Mathematical Modelling and Numerical Analysis Will be set by the publisher
Modélisation Mathématique et Analyse Numérique

AN ANALYSIS OF GALERKIN PROPER ORTHOGONAL DECOMPOSITION FOR SUBDIFFUSION

Bangti Jin[1] and Zhi Zhou[2]

Abstract. In this work, we develop a novel Galerkin-L1-POD scheme for the subdiffusion model with a Caputo fractional derivative of order $\alpha \in (0,1)$ in time, which is often used to describe anomalous diffusion processes in heterogeneous media. The nonlocality of the fractional derivative requires storing all the solutions from time zero. The proposed scheme is based on continuous piecewise linear finite elements, L1 time stepping, and proper orthogonal decomposition (POD). By constructing an effective reduced-order scheme using problem-adapted basis functions, it can significantly reduce the computational complexity and storage requirement. We shall provide a complete error analysis of the scheme under realistic regularity assumptions by means of a novel energy argument. Extensive numerical experiments are presented to verify the convergence analysis and the efficiency of the proposed scheme.

1991 Mathematics Subject Classification. 65M60, 65M15.

January 26, 2016.

1. Introduction

In this work, we consider the following model initial-boundary value problem for $u(x,t)$:

$$\begin{aligned} \partial_t^\alpha u - \Delta u &= f, \quad \text{in } \Omega \qquad T \geq t > 0, \\ u &= 0, \quad \text{on } \partial\Omega \quad T \geq t > 0, \\ u(0) &= v, \quad \text{in } \Omega, \end{aligned} \tag{1.1}$$

where Ω is a bounded convex polygonal domain in $\mathbb{R}^d$ $(d = 1, 2, 3)$ with a boundary $\partial\Omega$ and v is a given function defined on the domain Ω and $T > 0$ is a fixed value. Here $\partial_t^\alpha u$ $(0 < \alpha < 1)$ denotes the left-sided Caputo fractional derivative of order α with respect to t and it is defined by (see, e.g. [15, pp. 91])

$$\partial_t^\alpha u(t) = \frac{1}{\Gamma(1-\alpha)} \int_0^t (t-s)^{-\alpha} \frac{d}{ds} u(s)\, ds, \tag{1.2}$$

where $\Gamma(\cdot)$ is Euler's Gamma function defined by $\Gamma(x) = \int_0^\infty s^{x-1} e^{-s} ds$ for $x > 0$.

Keywords and phrases: fractional diffusion, energy argument, proper orthogonal decomposition, error estimates

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

[2] Department of Applied Physics and Applied Mathematics, Columbia University, New York, NY, 10027, USA (zhizhou0125@gmail.com)

In recent years, the model (1.1) has received much interest in physical modeling, mathematical analysis and numerical simulation. The main engine that has fueled these developments is its extraordinary capability for describing anomalously slow diffusion processes, in which the mean square variance of particle displacements grows sublinearly with time, instead of linear growth for a Gaussian process. At a microscopic level, the particle motion is more adequately described by continuous time random walk, whose macroscopic counterpart is a differential equation with a fractional derivative in time [24]. Nowadays the model has been successfully employed in many applications, e.g., thermal diffusion in fractal domains [26], ion transport in column experiments [6], and non-Fickian transport in geological formation [2], to name just a few.

Numerically, the presence of the fractional derivative $\partial_t^\alpha u$ has two important consequences. First, the nonlocality in time incurs huge storage requirement as well as much increased computational efforts along the evolution of the time. Second, the solution operator has only very limited smoothing property: the problem has at best order two smoothing in space [31], and the first derivative in time is usually unbounded, cf. Theorem A.1 in the appendix. These represent the main technical challenges in the development and analysis of robust numerical schemes for reliably simulating subdiffusion. The challenges are especially severe for "multi-query" applications, e.g., inverse problems and optimal control, where repeated solutions of "analogous" forward problems are required, e.g., due to variation in problem parameters or inputs. To reduce the storage requirement, a number of useful strategies have been proposed, e.g., short-memory principle and panel clustering [4, 21, 23, 28].

In this work, we shall develop an efficient strategy, called the Galerkin-L1-POD scheme, for reliably simulating the subdiffusion model (1.1) by coupling the Galerkin finite element method (FEM) with proper orthogonal decomposition (POD) to reduce the computational complexity of repeatedly simulating subdiffusion, which is important for solving related inverse problems and optimal control. POD is a popular model reduction technique, and it has achieved great success in reducing the complexity of mathematical models governed by differential equations; see [1, 3, 17, 18, 29, 33] for a rather incomplete list. It is especially attractive in optimal control [8, 16, 19, 30] and parameter inversion [10, 25]. To the best of our knowledge, this work represents the first application of the POD for the subdiffusion model (1.1) with a complete error analysis.

Next we describe the proposed scheme. Let $\mathcal{T}_h$ be a shape regular quasi-uniform partition of the domain Ω, and X_h be the associated continuous piecewise linear finite element space. Meanwhile, we discretize the Caputo fractional derivative $\partial_t^\alpha u(t)$ by the L1 approximation $\bar\partial_\tau^\alpha u(t_n)$ (with a time step size τ) [20, 35]

$$\bar\partial_\tau^\alpha u(t_n) = \sum_{j=0}^{n-1} b_j \frac{u(t_{n-j}) - u(t_{n-j-1})}{\tau^\alpha \Gamma(2-\alpha)},$$

where the weights $\{b_j\}$ are defined by (2.4). With the Galerkin FEM in space and L1 approximation in time, we arrive at the following fully discrete scheme: find $U_h^n \in X_h$ for $n = 1, 2, \ldots, N$

$$(\bar\partial_\tau^\alpha U_h^n, \varphi) + (\nabla U_h^n, \nabla\varphi) = (f(t_n), \varphi) \quad \forall \varphi \in X_h,$$

with $U_h^0 \in X_h$ being an approximation to the initial data v, where $(\cdot,\cdot)$ denotes the $L^2(\Omega)$ inner product. The term $\bar\partial_\tau^\alpha U_h^n$ involves all solutions $\{U_h^i\}_{i=0}^{n-1}$ preceding the current time step n, indicating the computational challenge. In this work, we shall adopt the POD methodology to overcome the challenge. Specifically, we take the fully discrete solutions $\{U_h^n\}_{n=0}^N$ and fractional difference quotients $\{\bar\partial_\tau^\alpha U_h^n\}_{n=1}^N$ as snapshots to generate an optimal orthonormal basis $\{\psi_j\}_{j=1}^r$. Since these snapshots are sampled from the solution manifold, the POD basis is automatically adapted to the characteristics of the manifold and is expected to have good approximation property. Then we employ a Galerkin framework using the POD space X_h^m, $m \le r$, spanned by the first m POD basis functions, i.e., find $U_m^n \in X_h^m$, $n = 1, 2 \ldots, N$ such that

$$(\bar\partial_\tau^\alpha U_m^n, \varphi) + (\nabla U_m^n, \nabla\varphi) = (f(t_n), \varphi) \quad \forall \varphi \in X_h^m,$$

with $U_m^0 \in X_h^m$ being an approximation to U_h^0. In the reduced order formulation, the degree of freedom is m, the number of POD basis functions, which is usually much smaller than that of the full Galerkin formulation.

Hence, it yields an enormous reduction in computational complexity and storage requirement. We shall provide a complete a priori convergence analysis of the scheme. Our main theoretical result is given in Theorem 3.6. For example for the POD approximation $\{U_m^n\}_{n=1}^N$ generated using the $H_0^1(\Omega)$-POD basis, the following error estimate holds (with $\ell_h = |\log h|$)

$$\frac{1}{N}\sum_{n=1}^{N} \|u(t_n) - U_m^n\|_{L^2(\Omega)}^2 \le c_T\big(\tau^{2\alpha} + h^4\ell_h^4 + \sum_{j=m+1}^{r} \widetilde{\lambda}_j\big),$$

where $\{\widetilde{\lambda}_j\}_{j=1}^r$ are the descendingly ordered eigenvalues of the correlation matrix $\widetilde{K}$ (see Section 2.3 for details) under suitable verifiable regularity conditions on the source term f and the initial data v.

This error estimate consists of three components: spatial error $O(h^2\ell_h^2)$, temporal error $O(\tau^\alpha)$ and POD error $(\sum_{j=m+1}^r \widetilde{\lambda}_j)^{1/2}$. While nearly optimal error estimates due to the spatially semidiscrete Galerkin FEM is available [11–13], it is not the case for temporal discretization by the L1 time stepping. The L1 scheme was first analyzed in [20, 35], where the local truncation error was shown to be $O(\tau^{2-\alpha})$ for twice continuously differentiable (in time) solutions, which is fairly restrictive, cf. Remark A.1. Recently some error bounds that are expressed directly in terms of data regularity for the homogeneous problem were shown using a generating function approach [14], however, the analysis does not extend straightforwardly to the inhomogeneous case.

In this work we shall develop a novel energy argument for the L1 time stepping to overcome the technical challenge in the convergence analysis, which represents the main technical novelty. We shall derive optimal error estimates under realistic regularity conditions, and the analysis covers both smooth and nonsmooth problem data, cf. Theorem 3.5. Further, the stability result plays an essential role in deriving error estimates due to the POD approximation. All the theoretical results are fully confirmed by extensive numerical experiments.

The rest of the paper is organized as follows. In Section 2 we develop an efficient Galerkin-L1-POD scheme, and in Section 3, provide a complete error analysis of the scheme. In Section 4, extensive numerical experiments for one- and two-dimensional examples are presented to verify the convergence analysis. Finally, in an appendix, we briefly discuss the temporal regularity results for problem (1.1). Throughout, the notation c, with or without a subscript, denotes a generic constant, which may differ at different occurrences, but it is always independent of the solution u, the mesh size h, time step size τ, and the number m of POD basis functions.

2. An efficient Galerkin-L1-POD scheme

In this section, we develop an efficient numerical scheme, termed as the Galerkin-L1-POD scheme, for problem (1.1). It is based on the following three components: standard Galerkin method with continuous piecewise linear finite elements in space, L1 approximation in time and proper orthogonal decomposition in the snapshot space, which we shall describe separately in the following three subsections.

2.1. Space discretization by the Galerkin FEM

First we describe the spatial discretization based on the Galerkin FEM. Let $\mathcal{T}_h$ be a shape regular and quasi-uniform triangulation of the domain Ω into d-simplexes, known as finite elements and denoted by T. Then over the triangulation $\mathcal{T}_h$ we define a continuous piecewise linear finite element space X_h by

$$X_h = \big\{v_h \in H_0^1(\Omega) :\ v_h|_T \text{ is a linear function},\ \forall T \in \mathcal{T}_h\big\}.$$

On the space X_h, we define the $L^2(\Omega)$-orthogonal projection $P_h : L^2(\Omega) \to X_h$ by $(P_h\varphi, \chi) = (\varphi, \chi)$ for all $\chi \in X_h$. Then the semidiscrete Galerkin scheme for problem (1.1) reads: find $u_h(t) \in X_h$ such that

$$(\partial_t^\alpha u_h, \chi) + (\nabla u_h, \nabla\chi) = (f, \chi) \quad \forall \chi \in X_h,\ t > 0, \tag{2.1}$$

with $u_h(0) = v_h \in X_h$. Upon introducing the discrete Laplacian $\Delta_h : X_h \to X_h$ defined by $-(\Delta_h\varphi, \chi) = (\nabla\varphi, \nabla\chi)$ for all φ, $\chi \in X_h$, the semidiscrete scheme (2.1) can be rewritten into

$$\partial_t^\alpha u_h(t) + A_h u_h(t) = f_h(t) \quad t > 0, \tag{2.2}$$

with $u_h(0) = v_h \in X_h$, $f_h = P_h f$ and $A_h = -\Delta_h$.

2.2. Time discretization by L1 scheme

For the time discretization, we divide the interval $[0, T]$ into N equally spaced subintervals with a time step size $\tau = T/N$, and $t_n = n\tau$, $n = 0, \ldots, N$. Then the L1 scheme [20, 35] approximates the Caputo fractional derivative $\partial_t^\alpha u(x, t_n)$ by

$$\begin{aligned}
\partial_t^\alpha u(x,t_n) &= \frac{1}{\Gamma(1-\alpha)} \sum_{j=0}^{n-1} \int_{t_j}^{t_{j+1}} \frac{\partial u(x,s)}{\partial s} (t_n - s)^{-\alpha}\, ds \\
&\approx \frac{1}{\Gamma(1-\alpha)} \sum_{j=0}^{n-1} \frac{u(x,t_{j+1}) - u(x,t_j)}{\tau} \int_{t_j}^{t_{j+1}} (t_n - s)^{-\alpha} ds \\
&= \sum_{j=0}^{n-1} b_j \frac{u(x,t_{n-j}) - u(x,t_{n-j-1})}{\tau^\alpha \Gamma(2-\alpha)} =: \bar\partial_\tau^\alpha u(t_n),
\end{aligned} \tag{2.3}$$

where the weights $\{b_j\}$ are given by

$$b_j = (j+1)^{1-\alpha} - j^{1-\alpha}, \quad j = 0, 1, \ldots, n-1. \tag{2.4}$$

Then the fully discrete scheme reads: given $U_h^0 = v_h \in X_h$ and $F_h^n = P_h f(t_n) \in X_h$, with $c_\alpha = \Gamma(2-\alpha)$, find $U_h^n \in X_h$ for $n = 1, 2, \ldots, N$ such that

$$(b_0 I + c_\alpha \tau^\alpha A_h) U_h^n = b_{n-1} U_h^0 + \sum_{j=1}^{n-1} (b_{j-1} - b_j) U_h^{n-j} + c_\alpha \tau^\alpha F_h^n. \tag{2.5}$$

The computational challenge of the fully discrete scheme (2.5) is obvious: To compute the numerical solution U_h^n at t_n, the solutions $\{U_h^k\}_{k=0}^{n-1}$ at all preceding time instances are required, as a result of the nonlocality of the Caputo fractional derivative $\partial_t^\alpha u$. Hence, the computational complexity and storage requirement grow linearly as the number n of time steps increases, which poses a significant challenge especially for high-dimensional problems and multi-query applications. This naturally motivates the development of cheap reduced order models by the POD methodology so as to reduce the effective degree of freedom.

2.3. Galerkin-L1-POD scheme

Now we develop an efficient Galerkin approximation scheme based on proper orthogonal decomposition (POD) to circumvent the challenge. We shall first describe the general framework of the POD methodology, and then discuss its application to the subdiffusion equation.

POD is a powerful model reduction technique for complex models, especially time/parameter dependent partial differential equations. It resides on the empirical observation that despite the large apparent dimensionality of the solution space (e.g., the degree of freedom of the finite element approximation), the solution actually lives on an effectively much lower dimensional (possibly highly nonlinear) manifold. POD constructs a problem adapted basis for efficiently approximating the manifold using samples from the manifold, often known as "snapshots", which can be either solutions at different time instances, different parameter values, or samples

generated using relevant physical experiments. The POD basis functions are then employed within either a Galerkin or Petrov-Galerkin framework to generate a reduced-order model.

Now we recall the general framework of POD. Let X be a real Hilbert space endowed with an inner product $(\cdot,\cdot)_X$ and norm $\|\cdot\|_X$. Now for $N\in\mathbb{N}$, let $\{y_n\}_{n=1}^N\subset X$ be an ensemble of snapshots and at least one of them is assumed to be nonzero. Then we set $\mathfrak{U}=\operatorname{span}\{y_1,y_2,...,y_N\}\subset X$. Let $\dim(\mathfrak{U})=r$ and let $\{\psi_j\}_{j=1}^r$ be an orthonormal basis of the snapshot space $\mathfrak{U}$. Then any element y_n can be written as

$$y_n=\sum_{j=1}^{r}(y_n,\psi_j)_X\psi_j,\ \ n=1,2,...,N.$$

POD chooses an orthonormal basis $\{\psi_j\}_{j=1}^m$ for $1\le m\le r$ to minimize the following ensemble average:

$$\min_{\{\psi_j\}_{j=1}^m}\frac{1}{N}\sum_{n=1}^{N}\|y_n-\sum_{j=1}^{m}(y_n,\psi_j)_X\psi_j\|_X^2. \tag{2.6}$$

A solution of problem (2.6) is called a POD-basis of rank m. This optimization problem is related to the correlation matrix $K\in\mathbb{R}^{N\times N}$ corresponding to the snapshots $\{y_n\}_{n=1}^N$, which is defined by

$$K_{ij}=N^{-1}(y_j,y_i)_X,\quad i,j=1,\dots,N. \tag{2.7}$$

By its very construction, the matrix K is symmetric positive semidefinite, and its eigenvectors can be chosen to be orthonormal (in the inner product $(\cdot,\cdot)_X$). Further, the number of positive eigenvalues is equal to r, the dimensionality of the space $\mathfrak{U}$ spanned by the snapshots (or equivalently the rank of K). The following lemma gives the formula of the POD-basis and the corresponding approximation error within the ensemble [34].

Lemma 2.1. *Let $\lambda_1\ge\lambda_2\ge...\ge\lambda_r>0$ be the positive eigenvalues of the correlation matrix K and $v_1,...,v_r\in\mathbb{R}^N$ be the corresponding orthonormal eigenvectors. Then a POD basis of rank $m\le r$ is given by*

$$\psi_j=\frac{1}{\sqrt{\lambda_j}}\sum_{n=1}^{N}(v_j)_n y_n,$$

where $(v_j)_n$ denotes the n-th component of the eigenvector v_j. Moreover, the error is given by

$$\frac{1}{N}\sum_{n=1}^{N}\|y_n-\sum_{j=1}^{m}(y_n,\psi_j)_X\psi_j\|_X^2=\sum_{j=m+1}^{r}\lambda_j.$$

Following the abstract framework, for the subdiffusion model (1.1), we choose $2N+1$ snapshots as

$$y_n=U_h^{n-1},\quad n=1,2,...,N+1,$$

and the fractional difference quotients (FDQs)

$$y_n=\bar\partial_\tau^\alpha U_h^{n-N-1},\quad n=N+2,...,2N+1.$$

The inclusion of FDQs $\{\bar\partial_\tau^\alpha U_h^n\}$ into the snapshots $\mathfrak{U}$ is to improve the error estimate below: it allows directly bounding the error due to the POD approximation to the fractional derivative term $\bar\partial_\tau^\alpha U_h^n$, cf. Lemma 2.1. In the absence of these FDQs in the snapshots, the error estimate due to POD approximation would involve an additional factor $\tau^{-2\alpha}$; see Remark 3.3 for details. The use of difference quotients was first proposed by Kunisch and Volkwein [17] for the standard parabolic equation, and we refer interested readers to the recent

work [9] for extensive discussions. In this work, we shall follow the work [17], and employ the FDQs $\bar\partial_\tau^\alpha U_h^n$ in the construction of the POD basis.

In practice, there are several possible choices of the Hilbert space X, and we shall consider two popular ones in this work. Our first choice for the POD space is $X = H_0^1(\Omega)$ with the inner product $(u, v)_X = (\nabla u, \nabla v)$ for all $u,\ v \in H_0^1(\Omega)$. Then the correlation matrix $\widetilde{K}$ is given by

$$\widetilde{K}_{i,j} = (2N+1)^{-1}(\nabla y_j, \nabla y_i). \tag{2.8}$$

We denote the corresponding POD basis (called $H_0^1(\Omega)$ POD basis) by $\{\widetilde{\psi}_j\}_{j=1}^r$ and the subspace spanned by the first m $H_0^1(\Omega)$-POD basis functions by X_h^m, $m \le r$. Then Lemma 2.1 yields the following error estimate for the POD space X_h^m

$$\frac{1}{2N+1}\bigg(\sum_{n=0}^N \|U_h^n - \sum_{j=1}^m (\nabla U_h^n, \nabla\widetilde{\psi}_j)\widetilde{\psi}_j\|_{H_0^1(\Omega)}^2 + \sum_{n=1}^N \|\bar\partial_\tau^\alpha U_h^n - \sum_{j=1}^m (\nabla\bar\partial_\tau^\alpha U_h^n, \nabla\widetilde{\psi}_j)\widetilde{\psi}_j\|_{H_0^1(\Omega)}^2\bigg) = \sum_{j=m+1}^r \widetilde{\lambda}_j, \tag{2.9}$$

where $\{\widetilde{\lambda}_j\}_{j=1}^r$ are the descendingly ordered eigenvalues of the correlation matrix $\widetilde{K}$. The second choice is $X = L^2(\Omega)$ with the standard inner product. The correlation matrix $\widehat{K}$ is given by

$$\widehat{K}_{ij} = (2N+1)^{-1}(y_j, y_i). \tag{2.10}$$

Likewise, we denote the corresponding POD basis (called $L^2(\Omega)$-POD basis) by $\{\widehat{\psi}\}_{j=1}^r$, and by slightly abusing the notation, the subspace spanned by the first m $L^2(\Omega)$ POD basis functions by X_h^m. Then in view of Lemma 2.1, the POD space X_h^m satisfies the following error estimate

$$\frac{1}{2N+1}\bigg(\sum_{n=0}^N \|U_h^n - \sum_{j=1}^m (U_h^n, \widehat{\psi}_j)\widehat{\psi}_j\|_{L^2(\Omega)}^2 + \sum_{n=1}^N \|\bar\partial_\tau^\alpha U_h^n - \sum_{j=1}^m (\bar\partial_\tau^\alpha U_h^n, \widehat{\psi}_j)\widehat{\psi}_j\|_{L^2(\Omega)}^2\bigg) = \sum_{j=m+1}^r \widehat{\lambda}_j, \tag{2.11}$$

where $\{\widehat{\lambda}_j\}_{j=1}^r$ are the descendingly order eigenvalues of the correlation matrix $\widehat{K}$.

Next we define the Ritz projection operator $R_h^m : X_h \to X_h^m$ by

$$(\nabla R_h^m\chi, \nabla\varphi) = (\nabla\chi, \nabla\varphi) \quad \forall\varphi \in X_h^m, \tag{2.12}$$

where $\chi \in X_h \subset H_0^1(\Omega)$. The $H^1(\Omega)$-stability of the projection operator R_h^m on the space X_h is immediate

$$\|\nabla R_h^m\chi\|_{L^2(\Omega)} \le \|\nabla\chi\|_{L^2(\Omega)} \quad \forall\chi \in X_h.$$

Given the POD basis, one can exploit it for model reduction in several different ways. One natural choice is to use a Galerkin approach, which yields the following reduced-order formulation: with $U_m^0 = R_h^m v_h \in X_h^m$, find $U_m^n \in X_h^m$, $n = 1, 2, ..., N$ such that

$$(\bar\partial_\tau^\alpha U_m^n, \varphi_m) + (\nabla U_m^n, \nabla\varphi_m) = (f(t_n), \varphi_m) \quad \forall\varphi_m \in X_h^m, \tag{2.13}$$

or equivalently with $c_\alpha = \Gamma(2-\alpha)$,

$$b_0(U_m^n, \varphi_m) + c_\alpha\tau^\alpha(\nabla U_m^n, \nabla\varphi_m) = b_{n-1}(U_m^0, \varphi_m) + \sum_{j=1}^{n-1}(b_{j-1} - b_j)(U_m^{n-j}, \varphi_m) + c_\alpha\tau^\alpha(f(t_n), \varphi_m) \quad \forall\varphi_m \in X_h^m.$$

The existence and uniqueness of the POD approximation $\{U_m^n\}_{n=1}^N$ follows directly by an energy argument (see Section 3 below). In the Galerkin framework, the stiffness matrix of the reduced-order formulation (2.13) is the

projection of that of the global one (2.5) into the POD space X_h^m. It is worth mentioning that the degree of freedom of the reduced system (2.13) is m, i.e., the number of POD basis functions in X_h^m, which is usually much smaller than that of (2.5), i.e., the number of finite element basis functions. This shows clearly the enormous gain in the computational complexity and storage requirement of the proposed scheme.

3. Error analysis

In this part, we provide a complete error analysis of the proposed scheme (2.13). The discretization error consists of three sources: the spatial discretization, temporal discretization and POD approximation. It is known that the semidiscrete solution u_h satisfies the following nearly optimal error estimate [12,13], where the operator A is the negative Laplacian operator $-\Delta$ with a zero Dirichlet boundary condition. The log factor ℓ_h^2 in the error estimate is due to the limited smoothing property of the solution operator for subdiffusion, and the prefactor $t^{-\alpha(1-\sigma)}$, for $t \to 0$, reflects the corresponding solution singularity.

Theorem 3.1. *Let u be the solution of problem (1.1) with $A^\sigma v \in L^2(\Omega)$, $0 < \sigma \leq 1$, and $f \in L^\infty(0,T;L^2(\Omega))$, and u_h be the solution of problem (2.2) with $v_h = P_h v$ and $f_h = P_h f$. Then there holds with $\ell_h = |\log h|$*

$$\|u(t) - u_h(t)\|_{L^2(\Omega)} \leq ch^2\ell_h^2 \left(t^{-\alpha(1-\sigma)} \|A^\sigma v\|_{L^2(\Omega)} + \|f\|_{L^\infty(0,T;L^2(\Omega))} \right).$$

Below we derive the errors due to the temporal approximation and the POD approximation that are expressed in terms of the data regularity directly. The main novel ingredient in the convergence analysis is to establish a suitable stability result for the L1 time stepping under realistic assumptions on the data regularity. To this end, we shall develop a novel energy argument, based on the monotonicity of a suitable quadrature rule.

3.1. Error analysis of the L1 scheme

Now we develop a novel energy argument for analyzing the L1 approximation. We begin with a weighted inequality for the weights $\{b_j\}$, which is crucial for establishing the monotonicity of the quadrature below.

Lemma 3.1. *Let $\{b_j\}$ be defined by (2.4). Then for $j = 2, \ldots, n-1$, there holds*

$$(j-1)n^{\alpha-2}b_{j-1} + (n-j)n^{\alpha-2}b_j \leq (n+1)^{\alpha-1}b_j.$$

Proof. Using the definition of the weights b_j, the assertion is equivalent to: for all $j = 2, \ldots, n-1$:

$$\int_0^1 (j-1+t)^{-\alpha}(j-1) - \left(n\left(1+n^{-1}\right)^{\alpha-1} - n + j \right)(j+t)^{-\alpha} dt \leq 0,$$

that is,

$$\int_0^1 \frac{g(t)}{(j-1+t)^\alpha (j+t)^\alpha} dt \leq 0,$$

where the function $g : [0,1] \to \mathbb{R}$ is defined by $g(t) = (j-1)(j+t)^\alpha - (j-1+t)^\alpha(n(1+n^{-1})^{\alpha-1} - n + j)$, with its $g'(t)$ given by

$$g'(t) = \alpha \left[\frac{j-1}{(j+t)^{1-\alpha}} - \frac{n(1+n^{-1})^{\alpha-1} - n + j}{(j-1+t)^{1-\alpha}} \right].$$

For $\alpha \in (0,1)$, there holds $n(1+n^{-1})^{\alpha-1} - n + j \geq n^2(n+1)^{-1} - n + j = j - n(n+1)^{-1} > j-1$. Hence we deduce $g'(t) < 0$ on the interval $[0,1]$. It suffices to show that $g(0) \leq 0$. Obviously,

$$g(0) = (j-1)^\alpha \underbrace{((j-1)^{1-\alpha} j^\alpha - j + n\left(1 - (1+n^{-1})^{\alpha-1}\right))}_{\mathrm{I}}.$$

The term I in the bracket can be rewritten as

$$\mathrm{I} = j\left((1-j^{-1})^{1-\alpha}-1\right)+n\left(1-(1+n^{-1})^{\alpha-1}\right).$$

We claim that the function $\tilde{g}(j) = j\left(1-(1-j^{-1})^{1-\alpha}\right)$ is monotonically decreasing in j. To see this, let $h(t):(0,1)\to\mathbb{R}$, with $h(t)=t^{-1}(1-(1-t)^{1-\alpha})$. Then $h'(t)=-t^{-2}\left(1-(1-t)^{-\alpha}(1-\alpha t)\right)$. Next consider the function $\tilde{h}(t):(0,1)\to\mathbb{R}$, with $\tilde{h}(t)=(1-t)^{\alpha}$. Then $\tilde{h}'(t)=-\alpha(1-t)^{\alpha-1}$ and $\tilde{h}''(t)=(\alpha-1)\alpha(1-t)^{\alpha-2}<0$, namely, the function $\tilde{h}$ is concave. Then the concavity implies $\tilde{h}(t)\le\tilde{h}(0)+\tilde{h}'(0)t$, which gives $(1-t)^{\alpha}\le 1-\alpha t$. Consequently, $h'(t)\ge -t^{-2}(1-(1-\alpha t)^{-1}(1-\alpha t))\ge 0$, and hence h is monotonically increasing, and the monotonicity of the function $\tilde{g}(j)$ follows. Hence, by the trivial inequality $(n-1)/n<n/(n+1)$, we have

$$\begin{aligned}\mathrm{I} &< n((1-n^{-1})^{1-\alpha}-1)+n(1-(1+n^{-1})^{\alpha-1})\\ &= n\big((1-n^{-1})^{1-\alpha}-(1-(n+1)^{-1})^{1-\alpha}\big)<0,\end{aligned}$$

which concludes the proof of the lemma. □

Now we give an important monotonicity relation of a weighted rectangular quadrature approximation.

Theorem 3.2. *Let the function $f:[0,1]\to\mathbb{R}$ be convex and nonnegative with $f(0)=0$, and $\alpha\in(0,1)$. For any $n\in\mathbb{N}$, let $x_j=\frac{j}{n}$, $j=0,\ldots,n$, and $y_j=\frac{j}{n+1}$, $j=0,\ldots,n+1$. Then there holds*

$$n^{\alpha-1}\sum_{j=0}^{n-1}b_j f(x_j)\le(n+1)^{\alpha-1}\sum_{j=0}^{n}b_j f(y_j).$$

Proof. First we observe the trivial inequalities $\frac{j}{n+1}<\frac{j}{n}<\frac{j+1}{n+1}$, i.e., $y_j<x_j<y_{j+1}$, for $j=1,\ldots,n-1$. There also holds the trivial identity

$$x_j:=\frac{j}{n}=\frac{n-j}{n}\frac{j}{n+1}+\frac{j}{n}\frac{j+1}{n+1}=:\frac{n-j}{n}y_j+\frac{j}{n}y_{j+1}.$$

Now by the convexity of the function f, we deduce

$$f(x_j)=f\left(\frac{n-j}{n}y_j+\frac{j}{n}y_{j+1}\right)\le\frac{n-j}{n}f(y_j)+\frac{j}{n}f(y_{j+1}),\quad j=1,\ldots,n.$$

With the assumption $f(0)=0$, it suffices to consider $j\ge 1$ in the sum. Hence

$$\begin{aligned}n^{\alpha-1}\sum_{j=1}^{n-1}b_j f(x_j)&\le n^{\alpha-1}\sum_{j=1}^{n-1}b_j\left(\frac{n-j}{n}f(y_j)+\frac{j}{n}f(y_{j+1})\right)\\ &= n^{\alpha-1}\left(b_1\frac{n-1}{n}f(y_1)+\sum_{j=2}^{n-1}\left(b_{j-1}\frac{j-1}{n}+\frac{n-j}{n}b_j\right)f(y_j)+b_{n-1}\frac{n-1}{n}f(y_n)\right).\end{aligned}$$

To show the desired assertion, we consider the following three cases separately, first, last and middle terms. For the first term, in view of the nonnegativity of the function f, it suffices to show $n^{\alpha-1}\frac{n-1}{n}b_1\le(n+1)^{\alpha-1}b_1$, which however follows from $\alpha\in(0,1)$ and

$$(n+1)^{1-\alpha}n^{\alpha-1}\frac{n-1}{n}=\left(\frac{n+1}{n}\right)^{1-\alpha}\frac{n-1}{n}=\left(\frac{n^2-1}{n^2}\right)^{1-\alpha}\left(\frac{n-1}{n}\right)^{\alpha}<1.$$

For the last term, we have

$$n^{\alpha-1}\frac{n-1}{n}b_{n-1} = n^{\alpha-1}(n^{1-\alpha}-(n-1)^{1-\alpha})\frac{n-1}{n} = 1-\frac{1}{n}-\left(\frac{n-1}{n}\right)^{2-\alpha},$$

and meanwhile

$$(n+1)^{\alpha-1}b_n = (n+1)^{\alpha-1}((n+1)^{1-\alpha}-n^{1-\alpha}) = 1-\left(\frac{n}{n+1}\right)^{1-\alpha}.$$

Hence, it suffices to show $n^{-1}+(1-n^{-1})^{2-\alpha}-(1-(n+1)^{-1})^{1-\alpha}>0$ for $n>1$. Let $g:[0,1]\to\mathbb{R}$ by $g(t)=t+(1-t)^{2-\alpha}-(1+t)^{\alpha-1}$. Then $g(0)=0$, and $g'(t)=1-(2-\alpha)(1-t)^{1-\alpha}-(\alpha-1)(1+t)^{\alpha-2}$. Clearly $g'(0)=0$ and further $g''(t)=(2-\alpha)(1-\alpha)((1-t)^{-\alpha}-(1+t)^{\alpha-3})>0$, which in particular implies $g'(t)\geq 0$ on the interval $[0,1]$. To conclude the proof, it suffices to show the inequality for the middle terms, i.e., for $j=2,\ldots,n-1$

$$n^{\alpha-1}\frac{j-1}{n}b_{j-1}+n^{\alpha-1}\frac{n-j}{n}b_j\leq(n+1)^{\alpha-1}b_j,$$

which however is already shown in Lemma 3.1. □

The following result is a direct corollary from Theorem 3.2, and it will play a crucial role in establishing the stability result in Theorem 3.3 below.

Lemma 3.2. *For any $\alpha\in(0,1)$, let b_j be defined in (2.4). Then for any $n\in\mathbb{N}$, there holds*

$$\sum_{j=1}^{n}(b_{j-1}-b_j)(n+1-j)^{\alpha-1}\leq(n+1)^{\alpha-1}.$$

Proof. Consider the function $f(x)=(1-x)^{\alpha-1}-1$. Then it satisfies $f(x)\geq 0$, $f(0)=0$, and also $f''(x)>0$, i.e., convex. Hence, by Theorem 3.2, we have

$$n^{\alpha-1}\sum_{j=0}^{n-1}b_j\big((1-jn^{-1})^{\alpha-1}-1\big)\leq(n+1)^{\alpha-1}\sum_{j=1}^{n}b_j\big((1-j(n+1)^{-1})^{\alpha-1}-1\big). \tag{3.1}$$

Meanwhile, it can be verified directly that for all $n\in\mathbb{N}^+$, $n^{\alpha-1}\sum_{j=0}^{n-1}b_j=(1-\alpha)\int_0^1 x^{-\alpha}\,dx=1$, i.e., $n^{\alpha-1}\sum_{j=0}^{n-1}b_j=(n+1)^{\alpha-1}\sum_{j=0}^{n}b_j$. Plugging the preceding identity into (3.1) yields

$$n^{\alpha-1}\sum_{j=0}^{n-1}b_j\left(1-jn^{-1}\right)^{\alpha-1}\leq(n+1)^{\alpha-1}\sum_{j=0}^{n}b_j\left(1-j(n+1)^{-1}\right)^{\alpha-1}.$$

which upon rearranging terms gives the desired assertion. □

Next we give an important $L^2(\Omega)$ stability result. The stability estimate puts more weights on the source term F_h^k as the index k gets close to the current time step n, in a manner analogous to the continuous problem.

Theorem 3.3. *Let U_h^n, $n=1,2,\ldots,N$, be the solution of the fully discrete scheme (2.5). Then with $c_\alpha=\Gamma(2-\alpha)$, for $n=1,2,\ldots,N$, we have the following stability estimate*

$$\|U_h^n\|_{L^2(\Omega)}\leq\|v_h\|_{L^2(\Omega)}+c_\alpha\tau^\alpha\sum_{k=0}^{n-1}(n-k)^{\alpha-1}\|F_h^{k+1}\|_{L^2(\Omega)}. \tag{3.2}$$

Proof. We show the assertion by mathematical induction. First we consider the case $n = 1$. Multiplying both sides of (2.5) by U_h^1 and integrating over the domain Ω yield

$$\|U_h^1\|_{L^2(\Omega)}^2 + c_\alpha \tau^\alpha \|\nabla U_h^1\|_{L^2(\Omega)}^2 = (U_h^0, U_h^1) + c_\alpha \tau^\alpha (F_h^1, U_h^1).$$

Then the Cauchy-Schwartz inequality and Young's inequality give

$$\|U_h^1\|_{L^2(\Omega)} \le \|U_h^0\|_{L^2(\Omega)} + c_\alpha \tau^\alpha \|F_h^1\|_{L^2(\Omega)}.$$

Now assume the estimate holds up to some $n \ge 1$. A similar argument yields

$$\begin{aligned}
\|U_h^{n+1}\|_{L^2(\Omega)} &\le b_n \|U_h^0\|_{L^2(\Omega)} + \sum_{j=1}^{n} (b_{j-1} - b_j) \|U^{n+1-j}\|_{L^2(\Omega)} + c_\alpha \tau^\alpha \|F^{n+1}\|_{L^2(\Omega)} \\
&\le b_n \|U_h^0\|_{L^2(\Omega)} + \sum_{j=1}^{n} (b_{j-1} - b_j) \bigg(\|U_h^0\|_{L^2(\Omega)} + c_\alpha \tau^\alpha \|F^{n+1}\|_{L^2(\Omega)} \\
&\quad + c_\alpha \tau^\alpha \sum_{k=0}^{n-j} (n+1-j-k)^{\alpha-1} \|F_h^{k+1}\|_{L^2(\Omega)} \bigg) \\
&= \|U_h^0\|_{L^2(\Omega)} + c_\alpha \tau^\alpha \sum_{j=1}^{n} (b_{j-1} - b_j) \sum_{k=0}^{n-j} (n+1-j-k)^{\alpha-1} \|F_h^{k+1}\|_{L^2(\Omega)} + c_\alpha \tau^\alpha \|F^{n+1}\|_{L^2(\Omega)}.
\end{aligned}$$

Then by changing the order of summation and applying Lemma 3.2 we have

$$\begin{aligned}
\sum_{j=1}^{n} (b_{j-1} - b_j) \sum_{k=0}^{n-j} (n+1-j-k)^{\alpha-1} \|F_h^{k+1}\|_{L^2(\Omega)} &= \sum_{k=0}^{n-1} \|F_h^{k+1}\|_{L^2(\Omega)} \sum_{j=1}^{n-k} (b_{j-1} - b_j)(n+1-j-k)^{\alpha-1} \\
&\le \sum_{k=0}^{n-1} \|F_h^{k+1}\|_{L^2(\Omega)} (n+1-k)^{\alpha-1},
\end{aligned}$$

and consequently

$$\begin{aligned}
\|U_h^{n+1}\|_{L^2(\Omega)} &\le \|U_h^0\|_{L^2(\Omega)} + c_\alpha \tau^\alpha \sum_{k=0}^{n-1} (n+1-k)^{\alpha-1} \|F_h^{k+1}\|_{L^2(\Omega)} + c_\alpha \tau^\alpha \|F^{n+1}\|_{L^2(\Omega)} \\
&= \|U_h^0\|_{L^2(\Omega)} + c_\alpha \tau^\alpha \sum_{k=0}^{n} (n+1-k)^{\alpha-1} \|F_h^{k+1}\|_{L^2(\Omega)},
\end{aligned}$$

which completes the induction step and the desired assertion follows. □

The next lemma gives one useful estimate for bounding the local truncation error.

Lemma 3.3. *For any $\delta \in (0, \alpha]$, there exists a constant $c > 0$, independent of n, such that for all $n \ge 2$*

$$\sum_{k=1}^{n-1} \left((n-k)^{1-\alpha} - (n-k-1)^{1-\alpha} \right) k^{\delta-2} \le c(n-1)^{-\alpha}.$$

Proof. The case $n = 2$ is trivial, and we consider only $n \geq 3$. Let $d_k = ((n-k)^{1-\alpha} - (n-k-1)^{1-\alpha})k^{\delta-2}$. First, we observe that for $k = 1$

$$\begin{aligned} d_1 &= (n-1)^{1-\alpha} - (n-2)^{1-\alpha} = (1-\alpha)\int_1^2 (n-s)^{-\alpha}\,ds \\ &\leq c(n-2)^{-\alpha} \leq c((n-1)/3)^{-\alpha} \leq c(n-1)^{-\alpha}. \end{aligned}$$

The sum of the remaining terms can be bounded directly by

$$\begin{aligned} \sum_{k=2}^{n-1} d_k &= \sum_{k=2}^{n-1}(1-\alpha)k^{\delta-2}\int_k^{k+1}(n-s)^{-\alpha}\,ds \leq c\sum_{k=2}^{n-1}\int_k^{k+1}(n-s)^{-\alpha}(s-1)^{\delta-2}\,ds \\ &= c\int_2^n (n-s)^{-\alpha}(s-1)^{\delta-2}\,ds = c\int_1^{n-1}(n-s-1)^{-\alpha}s^{\delta-2}\,ds \\ &= c\int_1^{\frac{n-1}{2}}(n-s-1)^{-\alpha}s^{\delta-2}ds + c\int_{\frac{n-1}{2}}^{n-1}(n-s-1)^{-\alpha}s^{\delta-2}\,ds := \mathrm{I} + \mathrm{II}. \end{aligned}$$

Then the desired result follows from

$$\mathrm{I} \leq c\int_1^{\frac{n-1}{2}}(n-s-1)^{-\alpha}s^{\delta-2}\,ds \leq c(n-1)^{-\alpha}\int_1^{\frac{n-1}{2}} s^{\delta-2}\,ds \leq c(n-1)^{-\alpha}$$

and

$$\mathrm{II} \leq c(n-1)^{\delta-2}\int_{\frac{n-1}{2}}^{n-1}(n-s-1)^{-\alpha}\,ds \leq c(n-1)^{\delta-\alpha-1} \leq c(n-1)^{-\alpha}.$$

□

Next we derive an error bound on the local truncation error r_n defined by

$$r_n = \|\partial_t^\alpha u_h(t_n) - \bar\partial_\tau^\alpha u_h(t_n)\|_{L^2(\Omega)}, \qquad n = 1, 2, ..., N. \tag{3.3}$$

In view of Theorems A.1 and A.2 in the appendix, we make the following temporal regularity assumption.

Assumption 3.1. *The solution u satisfies the following smoothing properties*

$$\|u(t)\|_{L^2(\Omega)} \leq c \quad \textit{and} \quad \|\partial_t^m u(t)\|_{L^2(\Omega)} \leq ct^{\delta-m},$$

where $\delta > 0$ and the integer $m \geq 1$.

Remark 3.1. *By Theorems A.1 and A.2, the regularity condition in Assumption 3.1 holds with $\delta = \sigma\alpha$, $\sigma \in (0, 1]$, for initial data $v \in D(A^\sigma)$ and source term $f \in W^{2,\infty}(0, T; L^2(\Omega))$. Under these conditions, Assumption 3.1 holds also for the semidiscrete Galerkin approximation u_h, with a constant c independent of h.*

Lemma 3.4. *Let Assumption 3.1 hold, and r_n be the local truncation error defined by (3.3). Then*

$$r_n \leq \begin{cases} c\tau^{\delta-\alpha} & \textit{if } n = 1, \\ c(n-1)^{-\alpha}\tau^{\delta-\alpha} & \textit{if } n \geq 2. \end{cases}$$

Proof. Using Assumption 3.1, for $n = 1$, we have the following estimate (with $c'_\alpha = 1/\Gamma(2-\alpha)$)

$$\begin{aligned} r_1 &\le c'_\alpha \tau^{-1} \left\| \int_0^\tau (\tau - s)^{-\alpha} \int_0^\tau (u'_h(s) - u'_h(y))\, dy\, ds \right\|_{L^2(\Omega)} \\ &\le c'_\alpha \tau^{-1} \int_0^\tau (\tau - s)^{-\alpha} \int_0^\tau \|u'_h(s)\|_{L^2(\Omega)} + \|u'_h(y)\|_{L^2(\Omega)}\, dy\, ds \\ &\le c'_\alpha \tau^{-1} \int_0^\tau (\tau - s)^{-\alpha} \int_0^\tau (s^{\delta-1} + y^{\delta-1})\, dy\, ds \le c\tau^{\delta-\alpha}. \end{aligned} \tag{3.4}$$

Now we consider the case $n \ge 2$. Then

$$\begin{aligned} r_n &= c'_\alpha \| \sum_{k=0}^{n-1} \int_{t_k}^{t_{k+1}} (t_n - s)^{-\alpha} \left(u'_h(s) - \frac{u_h(t_{k+1}) - u_h(k)}{\tau} \right) ds \|_{L^2(\Omega)} \\ &\le c \sum_{k=0}^{n-1} \| \int_{t_k}^{t_{k+1}} (t_n - s)^{-\alpha} \left(u'_h(s) - \frac{u_h(t_{k+1}) - u_h(k)}{\tau} \right) ds \|_{L^2(\Omega)} := c \sum_{k=0}^{n-1} r_{n,k}. \end{aligned}$$

The first term $r_{n,0}$ can be bounded using Assumption 3.1 and the argument for (3.4) as

$$\begin{aligned} r_{n,0} &\le c \int_0^{t_1} (t_n - s)^{-\alpha} \|u'_h(s)\|_{L^2(\Omega)}\, ds + c\tau^{-1} \int_0^{t_1} (t_n - s)^{-\alpha} \int_0^{t_1} \|u'_h(y)\|_{L^2(\Omega)}\, dy\, ds \\ &\le c(t_n - t_1)^{-\alpha} \int_0^{t_1} s^{\delta-1}\, ds + c\tau^{\delta-1} \int_0^{t_1} (t_n - s)^{-\alpha}\, ds \le c(n-1)^{-\alpha} \tau^{\delta-\alpha}. \end{aligned} \tag{3.5}$$

Next we derive estimates for $r_{n,k}$, $k = 1, 2, ..., n-1$. To this end, we use the identity

$$u'_h(s) - \frac{u_h(t_{k+1}) - u_h(k)}{\tau} = \frac{1}{\tau} \int_{t_k}^{t_{k+1}} u'_h(s) - u'_h(y)\, dy = \frac{1}{\tau} \int_{t_k}^{t_{k+1}} \int_y^s u''_h(z)\, dz\, dy$$

and apply Assumption 3.1 such that $\|u''_h(z)\|_{L^2(\Omega)} \le c t_k^{\delta-2}$ with c independent of t and h to deduce

$$\left\| u'_h(s) - \frac{u_h(t_{k+1}) - u_h(k)}{\tau} \right\|_{L^2(\Omega)} \le \frac{1}{\tau} \int_{t_k}^{t_{k+1}} \int_{\min(s,y)}^{\max(s,y)} \|u''_h(z)\|_{L^2(\Omega)}\, dz\, dy \le c\tau t_k^{\delta-2}.$$

Thus we obtain

$$\begin{aligned} r_{n,k} &\le c\tau t_k^{\delta-2} \int_{t_k}^{t_{k+1}} (t_n - s)^{-\alpha}\, ds = c\tau^{2-\alpha} t_k^{\delta-2} \left((n-k)^{1-\alpha} - (n-k-1)^{1-\alpha} \right) \\ &= c\tau^{\delta-\alpha} k^{\delta-2} \left((n-k)^{1-\alpha} - (n-k-1)^{1-\alpha} \right). \end{aligned}$$

Then by Lemma 3.3 we deduce

$$\sum_{k=1}^{n-1} r_{n,k} \le c\tau^{\delta-\alpha} \sum_{k=1}^{n-1} k^{\delta-2} \left((n-k)^{1-\alpha} - (n-k-1)^{1-\alpha} \right) \le c\tau^{\delta-\alpha} (n-1)^{-\alpha}.$$

This together with (3.5) yields the desired estimate and hence completes the proof. □

Next we derive the error estimate $e_h^n = u_h(t_n) - U_h^n$, $n = 1, 2, ..., N$. First, we observe that the nodal error e_h^n satisfies $e_h^0 = 0$ and the following error equation

$$\bar\partial_t^\alpha e_h^n + A_h e_h^n = \bar\partial_t^\alpha u_h(t_n) - \partial_t^\alpha u_h(t_n).$$

The next theorem gives an optimal (uniform in time t) error estimate for the fully discrete scheme (2.5).

Theorem 3.4. *Assume $f \in W^{2,\infty}(0,T;L^2(\Omega))$ and $v \in D(A^\sigma)$, with $0 < \sigma \le 1$. Let u_h and U_h^n be the solutions of problems (2.2) and (2.5), respectively. Then there holds*

$$\|u_h(t_n) - U_h^n\|_{L^2(\Omega)} \le c\tau^{\sigma\alpha} \left(\|A^\sigma v\|_{L^2(\Omega)} + \|f\|_{W^{2,\infty}(0,T;L^2(\Omega))} \right).$$

Proof. By Theorem 3.3 and Lemma 3.4, with $\delta = \sigma\alpha$, we have

$$\begin{aligned}\|u_h(t_n) - U_h^n\|_{L^2(\Omega)} &\le c\tau^\alpha \sum_{k=0}^{n-1} (n-k)^{\alpha-1} \|\bar\partial_\tau^\alpha u_h(t_{k+1}) - \partial_t^\alpha u_h(t_{k+1})\|_{L^2(\Omega)} \\ &\le c\tau^{\sigma\alpha} \left(\|A^\sigma v\|_{L^2(\Omega)} + \|f\|_{W^{2,\infty}(0,T;L^2(\Omega))} \right) \left(1 + \sum_{k=1}^{n-1} (n-k)^{\alpha-1} k^{-\alpha} \right).\end{aligned}$$

Then the following uniform bound

$$\sum_{k=1}^{n-1} (n-k)^{\alpha-1} k^{-\alpha} = \frac{1}{n} \sum_{k=1}^{n-1} \left(1 - \frac{k}{n}\right)^{\alpha-1} \left(\frac{k}{n}\right)^{-\alpha} \le \int_0^1 (1-x)^{\alpha-1} x^{-\alpha} dx \le c$$

yields the desired estimate. □

Last, we can state an error estimate on the fully discrete approximation U_h^n, which follows from Theorems 3.1 and 3.4 by the triangle inequality.

Theorem 3.5. *Assume $f \in W^{2,\infty}(0,T;L^2(\Omega))$ and $v \in D(A^\sigma)$, with $0 < \sigma \le 1$. Let u and U_h^n be the solutions of problems (1.1) and (2.5), respectively. Then with $\ell_h = |\log h|$, there holds*

$$\|u(t_n) - U_h^n\|_{L^2(\Omega)} \le c(h^2 \ell_h t^{-\alpha(1-\sigma)} + \tau^{\sigma\alpha}) \|A^\sigma v\|_{L^2(\Omega)} + c(h^2 \ell_h^2 + \tau^{\sigma\alpha}) \|f\|_{W^{2,\infty}(0,T;L^2(\Omega))}.$$

3.2. **Error analysis of the POD approximation**

Next we derive the error estimates for the POD approximation U_m^n. First we recall an approximation property of the Ritz projection operator R_h^m defined in (2.12) within the ensemble [17, Lemma 3 and Corrolary 3].

Lemma 3.5. *For every $m = 1, ..., r$, the Ritz projection operator R_h^m satisfies*

$$\frac{1}{N} \sum_{n=1}^{N} \left(\|\nabla(U_h^n - R_h^m U_h^n)\|_{L^2(\Omega)}^2 + \|\nabla(\bar\partial_\tau^\alpha U_h^n - \bar\partial_\tau^\alpha R_h^m U_h^n)\|_{L^2(\Omega)}^2 \right) \le c \sum_{j=m+1}^{r} \widetilde\lambda_j$$

and

$$\frac{1}{N} \sum_{n=1}^{N} \left(\|\nabla(U_h^n - R_h^m U_h^n)\|_{L^2(\Omega)}^2 + \|\nabla(\bar\partial_\tau^\alpha U_h^n - \bar\partial_\tau^\alpha R_h^m U_h^n)\|_{L^2(\Omega)}^2 \right) \le c h^{-2} \sum_{j=m+1}^{r} \widehat\lambda_j$$

where $\{\widetilde\lambda_j\}_{j=1}^r$ and $\{\widehat\lambda_j\}_{j=1}^r$ denote the eigenvalues of $\widetilde K$ and $\widehat K$ defined in (2.8) and (2.10), respectively.

Now we can give the error estimate for the POD approximation U_m^n for smooth problem data. The result indicates that the error incurred by using the POD basis in place of the full Galerkin FEM basis is determined by the eigenvalues corresponding to the eigenfunctions that are not included in constructing the POD approximation. In particular, if the eigenvalues of the correlation matrix decay rapidly, then a small number of POD basis functions in the Galerkin POD scheme (2.13) suffice the desired accuracy.

Theorem 3.6. *Let u and U_m^n be the solutions of (1.1) and (2.13), respectively, and suppose that $v \in D(A)$, and $f \in W^{2,\infty}(0,T;L^2(\Omega))$. Then there holds*

$$\frac{1}{N}\sum_{n=1}^{N}\|u(t_n)-U_m^n\|_{L^2(\Omega)}^2 \le c_T\bigg(\tau^{2\alpha}+h^4\ell_h^4+\sum_{j=m+1}^{r}\widetilde{\lambda}_j\bigg) \tag{3.6}$$

and

$$\frac{1}{N}\sum_{n=1}^{N}\|u(t_n)-U_m^n\|_{L^2(\Omega)}^2 \le c_T\bigg(\tau^{2\alpha}+h^4\ell_h^4+h^{-2}\sum_{j=m+1}^{r}\widehat{\lambda}_j\bigg), \tag{3.7}$$

where $\{\widetilde{\lambda}_j\}_{j=1}^r$ and $\{\widehat{\lambda}_j\}_{j=1}^r$ denote the eigenvalues of $\widetilde{K}$ and $\widehat{K}$ defined in (2.8) and (2.10), respectively.

Proof. We split the error $e_m^n = u(t_n) - U_m^n$ into

$$e_m^n = (u(t_n)-U_h^n)+(U_h^n-U_m^n),$$

and the first term can be bounded using Theorem 3.5, i.e.,

$$\frac{1}{N}\sum_{n=1}^{N}\|u(t_n)-U_h^n\|_{L^2(\Omega)}^2 \le c\left(\tau^{2\alpha}+h^4\ell_h^4\right).$$

Hence it suffices to establish a bound for the second term $U_h^n - U_m^n$. Now we consider the splitting

$$U_h^n-U_m^n = (U_h^n-R_h^mU_h^n)+(R_h^mU_h^n-U_m^n) := \rho^n+\theta^n.$$

Then Lemma 3.5 yields the following bound on ρ^n as

$$\frac{1}{N}\sum_{n=1}^{N}\|\rho^n\|_{L^2(\Omega)}^2 \le c\sum_{j=m+1}^{r}\widetilde{\lambda}_j \qquad \text{and} \qquad \frac{1}{N}\sum_{n=1}^{N}\|\rho^n\|_{L^2(\Omega)}^2 \le ch^{-2}\sum_{j=m+1}^{r}\widehat{\lambda}_j, \tag{3.8}$$

for the $H_0^1(\Omega)$- and $L^2(\Omega)$-POD basis, respectively. Next we derive an estimate on the component θ^n. Using (2.13), the definition of the Ritz projection operator R_h^m, and the fact that $\varphi_m \in X_h^m \subset X_h$, we have

$$\begin{aligned}(\bar\partial_\tau^\alpha\theta^n,\varphi_m)+(\nabla\theta^n,\nabla\varphi_m) &= (\bar\partial_\tau^\alpha R_h^mU_h^n,\varphi_m)+(\nabla R_h^mU_h^n,\nabla\varphi_m)-(\bar\partial_\tau^\alpha U_m^n,\varphi_m)-(\nabla U_m^n,\nabla\varphi_m)\\ &= (\bar\partial_\tau^\alpha R_h^mU_h^n,\varphi_m)+(\nabla U_h^n,\nabla\varphi_m)-(f(t_n),\varphi_m)\\ &= (\bar\partial_\tau^\alpha(R_h^mU_h^n-U_h^n),\varphi_m) = -(\bar\partial_\tau^\alpha\rho^n,\varphi_m)\end{aligned}$$

and $\theta^0 = 0$. The stability result in Theorem 3.3 yields

$$\|\theta^n\|_{L^2(\Omega)} \le c\tau^\alpha\sum_{k=0}^{n-1}(n-k)^{\alpha-1}\|\bar\partial_\tau^\alpha\rho^{k+1}\|_{L^2(\Omega)}.$$

Appealing to Young's inequality for the Laplace type discrete convolution [7, Theorem 20.18], i.e.,

$$\sum_{n=0}^{N}\left(\sum_{k=0}^{n} a_{n-k}b_k\right)^2 \le \left(\sum_{n=0}^{N} a_n\right)^2 \sum_{n=0}^{N} b_n^2, \tag{3.9}$$

we deduce

$$\sum_{n=1}^{N}\left(\sum_{k=0}^{n-1}(n-k)^{\alpha-1}\|\bar\partial_\tau^\alpha \rho^{k+1}\|_{L^2(\Omega)}\right)^2 \le \left(\sum_{n=1}^{N} n^{\alpha-1}\right)^2 \sum_{n=1}^{N}\|\bar\partial_\tau^\alpha \rho^n\|_{L^2(\Omega)}^2 \le cN^{2\alpha}\sum_{n=1}^{N}\|\bar\partial_\tau^\alpha \rho^n\|_{L^2(\Omega)}^2.$$

Then by Lemma 3.5, we have

$$\frac{1}{N}\sum_{n=1}^{N}\|\theta^n\|_{L^2(\Omega)}^2 \le \frac{c\tau^{2\alpha}N^{2\alpha}}{N}\sum_{n=1}^{N}\|\bar\partial_\tau^\alpha \rho^n\|_{L^2(\Omega)}^2 = \frac{cT^{2\alpha}}{N}\sum_{n=1}^{N}\|\bar\partial_\tau^\alpha \rho^n\|_{L^2(\Omega)}^2 \le c_T\sum_{j=m+1}^{r}\widetilde{\lambda}_j.$$

Likewise, for the $L^2(\Omega)$-POD basis, we deduce

$$\frac{1}{N}\sum_{n=1}^{N}\|\theta^n\|_{L^2(\Omega)}^2 \le \frac{cT^{2\alpha}}{N}\sum_{n=1}^{N}\|\bar\partial_\tau^\alpha \rho^n\|_{L^2(\Omega)}^2 \le c_T h^{-2}\sum_{j=m+1}^{r}\widehat{\lambda}_j.$$

This completes the proof of the theorem. □

The error estimate in Theorem 3.6 covers only smooth initial data $v \in D(A)$. In the case of nonsmooth initial data $v \in D(A^\sigma)$, $0 < \sigma < 1$, one can derive an analogous error estimate; see the following remark. We note that the regularity of problem data (or solution) does not enter the error estimate due to the POD approximation directly. Hence, in principle, the approach is capable of handling nonsmooth problem data, if the solution singularity is built-in in the ensemble of snapshots and thus captured by the POD basis directly.

Remark 3.2. *We comment on nonsmooth problem data. Consider the $H_0^1(\Omega)$ POD for $f \in W^{2,\infty}(0,T;L^2(\Omega))$ and nonsmooth initial data $v \in D(A^\sigma)$, $0 < \sigma < 1$. Then in view of Theorem 3.5, we have*

$$\frac{1}{N}\sum_{n=1}^{N}\|u(t_n) - U_h^n\|_{L^2(\Omega)}^2 \le c\big(\tau^{2\sigma\alpha} + h^4\ell_h^4\frac{1}{N}\sum_{n=1}^{N} t_n^{-2\alpha(1-\sigma)}\big).$$

Meanwhile, the summation can be bounded as

$$\frac{1}{N}\sum_{n=1}^{N} t_n^{-2\alpha(1-\sigma)} = \frac{\tau^{-2\alpha(1-\sigma)}}{N}\sum_{n=1}^{N} n^{-2\alpha(1-\sigma)} \le \frac{\tau^{-2\alpha(1-\sigma)}}{N}\int_1^N s^{-2\alpha(1-\sigma)}ds \le c_T\ell_{\alpha,\sigma,\tau},$$

where the constant $\ell_{\alpha,\sigma,\tau}$ is given by

$$\ell_{\alpha,\sigma,\tau} = \begin{cases} \tau^{1-2\alpha(1-\sigma)}, & \alpha(1-\sigma) > 1/2,\\ \log\frac{T}{\tau}, & \alpha(1-\sigma) = 1/2,\\ 1, & \alpha(1-\sigma) < 1/2. \end{cases}$$

Consequently, by repeating the arguments in Theorem 3.6, we obtain the following error estimate for the POD approximation $\{U_m^n\}$ (with the $H_0^1(\Omega)$ POD basis)

$$\frac{1}{N}\sum_{n=1}^{N}\|u(t_n)-U_m^n\|_{L^2(\Omega)}^2 \le c_T\bigg(\tau^{2\sigma\alpha}+h^4\ell_h^4\ell_{\alpha,\sigma,\tau}+\sum_{j=m+1}^{r}\widetilde{\lambda}_j\bigg),$$

and a similar error estimate holds for the $L^2(\Omega)$ POD basis. Interestingly, for the case $\alpha(1-\sigma)<1/2$, the error estimate in the space remains uniform with respect to the time step size τ.

Last we briefly comment on the case when the FDQs are not included in the snapshots.

Remark 3.3. *In our construction of the POD basis, we have included the FDQs in the snapshots. When the FDQs $\bar\partial_\tau^\alpha U_h^n$, $n=1,2,...,N$, are not contained in the snapshot set, the error formula (2.9) for $H_0^1(\Omega)$ POD basis becomes*

$$\frac{1}{N+1}\sum_{n=0}^{N}\|U_h^n-\sum_{j=1}^{m}(\nabla U_h^n,\nabla\widetilde{\psi}_j)\widetilde{\psi}_j\|_{H_0^1(\Omega)}^2=\sum_{j=m+1}^{r}\widetilde{\lambda}_j.$$

Further for the FDQs we have

$$\frac{1}{N}\sum_{n=1}^{N}\|\bar\partial_\tau^\alpha U_h^n-\sum_{j=1}^{m}(\nabla\bar\partial_\tau^\alpha U_h^n,\nabla\widetilde{\psi}_j)\widetilde{\psi}_j\|_{H_0^1(\Omega)}^2=\frac{1}{N}\sum_{n=1}^{N}\|\bar\partial_\tau^\alpha\big(U_h^n-\sum_{j=1}^{m}(\nabla U_h^n,\nabla\widetilde{\psi}_j)\widetilde{\psi}_j\big)\|_{H_0^1(\Omega)}^2$$

Let $\overline{U}_h^n=U_h^n-\sum_{j=1}^{m}(\nabla U_h^n,\nabla\widetilde{\psi}_j)\widetilde{\psi}_j$. By the monotonicity of the weights $\{b_j\}$, we have

$$\begin{aligned}\Big\|\bar\partial_\tau^\alpha\overline{U}_h^n\Big\|_{H_0^1(\Omega)}^2&\le c_\alpha\tau^{-2\alpha}\bigg(b_0\|\overline{U}_h^n\|_{H_0^1(\Omega)}+b_{n-1}\|\overline{U}_h^0\|_{H_0^1(\Omega)}+\sum_{j=1}^{n-1}(b_{j-1}-b_j)\|\overline{U}_h^{n-j}\|_{H_0^1(\Omega)}\bigg)^2\\&\le c_\alpha\tau^{-2\alpha}b_n^2\|\overline{U}_h^0\|_{H_0^1(\Omega)}^2+c_\alpha\tau^{-2\alpha}\bigg(\sum_{j=0}^{n}g_j\|\overline{U}_h^{n-j}\|_{H_0^1(\Omega)}\bigg)^2,\end{aligned}$$

with $g_j=b_{j-1}-b_j$ and $b_{-1}=2$. Then by Young's inequality for discrete convolution, cf. (3.9), we arrive at

$$\frac{1}{N}\sum_{n=1}^{N}\bigg(\sum_{j=0}^{n}g_j\|\overline{U}_h^{n-j}\|_{H_0^1(\Omega)}\bigg)^2\le\frac{1}{N}\bigg(\sum_{n=0}^{N}g_j\bigg)^2\sum_{n=0}^{N}\|\overline{U}_h^{n-j}\|_{H_0^1(\Omega)}^2\le\frac{c}{N+1}\sum_{n=0}^{N}\|\overline{U}_h^n\|_{H_0^1(\Omega)}^2.$$

Meanwhile, by the Cauchy-Schwarz inequality, we have

$$\sum_{n=1}^{N}b_n^2=(1-\alpha)^{-2}\sum_{n=1}^{N}\bigg(\int_n^{n+1}s^{-\alpha}\,ds\bigg)^2\le c\int_1^{N+1}s^{-2\alpha}\,ds\le\begin{cases}cN^{1-2\alpha}, & \text{if } \alpha<1/2,\\ c\log N, & \text{if } \alpha=1/2,\\ c, & \text{if } \alpha>1/2.\end{cases}$$

Consequently, there holds

$$\frac{1}{N}\sum_{n=1}^{N}\|\bar\partial_\tau^\alpha U_h^n-\sum_{j=1}^{m}(\nabla\bar\partial_\tau^\alpha U_h^n,\nabla\widetilde{\psi}_j)\widetilde{\psi}_j\|_{H_0^1(\Omega)}^2\le c_T\big(\ell_{\alpha,\tau}\|\overline{U}_h^0\|_{H_0^1(\Omega)}^2+\tau^{-2\alpha}\sum_{j=m+1}^{r}\widetilde{\lambda}_j\big),$$

where the constant $\ell_{\alpha,\tau}$ *is given by*

$$\ell_{\alpha,\tau} = \begin{cases} 1 & \text{if } \alpha < 1/2, \\ \log \frac{T}{\tau} & \text{if } \alpha = 1/2, \\ \tau^{-2\alpha+1} & \text{if } \alpha > 1/2. \end{cases}$$

For $\alpha \le 1/2$*, the term involving the initial data* U_h^0 *is of higher order in comparison with the last term. Hence the error for* $H_0^1(\Omega)$ *Galerkin POD* (2.13) *(without FDQs in the snapshots) can be bounded by*

$$\frac{1}{N}\sum_{n=1}^{N} \|u(t_n) - U_m^n\|_{L^2(\Omega)}^2 \le c_T\bigg(\tau^{2\alpha} + h^4\ell_h^4 + \ell_{\alpha,\tau}\|U_h^0 - \sum_{j=1}^{m}(\nabla U_h^0, \nabla\widetilde{\psi}_j)\widetilde{\psi}_j\|_{H_0^1(\Omega)}^2 + \tau^{-2\alpha}\sum_{j=m+1}^{r}\widetilde{\lambda}_j\bigg).$$

In comparison with the error estimate (3.6) *with FDQs from Theorem 3.6, this estimate contains an extra factor* $\tau^{-2\alpha}$ *and an approximation error of the initial data* v_h *(within the POD basis* X_h^m*). For the fractional order* $\alpha \to 1$*, the factor recovers that for the classical diffusion equation [17].*

4. Numerical results

Now we present numerical results to verify the convergence theory in Section 3 and the efficiency of the proposed Galerkin-L1-POD scheme.

4.1. Numerical results for one-dimensional examples

First we present numerical results for one-dimensional examples to verify the convergence analysis in Section 3. We consider the subdiffusion model in the following two cases:

(a) $\Omega = (0,1)$, $v = x(1-x) \in D(A)$, and $f(x,t) = e^{t\cos(2\pi x)} \in W^{2,\infty}(0,T;L^2(\Omega))$;
(b) $\Omega = (0,1)$, $v = \chi_{(0,1/2)}(x) \in D(A^{1/4-\epsilon})$ for $\epsilon \in (0,1/4)$, and $f(x,t) = e^{t\cos(2\pi x)} \in W^{2,\infty}(0,T;L^2(\Omega))$.

In the computations, we divide the unit interval Ω into M equally spaced subintervals with a mesh size $h = 1/M$. Likewise, we fix the time step size τ at $\tau = T/N$.

First we examine the temporal convergence by setting $T = 0.1$ (the spatial convergence was already examined in [12,13]). We take a small mesh size $h = 10^{-3}$, so that the spatial discretization error is negligible. The exact solution can be expressed in terms of the Mittag-Leffler function $E_{\alpha,\beta}(z)$, cf. (A.1), which can be evaluated efficiently by an algorithm developed in [32]. The numerical results by the fully discrete scheme (2.5) are given in Table 1. In the table, `rate` refers to the empirical rate when the time step size τ halves, and the numbers in the bracket denote the theoretical predictions from Theorem 3.5. For cases (a) and (b), the empirical rate is $O(\tau^{\alpha})$ and $O(\tau^{\alpha/4})$, respectively, which agree well with the theoretical ones. The convergence rate of the L1 scheme improves with the smoothness of the initial data v (while keeping the smooth right hand side f fixed) and the increase of the fractional order α, since the solution regularity improves accordingly.

Table 1. The maximum error $e_{\max} = \max_{1\le n\le N}\|U_h^n - u(t_n)\|_{L^2(\Omega)}$ for initial data (a) and (b) with $T = 0.1$, $h = 10^{-3}$, $\tau = T/N$.

α	N	1000	2000	4000	8000	16000	32000	rate
0.35	(a)	2.67e-3	2.27e-3	1.90e-3	1.58e-3	1.29e-3	1.05e-3	$\approx$ 0.29 (0.35)
	(b)	2.48e-2	2.41e-2	2.29e-2	2.15e-2	1.99e-2	1.82e-2	$\approx$ 0.10 (0.09)
0.5	(a)	9.26e-4	6.73e-4	4.86e-4	3.50e-4	2.51e-4	1.80e-4	$\approx$ 0.48 (0.50)
	(b)	2.03e-2	1.81e-2	1.64e-2	1.50e-2	1.37e-2	1.26e-2	$\approx$ 0.13 (0.13)
0.75	(a)	1.82e-4	1.09e-4	6.43e-5	3.77e-5	2.17e-5	1.25e-5	$\approx$ 0.76 (0.75)
	(b)	2.52e-2	2.20e-2	1.91e-2	1.64e-2	1.39e-2	1.15e-2	$\approx$ 0.21 (0.19)

Next we illustrate the proposed Galerkin-L1-POD scheme, and the numerical results are given in Table 2 for the choice $T = 1$ and $N = 200$. Here the average error e and the POD approximation error e^m are defined by

$$e = \frac{1}{N}\sum_{n=1}^{N} \|U_h^n - u(t_n)\|_{L^2(\Omega)}^2 \quad \text{and} \quad e^m = \frac{1}{N}\sum_{n=1}^{N} \|U_h^n - U_m^n\|_{L^2(\Omega)}^2,$$

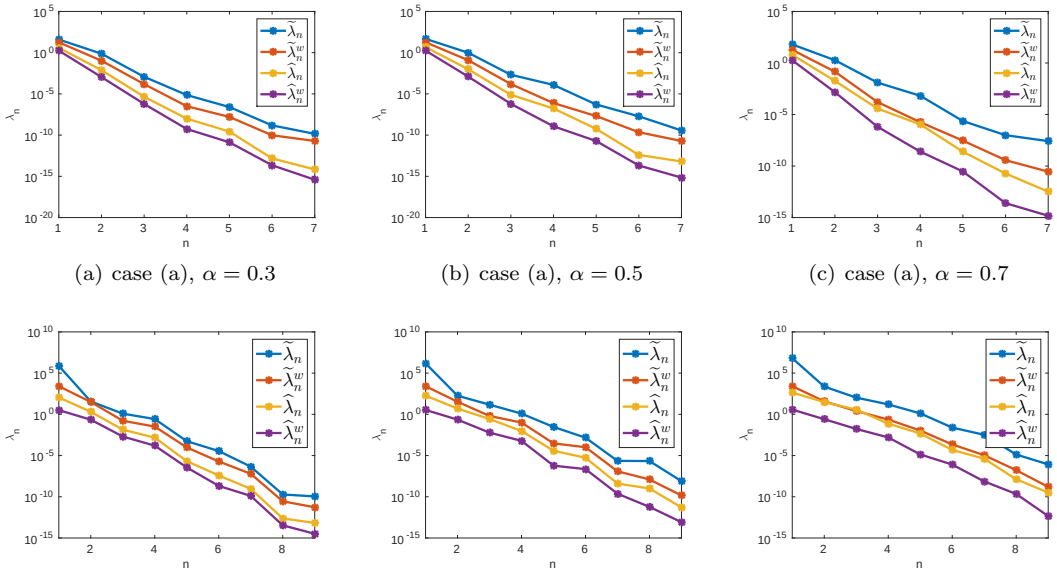
respectively. Like before, we use the notation $\widetilde{\ }$ and $\widehat{\ }$ over e^m to denote $H_0^1(\Omega)$- and $L^2(\Omega)$-POD basis, respectively, and the subscript w to indicate that the snapshots do not contain FDQs. For example, $\widetilde{e}^m$ and $\widetilde{e}_w^m$ denote the error between the full Galerkin solution U_h^n and the solution of the Galerkin POD formulation with m $H_0^1(\Omega)$ POD basis functions, with and without FDQs, respectively. For both cases (a) and (b), with three or four POD basis functions, the POD approximation error falls below the error due to temporal discretization, and the convergence is relatively independent of the fractional order α. The fast convergence of the Galerkin POD scheme is also expected from the exponential decay of the eigenvalues of the correlation matrix, cf. Fig. 1. Further, the inclusion of FDQs does not affect much the POD approximation error, with their errors within a factor of ten, even though their presence improves the apparent theoretical convergence rates, cf. Theorem 3.6 and Remark 3.3. The effect seems to be compensated by the smaller eigenvalues, cf. Fig. 1. These observations show the efficiency of the Galerkin POD scheme, which has only a degree of freedom of three or four at each time level, compared with one thousand for the standard Galerkin FEM.

For case (b), the Galerkin POD scheme requires slightly more POD basis functions in order to reach the same level of the accuracy. This is expected, since for nonsmooth data v, it can only be accurately described by more Fourier modes, and all these modes persist in the dynamics due to the "slow" decay of subdiffusion. Hence the solution manifold may exhibit richer structure than case (a), and consequently, more POD basis functions are needed to accurately capture the dynamics. However, the eigenvalues in the nonsmooth case decays also exponentially, cf. Fig. 1. Hence, the proposed scheme also works well with low regularity data.

The efficiency of the proposed scheme relies crucially on constructing "good" POD basis. To this end, we present the first five POD basis functions for case (b) in Fig. 2. The $H_0^1(\Omega)$- and $L^2(\Omega)$ POD basis take very different shapes: for the $H_0^1(\Omega)$ POD, the first basis function captures the singularity (caused by the discontinuous initial data), whereas the higher POD modes are very smooth. In contrast, for the $L^2(\Omega)$ POD, all the first five POD basis functions contain singularities (in the middle of the interval as well as oscillations around the end points). Namely, the $H_0^1(\Omega)$ POD seems to better aggregate the solution singularity (actually into one single POD basis). Nonetheless, the $L^2(\Omega)$ and $H_0^1(\Omega)$ POD-basis exhibit quite similar approximation property, and thus can provide equally good approximations of the solution manifold, cf. Table 2.

TABLE 2. The numerical results of the Galerkin POD for cases (a) and (b) with $T = 1$, $h = 10^{-3}$, $N = 200$, and with m POD basis functions.

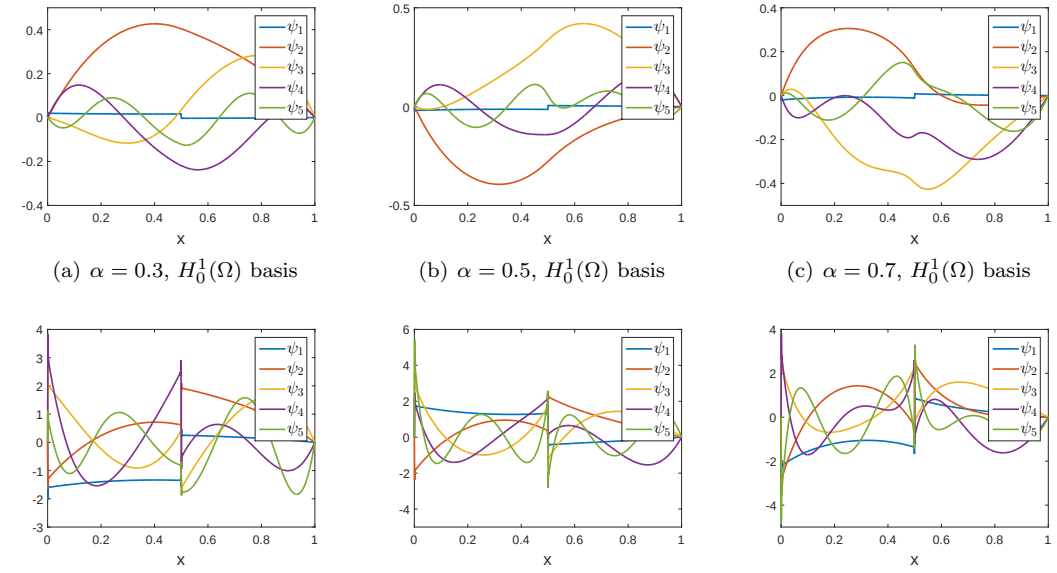
α	case	m	e	$\widetilde{e}^m$	$\widetilde{e}_w^m$	$\widehat{e}^m$	$\widehat{e}_w^m$
0.3	(a)	3	1.82e-7	9.34e-12	3.03e-12	9.45e-12	3.02e-12
		4	1.82e-7	4.72e-13	3.71e-14	4.83e-13	3.19e-14
	(b)	3	3.83e-6	4.65e-6	3.59e-6	4.36e-6	3.60e-6
		4	3.83e-6	2.73e-9	2.41e-9	2.73e-9	2.41e-9
0.5	(a)	3	4.46e-7	1.01e-10	6.25e-12	1.11e-10	6.22e-12
		4	4.46e-7	5.33e-13	8.87e-14	5.41e-13	8.28e-14
	(b)	3	1.70e-5	1.81e-5	6.70e-6	1.59e-5	7.08e-6
		4	1.70e-5	3.67e-8	6.70e-9	3.43e-8	6.69e-9
0.7	(a)	3	2.89e-7	4.70e-10	1.35e-11	4.98e-10	1.34e-11
		4	2.89e-7	1.33e-12	1.85e-13	1.29e-12	1.81e-13
	(b)	4	2.80e-5	2.51e-5	1.83e-7	1.45e-5	1.78e-7
		5	2.80e-5	2.49e-8	5.00e-9	2.42e-8	4.99e-9



(a) case (a), $\alpha = 0.3$ (b) case (a), $\alpha = 0.5$ (c) case (a), $\alpha = 0.7$

(d) case (b), $\alpha = 0.3$ (e) case (b), $\alpha = 0.5$ (f) case (b), $\alpha = 0.7$

FIGURE 1. The decay of eigenvalues of the correlation matrix in the 1D problem with $\alpha = 0.3$, 0.5 and 0.7. Here, $\widetilde{\lambda}_n$, $\widetilde{\lambda}_n^w$, $\widehat{\lambda}_n$, and $\widehat{\lambda}_n^w$ denote eigenvalues of correlation matrix for $H_0^1(\Omega)$ POD basis with or without FDQs and $L^2(\Omega)$ POD basis with or without FDQs, respectively.



(a) $\alpha = 0.3$, $H_0^1(\Omega)$ basis (b) $\alpha = 0.5$, $H_0^1(\Omega)$ basis (c) $\alpha = 0.7$, $H_0^1(\Omega)$ basis

(d) $\alpha = 0.3$, $L^2(\Omega)$ basis (e) $\alpha = 0.5$, $L^2(\Omega)$ basis (f) $\alpha = 0.7$, $L^2(\Omega)$ basis

FIGURE 2. The first five POD basis functions, in the $H_0^1(\Omega)$ and $L^2(\Omega)$ norms for case (b), with FDQs included in the basis construction.

4.2. **Numerical results for one two-dimensional example**

Now we present numerical results for the following two-dimensional example:

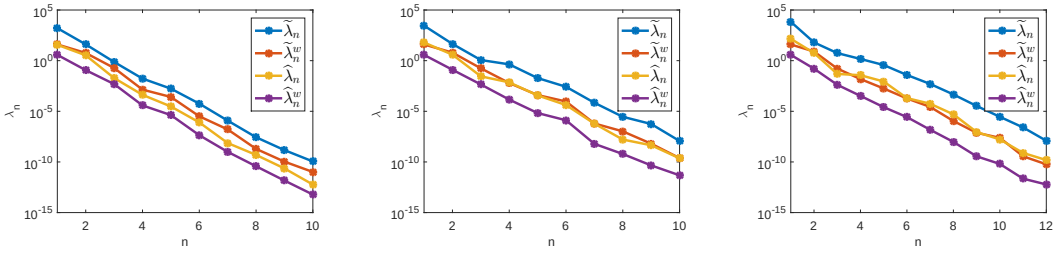
(c) $\Omega = (-1,1)^2 \backslash ([0,1]\times[-1,0])$, $v(x_1,x_2) = x_1(1+x_1)(1-x_1)\sin(2\pi x_2)$, $f(x_1,x_2,t) = e^{t\cos(2\pi x_1)\sin(\pi x_2)} \in W^{2,\infty}(0,T;L^2(\Omega))$, and $T = 1$.

In the computations, we divide the L-shaped domain Ω into a triangulation with a degree of freedom 10^4, and fix the time step size τ at $\tau = T/200$. A reentrant corner with an angle $\omega \in (\pi, 2\pi)$ induces a singularity associated with the corresponding stationary Poisson's problem [5]. In example (c), the angle $\omega = 3\pi/2$, and the reentrant corner gives rise to a singularity near the origin with a leading term of the form $r^{2/3}\sin(2\theta/3)$ in polar coordinates. Hence, we refine the mesh adaptively using the bisection rule [27, Section 4.1]. We compute the reference solution on a more refined mesh with 2×10^4 and $\tau = 1/1000$.

The numerical results are shown in Table 3. The POD scheme exhibits a fast convergence, and the error decreases steadily with the increase of the number m of POD basis functions. In particular, five or six POD basis functions suffice to resolve the solution manifold to an accuracy $O(10^{-9})$, which clearly shows the efficiency of the Galerkin POD scheme, when compared with the standard Galerkin FEM. The fast convergence follows also from the exponential decay of the eigenvalues of the correlation matrix, cf. Fig. 3. The decay rate of the spectrum is almost identical for the $L^2(\Omega)$ and $H^1(\Omega)$ POD basis, and independent of the presence of the FDQs. Hence, the presence of geometrical singularities in the domain does not influence the efficiency of the Galerkin-L1-POD scheme. Interestingly, we observe that with the increase of the fractional order α, the error increases slightly, which awaits further theoretical justification.

TABLE 3. The numerical results of the Galerkin POD for case (c), with $T = 1$, $N = 200$ and with m POD basis functions.

α	m	e	$\widetilde{e}^m$	$\widetilde{e}^m_w$	$\widehat{e}^m$	$\widehat{e}^m_w$
0.3	5	7.67e-7	5.36e-10	3.33e-10	5.17e-10	3.32e-10
	6	7.67e-7	6.40e-12	5.49e-12	6.39e-12	5.48e-12
0.5	5	4.75e-6	2.08e-8	8.23e-9	1.96e-8	8.18e-9
	6	4.75e-6	1.62e-10	4.82e-11	1.44e-10	4.79e-11
0.7	6	1.01e-5	2.05e-8	1.36e-9	1.38e-8	1.27e-9
	7	1.01e-5	9.11e-10	1.17e-10	6.09e-10	1.11e-10



(a) $\alpha = 0.3$ (b) $\alpha = 0.5$ (c) $\alpha = 0.7$

FIGURE 3. The decay of the eigenvalues of the correlation matrix for case (c) (2D problem on an L-shaped domain), with $\alpha = 0.3$, 0.5 and 0.7. Here $\widetilde{\lambda}_n$, $\widetilde{\lambda}^w_n$, $\widehat{\lambda}_n$, and $\widehat{\lambda}^w_n$ denote eigenvalues of correlation matrix for $H^1_0(\Omega)$ POD basis with or without FDQs and $L^2(\Omega)$ POD basis with or without FDQs, respectively.

4.3. **Numerical results for a perturbed problem**

Last, we illustrate the proposed Galerkin POD scheme with a perturbed problem, where the snapshots are generated using a problem setting different from the one of interest, as typically occurs in optimal control and inverse problems. Let $\delta_n(x) = n(2\cosh^2(nx))^{-1}$ be an approximate Dirac delta function.

(d) On the domain Ω is $\Omega = (0,1)^2$, we consider the following problem:

$$\partial_t^\alpha u - \Delta u + qu = f \quad \text{in } \Omega$$

with $q(x_1,x_2) = 1 + \cos(\pi x_1)\sin(2\pi x_2)$, $f(x_1,x_2,t) = \delta_2(x_1 - \frac{1}{2})\delta_2(x_2 - \frac{1}{2})e^{\cos(t)}$ and $v(x_1,x_2) = x_1(1-x_1)\sin(2\pi x_2)$ and $T = 1$. However, the snapshots are generated using a perturbed source term $\tilde{f}(x_1,x_2,t) = \delta_{10}(x_1 - \frac{1}{2})\delta_{10}(x_2 - \frac{1}{2})$.

In our computation, we divide the sides of the domain Ω into 100 equal subintervals, each of length 10^{-2}, thus dividing Ω into 10^4 small squares, and obtain a uniform triangulation by connecting parallel diagonals of each small square. The time step size τ is fixed as $\tau = T/200$.

TABLE 4. The numerical results of the Galerkin POD for case (d), with $T = 1$, $N = 200$ and with m POD basis functions.

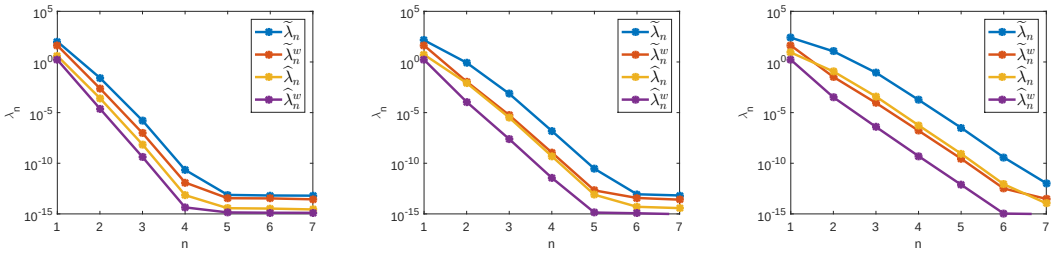
α	m	$\widetilde{e}^m$	$\widetilde{e}^m_w$	$\widehat{e}^m$	$\widehat{e}^m_w$
0.3	4	4.63e-7	4.64e-7	4.63e-7	4.64e-7
	5	3.32e-7	4.50e-7	3.21e-7	3.34e-7
0.5	4	4.47e-7	4.52e-7	4.47e-7	4.53e-7
	5	3.50e-7	3.46e-7	3.50e-8	3.45e-7
0.4	4	4.12e-7	4.32e-7	4.12e-7	4.32e-7
	5	3.81e-7	3.71e-7	3.80e-7	3.71e-7

Since the snapshots are generated from a perturbed problem, the error estimates in Theorem 3.6 do not apply directly. Nonetheless, one can still observe a fast decay of the POD approximation error, and with four to five POD basis functions, the error is already much smaller than the L1 time stepping, cf. Table 4, for both $L^2(\Omega)$- and $H_0^1(\Omega)$-POD basis and with/without FDQs. The high efficiency of the proposed scheme is attributed to the intrinsic low-dimensionality of the solution manifold, which is fully captured by the snapshots generated from the perturbed problem. This is also expected from the fast decay of the eigenvalues of the correlation matrix (from the perturbed problem) in Fig. 4. The solution profiles and corresponding errors are shown in Fig. 5. This example shows clearly the potential of the proposed approach for solving related inverse problems and optimal control, where many analogous forward problems have to be solved.

5. CONCLUDING REMARKS

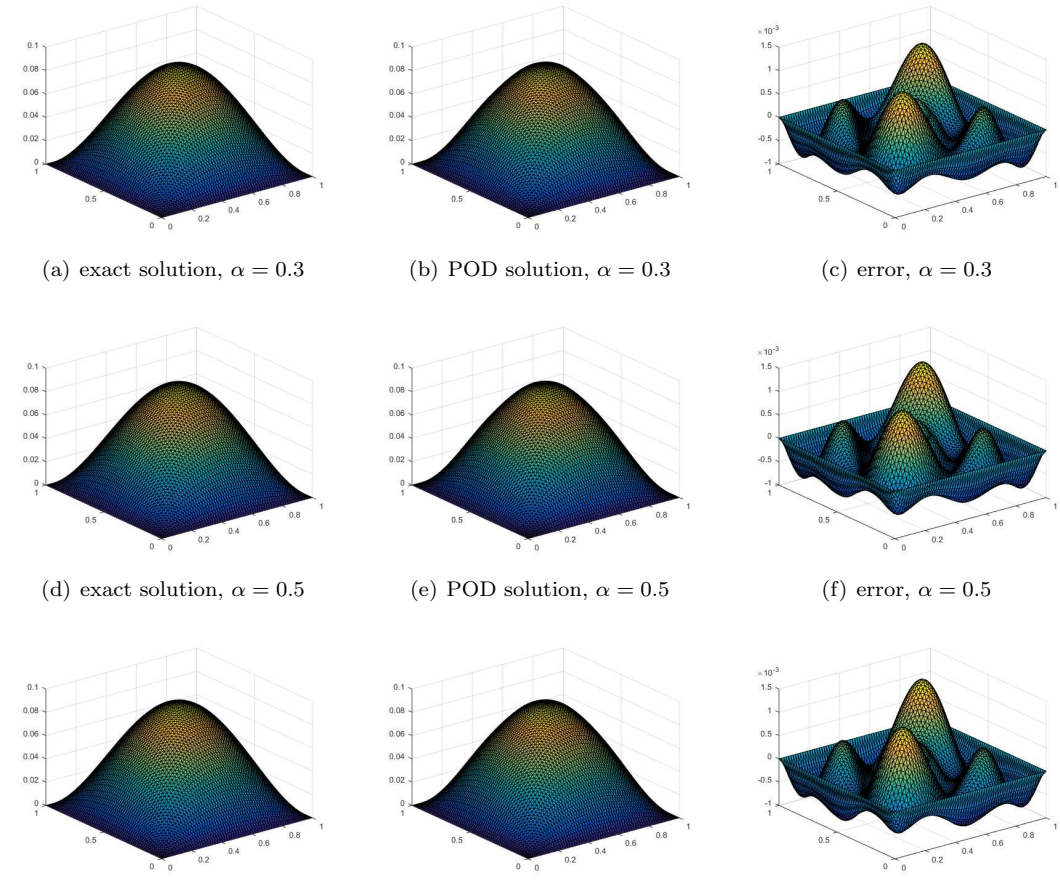
In this work, we have developed an efficient Galerkin-L1-POD scheme for solving the subdiffusion problem, by coupling the Galerkin finite element method, L1 time stepping and proper orthogonal decomposition. It realizes the computational efficiency by constructing an effective reduced-order model using POD, often with a very small degree of freedom. We provided a complete error analysis of the scheme, and derived optimal error estimates due to spatial discretization, temporal discretization and POD approximation. This is achieved by developing a novel energy argument for L1 time stepping. The extensive numerical experiments fully confirmed the convergence analysis and the efficiency and robustness of the scheme.

The work represents only a first step towards effective model reduction strategies for fractional differential equations. The choice of the three components in the proposed scheme is not unique. Alternatively, one may employ finite difference methods or spectral methods instead of the finite element method, and convolution quadrature type schemes instead of the L1 time scheme. The overall framework extends straightforwardly to



(a) $\alpha = 0.3$ (b) $\alpha = 0.5$ (c) $\alpha = 0.7$

FIGURE 4. The decay of the eigenvalues of the correlation matrix for case (d) with $\alpha = 0.3$, 0.5 and 0.7. Here $\widetilde{\lambda}_n$, $\widetilde{\lambda}_n^w$, $\widehat{\lambda}_n$, and $\widehat{\lambda}_n^w$ denote eigenvalues of correlation matrix for $H_0^1(\Omega)$ POD basis with or without FDQs and $L^2(\Omega)$ POD basis with or without FDQs, respectively.



(a) exact solution, $\alpha = 0.3$ (b) POD solution, $\alpha = 0.3$ (c) error, $\alpha = 0.3$

(d) exact solution, $\alpha = 0.5$ (e) POD solution, $\alpha = 0.5$ (f) error, $\alpha = 0.5$

(g) exact solution, $\alpha = 0.7$ (h) POD solution, $\alpha = 0.7$ (i) error, $\alpha = 0.7$

FIGURE 5. Exact and numerical solutions at $T = 1$ for case (d), where the POD solutions are obtained using $H_0^1(\Omega)$ POD basis with the FDQs.

these alternative choices, even though the convergence analysis will differ. Further, it is of much interest to extend the proposed scheme to more complex models, e.g., the multi-term model and the diffusion-wave model.

Acknowledgements

The work of the first author (B. Jin) is partly supported by EPSRC grant EP/M025160/1. The authors are grateful to the referees for their constructive comments.

Appendix A. Regularity theory for problem (1.1)

Now we describe temporal regularity results of problem (1.1) which plays an important role in the convergence analysis. Let $\{(\lambda_j,\varphi_j)\}_{j=1}^{\infty}$ be the eigenvalue pairs of the negative Laplacian $A=-\Delta$ with a homogeneous Dirichlet boundary condition, where the set $\{\varphi_j\}_{j=1}^{\infty}$ forms an orthonormal basis in $L^2(\Omega)$. Then by the standard separation of variable technique, we deduce that the solution u can be represented by

$$u(t)=E(t)v+\int_0^t \bar{E}(t-s)f(s)ds,$$

where the solution operators $E(t)$ and $\bar{E}(t)$ are given by

$$E(t)\psi=\sum_{j=1}^{\infty}E_{\alpha,1}(-\lambda_j t^{\alpha})(\psi,\varphi_j)\varphi_j \quad \text{and} \quad \bar{E}(t)\psi=\sum_{j=1}^{\infty}t^{\alpha-1}E_{\alpha,\alpha}(-\lambda_j t^{\alpha})(\psi,\varphi_j)\varphi_j, \tag{A.1}$$

respectively. Here the Mittag-Leffler function $E_{\alpha,\beta}(z)$, $\alpha>0$, $\beta\in\mathbb{R}$, is defined by [15, pp. 42] $E_{\alpha,\beta}(z)=\sum_{k=0}^{\infty}\frac{z^k}{\Gamma(k\alpha+\beta)}$. The following relations hold (see [31, Lemma 3.2] and [15, pp. 43, eq. (1.8.28)] for proofs).

Lemma A.1. *Let $\alpha\in(0,1)$, and $\beta\in\mathbb{R}$. The Mittag-Leffler function $E_{\alpha,\beta}(z)$ satisfies for $m\geq 1$*

$$\frac{d^m}{dt^m}E_{\alpha,1}(-\lambda t^{\alpha})=-\lambda t^{\alpha-m}E_{\alpha,\alpha+1-m}(-\lambda t^{\alpha}) \quad t>0,$$

and the following uniform bound on the negative real axis $\mathbb{R}^-$ holds

$$E_{\alpha,\beta}(z)\leq c(1+|z|)^{-1} \quad \forall z\in\mathbb{R}^-.$$

Now we can state the temporal regularity for the homogeneous problem.

Theorem A.1. *If $v\in D(A^{\sigma})$ and $f\equiv 0$, then*

$$\|\partial_t^m u\|_{L^2(\Omega)}\leq ct^{\sigma\alpha-m}\|A^{\sigma}v\|_{L^2(\Omega)}, \tag{A.2}$$

where if $\sigma\in(0,1]$, $m\geq 1$, and if $\sigma=0$, $0\leq m\leq 2$.

Proof. The case $\sigma=0$ has been shown [31, Corollary 2.6]. For $\sigma\in(0,1]$, by Lemma A.1, we have

$$\begin{aligned}\|\partial_t^m u\|_{L^2(\Omega)}^2&=\|\sum_{j=1}^{\infty}\frac{d^m}{dt^m}E_{\alpha,1}(-\lambda_j t^{\alpha})(v,\varphi_j)\varphi_j\|_{L^2(\Omega)}^2=\sum_{j=1}^{\infty}\lambda_j^2 t^{2\alpha-2m}E_{\alpha,\alpha-m+1}(-\lambda_j t^{\alpha})^2(v,\varphi_j)^2\\&=\sum_{j=1}^{\infty}(\lambda_j t^{\alpha})^{2-2\sigma}t^{2\sigma\alpha-2m}E_{\alpha,\alpha-m+1}(-\lambda_j t^{\alpha})^2(v,\varphi_j)^2\lambda_j^{2\sigma}\\&\leq ct^{2\sigma\alpha-2m}\sup_j\frac{(\lambda_j t^{\alpha})^{2-2\sigma}}{(1+\lambda_j t^{\alpha})^2}\sum_{j=1}^{\infty}(v,\varphi_j)^2\lambda_j^{2\sigma}\leq ct^{2\sigma\alpha-2m}\|A^{\sigma}v\|_{L^2(\Omega)}^2,\end{aligned}$$

where the last inequality follows from the inequality $\sup_j (\lambda_j t^\alpha)^{2-2\sigma}/(1+\lambda_j t^\alpha)^2 \le c$. □

Next we consider the inhomogeneous problem. We shall need the following estimate on $\bar{E}(t)$

Lemma A.2. *For any $t > 0$, we have for $\chi \in L^2(\Omega)$ and $m \ge 0$*

$$\|\partial_t^m \bar{E}(t)\chi\|_{L^2(\Omega)} \le ct^{\alpha-m-1}\|\chi\|_{L^2(\Omega)}.$$

Proof. The definition of the operator $\bar{E}$ in (A.1) and Lemma A.1 yield

$$\begin{aligned}\|\partial_t^m \bar{E}(t)\chi\|_{L^2(\Omega)}^2 &= \sum_{j=1}^\infty |t^{\alpha-m-1}E_{\alpha,\alpha-m}(-\lambda_j t^\alpha)|^2 |(\chi,\varphi_j)|^2\\ &\le ct^{2\alpha-2m-2}\sum_{j=1}^\infty |(\chi,\varphi_j)|^2 = ct^{2\alpha-2m-2}\|\chi\|_{L^2(\Omega)}^2,\end{aligned}$$

which completes the proof of the lemma. □

Now we can state the temporal regularity result for the inhomogeneous problem.

Theorem A.2. *If $v \equiv 0$ and $f \in W^{m,\infty}(0,T;L^2(\Omega))$ with some $m \in [0,2]$, then there holds*

$$\|\partial_t^m u\|_{L^2(\Omega)} \le c_T t^{\alpha-m}\|f\|_{W^{m,\infty}(0,T;L^2(\Omega))}, \quad 0 \le m \le 2. \tag{A.3}$$

Proof. Using the following convolution relation [22, Lemma 5.2]

$$t(f*g)' = f*g + (tf')*g + f*(tg')$$

and Lemma A.2, we deduce that for $t \in (0,T]$

$$\begin{aligned}t^m\|\partial_t^m u\|_{L^2(\Omega)} &\le c\sum_{p+q\le m}\int_0^t \|(t-s)^p\partial_t^p \bar{E}(t-s)(s^q f^{(m)}(s))\|_{L^2(\Omega)}\,ds\\ &\le c\sum_{p+q\le m}\int_0^t (t-s)^{\alpha-1}s^q\|f^{(m)}(s))\|_{L^2(\Omega)}\,ds \le c\|f\|_{W^{m,\infty}(0,T;L^2(\Omega))}\sum_{p+q\le m} t^{\alpha+q}.\end{aligned}$$

Since for $t \in (0,T]$, we have $\sum_{p+q\le m} t^{\alpha+q-m} \le c_T t^{\alpha-m}$, the desired assertion follows. □

Remark A.1. *Theorems A.1 and A.2 show the limited smoothing property of the subdiffusion model (1.1): for the homogeneous problem with $v \in D(A)$, the first order derivative in time t of the solution u exhibits a singularity of the form $t^{\alpha-1}$; and for the inhomogeneous problem with $f \in W^{2,\infty}(0,T;L^2(\Omega))$, the first-order derivative exhibits a similar singularity, despite the smoothness of f in time.*

References

[1] D. Amsallem and U. Hetmaniuk. Error estimates for Galerkin reduced-order models of the semi-discrete wave equation. *ESAIM Math. Model. Numer. Anal.*, 48(1):135–163, 2014.

[2] B. Berkowitz, A. Cortis, M. Dentz, and H. Scher. Modeling non-Fickian transport in geological formations as a continuous time random walk. *Rev. Geophys.*, 44(2):RG2003, 49 pp., 2006.

[3] D. Chapelle, A. Gariah, and J. Sainte-Marie. Galerkin approximation with proper orthogonal decomposition: new error estimates and illustrative examples. *ESAIM Math. Model. Numer. Anal.*, 46(4):731–757, 2012.

[4] N. J. Ford and A. C. Simpson. The numerical solution of fractional differential equations: speed versus accuracy. *Numer. Algorithms*, 26(4):333–346, 2001.

[5] P. Grisvard. *Elliptic Problems in Nonsmooth Domains*. Pitman, Boston, MA, 1985.

[6] Y. Hatano and N. Hatano. Dispersive transport of ions in column experiments: An explanation of long-tailed profiles. *Water Res. Research*, 34(5):1027–1033, 1998.
[7] E. Hewitt and K. A. Ross. *Abstract Harmonic Analysis. Vol. I: Structure of Topological Groups. Integration Theory, Group Representations*. Springer-Verlag, Berlin, 1963.
[8] M. Hinze and S. Volkwein. Error estimates for abstract linear-quadratic optimal control problems using proper orthogonal decomposition. *Comput. Optim. Appl.*, 39(3):319–345, 2008.
[9] T. Iliescu and Z. Wang. Are the snapshot difference quotients needed in the proper orthogonal decomposition? *SIAM J. Sci. Comput.*, 36(3):A1221–A1250, 2014.
[10] B. Jin. Fast Bayesian approach for parameter estimation. *Int. J. Numer. Methods Eng.*, 76(2):230–252, 2008.
[11] B. Jin, R. Lazarov, Y. Liu, and Z. Zhou. The Galerkin finite element method for a multi-term time-fractional diffusion equation. *J. Comput. Phys.*, 281:825–843, 2015.
[12] B. Jin, R. Lazarov, J. Pasciak, and Z. Zhou. Error analysis of semidiscrete finite element methods for inhomogeneous time-fractional diffusion. *IMA J. Numer. Anal.*, 35(2):561–582, 2015.
[13] 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.
[14] B. Jin, R. Lazarov, and Z. Zhou. An analysis of the L1 scheme for the subdiffusion equation with nonsmooth data. *IMA J. Numer. Anal.*, page dru063, 2015.
[15] A. Kilbas, H. Srivastava, and J. Trujillo. *Theory and Applications of Fractional Differential Equations*. Elsevier, Amsterdam, 2006.
[16] K. Kunisch and S. Volkwein. Control of the Burgers equation by a reduced-order approach using proper orthogonal decomposition. *J. Optim. Theory Appl.*, 102(2):345–371, 1999.
[17] K. Kunisch and S. Volkwein. Galerkin proper orthogonal decomposition methods for parabolic problems. *Numer. Math.*, 90(1):117–148, 2001.
[18] K. Kunisch and S. Volkwein. Galerkin proper orthogonal decomposition methods for a general equation in fluid dynamics. *SIAM J. Numer. Anal.*, 40(2):492–515, 2002.
[19] K. Kunisch and S. Volkwein. Proper orthogonal decomposition for optimality systems. *M2AN Math. Model. Numer. Anal.*, 42(1):1–23, 2008.
[20] Y. Lin and C. Xu. Finite difference/spectral approximations for the time-fractional diffusion equation. *J. Comput. Phys.*, 225(2):1533–1552, 2007.
[21] M. López-Fernández, C. Lubich, and A. Schädle. Adaptive, fast, and oblivious convolution in evolution equations with memory. *SIAM J. Sci. Comput.*, 30(2):1015–1037, 2008.
[22] W. McLean. Regularity of solutions to a time-fractional diffusion equation. *ANZIAM J.*, 52(2):123–138, 2010.
[23] W. McLean. Fast summation by interval clustering for an evolution equation with memory. *SIAM J. Sci. Comput.*, 34(6):A3039–A3056, 2012.
[24] R. Metzler and J. Klafter. The random walk's guide to anomalous diffusion: a fractional dynamics approach. *Phys. Rep.*, 339(1):77, 2000.
[25] P. Moireau and D. Chapelle. Reduced-order unscented Kalman filtering with application to parameter identification in large-dimensional systems. *ESAIM Control Optim. Calc. Var.*, 17(2):380–405, 2011.
[26] R. Nigmatulin. The realization of the generalized transfer equation in a medium with fractal geometry. *Phys. Stat. Sol. B*, 133:425–430, 1986.
[27] R. H. Nochetto, K. G. Siebert, and A. Veeser. Theory of adaptive finite element methods: an introduction. In *Multiscale, Nonlinear and Adaptive Approximation*, pages 409–542. Springer, 2009.
[28] I. Podlubny. *Fractional Differential Equations*. Academic Press, Inc., San Diego, CA, 1999.
[29] E. W. Sachs and M. Schu. *A priori* error estimates for reduced order models in finance. *ESAIM Math. Model. Numer. Anal.*, 47(2):449–469, 2013.
[30] E. W. Sachs and S. Volkwein. POD-Galerkin approximations in PDE-constrained optimization. *GAMM-Mitt.*, 33(2):194–208, 2010.
[31] K. Sakamoto and M. Yamamoto. Initial value/boundary value problems for fractional diffusion-wave equations and applications to some inverse problems. *J. Math. Anal. Appl.*, 382(1):426–447, 2011.
[32] H. Seybold and R. Hilfer. Numerical algorithm for calculating the generalized Mittag-Leffler function. *SIAM J. Numer. Anal.*, 47(1):69–88, 2009.
[33] J. R. Singler. New POD error expressions, error bounds, and asymptotic results for reduced order models of parabolic PDEs. *SIAM J. Numer. Anal.*, 52(2):852–876, 2014.
[34] L. Sirovich. Turbulence and the dynamics of coherent structures. part i: Coherent structures. *Quart. Appl. Math.*, 45(3):561–571, 1987.
[35] Z.-Z. Sun and X. Wu. A fully discrete scheme for a diffusion wave system. *Appl. Numer. Math.*, 56(2):193–209, 2006.