq ct^{-\alpha\frac{r-q}{2}}\|u_0\|_{L^p(\Omega;\dot{H}^q(D))} + ct^{(1-\frac{\kappa}{2})\alpha+\gamma-\frac{1}{2}}\|A^{\frac{r-\kappa}{2}}\|_{\mathcal{L}_2^0}.$$

Proof. By Lemma 4.1, it suffices to bound the stochastic integral. By Burkholder's inequality (2.3),

$$\Big(\mathbb{E}\|\int_0^t A^{\frac{r}{2}}\bar{E}(t-s)\,\mathrm{d}W(s)\|^p\Big)^{2/p} \le c\int_0^t \|A^{\frac{r}{2}}\bar{E}(s)\|_{\mathcal{L}_2^0}^2\,\mathrm{d}s \le \|A^{\frac{r-\kappa}{2}}\|_{\mathcal{L}_2^0}\int_0^t \|A^{\frac{\kappa}{2}}\bar{E}(s)\|\mathrm{d}s,$$

Then Lemma A.1, the representation (4.1) and the triangle inequality complete the proof. □

To study the temporal regularity of the mild solution in (4.1), we need an elementary inequality.

Lemma A.2. *For $0\le t_1<t_2$ and $\theta\in(1/2,3/2)$, then with $c=(3-2\theta)^{-\frac12}(\theta-\frac12)^{-1}$, there holds $\int_{t_1}^{t_2}\big(\int_0^{t_1}(t-s)^{2(\theta-2)}\mathrm{d}s\big)^{\frac12}\mathrm{d}t\le c(t_2-t_1)^{\theta-\frac12}$.*

Proof. Since $\theta\in(1/2,3/2)$, i.e., $3-2\theta>0$, straightforward computation gives for $t>t_1$ $\int_0^{t_1}(t-s)^{2(\theta-2)}\mathrm{d}s\le(3-2\theta)^{-1}(t-t_1)^{2\theta-3}$. Thus simple computation gives $\int_{t_1}^{t_2}\Big(\int_0^{t_1}(t-s)^{2(\theta-2)}\mathrm{d}s\Big)^{\frac12}\mathrm{d}t\le c(t_2-t_1)^{\theta-\frac12}$, completing the proof. □

Now we can state the temporal Hölder regularity of the mild solution in (4.1).

Theorem A.2. *Let condition (1.2) hold, and κ be defined in (A.1). Let $u(t)$ be defined in (4.1). Let $q,r\in[0,2]$ with $0\le r-q\le 2$, and $s\in[0,1]$ with $\max(0,\frac{2(\alpha+\gamma)-3}{\alpha}+\epsilon)\le r+s\le\kappa$. Then for any $0<t_1<t_2<T$ and $p\ge 2$ and $u_0\in L^p(\Omega;\dot H^q(D))$, there holds*

$$\|u(t_2)-u(t_1)\|_{L^p(\Omega;\dot H^r(D))}\le ct_1^{-(1+\alpha\frac{r-q}{2})}(t_2-t_1)\|u_0\|_{L^p(\Omega;\dot H^q(D))}+c(t_2-t_1)^{(1-\frac{r+s}{2})\alpha+\gamma-\frac12}\|A^{-\frac{s}{2}}\|_{\mathcal{L}_2^0}.$$

Proof. By the representation (4.1), we have the splitting

$$\begin{aligned} u(t_2)-u(t_1)=&(E(t_2)u_0-E(t_1)u_0)+\int_0^{t_1}(\bar{E}(t_2-s)-\bar{E}(t_1-s))\mathrm{d}W(s)\\ &+\int_{t_1}^{t_2}\bar{E}(t_2-s)\mathrm{d}W(s):=\mathrm{I}+\mathrm{II}+\mathrm{III}.\end{aligned}$$

Next we bound the three terms separately. The first term can be bounded directly by Lemma 4.1. Next, for $\alpha+\gamma\neq1$ (the case $\alpha+\gamma=1$ is similar), by stochastic Fubini theorem [14, Theorem 4.33],

$$\mathrm{II}=\int_0^{t_1}(\bar{E}(t_2-s)-\bar{E}(t_1-s))\mathrm{d}W(s)=\int_{t_1}^{t_2}\int_0^{t_1}\bar{E}'(t-s)\mathrm{d}W(s)\mathrm{d}t.$$

Thus by Burkholder's inequality (2.3) and Lemma 4.1, we have (with $\theta=\alpha+\gamma-\frac{r+s}{2}\alpha$)

$$\begin{aligned}\|\mathrm{II}\|_{L^p(\Omega;\dot H^r(D))}&\le\int_{t_1}^{t_2}\|\int_0^{t_1}A^{\frac r2}\bar{E}'(t-s)\mathrm{d}W(s)\|_{L^p(\Omega;H)}\mathrm{d}t\\ &\le c\int_{t_1}^{t_2}\Big(\int_0^{t_1}\|A^{\frac{r+s}{2}}\bar{E}'(t-s)\|\|A^{-\frac s2}\|_{\mathcal{L}_2^0}\Big)^2\mathrm{d}s\Big)^{\frac12}\mathrm{d}t\\ &=c\int_{t_1}^{t_2}\Big(\int_0^{t_1}(t-s)^{2(\theta-2)}ds\Big)^{\frac12}\mathrm{d}t\|A^{-\frac s2}\|_{\mathcal{L}_2^0}.\end{aligned}$$

For $1/2<\alpha+\gamma<3/2$, the condition on $r+s$ ensures $\theta\in(1/2,3/2)$. Similarly, for $\alpha+\gamma\ge3/2$ and $r\in(\frac{2(\alpha+\gamma)-3}{\alpha},2]$, we have also $\theta\in(1/2,3/2)$. Thus, Lemma A.2 implies

$$\|\mathrm{II}\|_{L^p(\Omega;\dot H^r(D))}\le c(t_2-t_1)^{\theta-\frac12}\|A^{-\frac s2}\|_{\mathcal{L}_2^0}.$$

Last, by Burkholder's inequality (2.3) and Lemma 4.1 with $p = r + s$ and $q = 0$, we deduce

$$\begin{aligned}\|\mathrm{III}\|^2_{L^p(\Omega;\dot H^r(D))} &\le c\int_{t_1}^{t_2}\|A^{\frac{r+s}{2}}\bar E(t_2-s)A^{-\frac{s}{2}}\|^2_{\mathcal{L}_2^0}\mathrm{d}s\\ &\le c\|A^{-\frac{s}{2}}\|^2_{\mathcal{L}_2^0}\int_{t_1}^{t_2}(t_2-s)^{2\theta-2}\,\mathrm{d}s = c(t_2-t_1)^{2\theta-1}\|A^{-\frac{s}{2}}\|^2_{\mathcal{L}_2^0}.\end{aligned}$$

Combining these estimates together completes the proof of the theorem. □

Remark A.2. *The condition* $\max(0, \frac{2(\alpha+\gamma)-3}{\alpha} + \epsilon) \le r + s \le \kappa$ *is only for* $\alpha + \gamma \ge 3/2$ *and restricts the discussion to Hölder continuity in time. For* $u_0 = 0$ *and trace class noise, i.e.,* $s = 0$*,* $r = 0$*,*

$$\|u(t_2) - u(t_1)\|_{L^2(\Omega;H)} \le c(t_2-t_1)^{\alpha+\gamma-\frac12}\|Q^{\frac12}\|_{HS}.$$

The case of $\alpha = 1$ *and* $\gamma = 0$ *recovers the well known regularity result of the stochastic heat equation.*