ERROR ESTIMATES FOR A SEMIDISCRETE FINITE ELEMENT METHOD FOR FRACTIONAL ORDER PARABOLIC EQUATIONS

BANGTI JIN, RAYTCHO LAZAROV, AND ZHI ZHOU

ABSTRACT. We consider the initial boundary value problem for the homogeneous time-fractional diffusion equation $\partial_t^\alpha u - \Delta u = 0$ ($0 < \alpha < 1$) with initial condition $u(x, 0) = v(x)$ and a homogeneous Dirichlet boundary condition in a bounded polygonal domain Ω. We shall study two semidiscrete approximation schemes, i.e., Galerkin FEM and lumped mass Galerkin FEM, by using piecewise linear functions. We establish optimal with respect to the regularity of the solution error estimates, including the case of nonsmooth initial data, i.e., $v \in L_2(\Omega)$.

1. INTRODUCTION

We consider the model initial–boundary value problem for the fractional order parabolic differential equation (FPDE) for $u(x, t)$:

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

where Ω is a bounded polygonal domain in $\mathbb{R}^d$ ($d = 1, 2, 3$) with a boundary $\partial\Omega$ and v is a given function on 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. [12, p. 91] or [22, p. 78])

$$
\partial_t^\alpha u(t) = \frac{1}{\Gamma(1-\alpha)} \int_0^t (t - \tau)^{-\alpha} \frac{d}{d\tau} u(\tau) \, d\tau,
$$

where $\Gamma(\cdot)$ is the Gamma function. Note that if the fractional order α tends to unity, the fractional derivative $\partial_t^\alpha u$ converges to the canonical first-order derivative $\frac{du}{dt}$ [12], and thus the problem (1.1) reproduces the standard parabolic equation. The model (1.1) is known to capture well the dynamics of anomalous diffusion (also known as sub-diffusion) in which the mean square variance grows slower than that in a Gaussian process [1], and has found a number of important practical applications. For example, it was introduced

by Nigmatulin [21] to describe diffusion in media with fractal geometry. A comprehensive survey on fractional order differential equations arising in dynamical systems in control theory, electrical circuits with fractance, generalized voltage divider, viscoelasticity, fractional-order multipoles in electromagnetism, electrochemistry, and model of neurons in biology is provided in [5]; see also [22].

The capabilities of FPDEs to accurately model such processes have generated considerable interest in deriving, analyzing and testing numerical methods for solving such problems. As a result, a number of numerical techniques were developed and their stability and convergence were investigated, see e.g. [6, 14, 15, 19, 20, 27]. Yuste and Acedo in [27] presented a numerical scheme by combining the forward time centered space method and the Grunwald-Letnikov method, and provided a von Neumann type stability analysis. By exploiting the variational framework introduced by Ervin and Roop, [9], Li and Xu [15] developed a spectral approximation method in both temporal and spatial variable, and established various a priori error estimates. Deng [6] analyzed the finite element method (FEM) for space- and time-fractional Fokker-Plank equation, and established a convergence rate of $O(\tau^{2-\alpha} + h^\mu)$, with $\alpha \in (0,1)$ and $\mu \in (1,2)$ being the temporal and spatial fractional order, respectively.

In all these useful studies, the error analysis was done by assuming that the solution is sufficiently smooth. The optimality of the established estimates with respect to the smoothness of the solution expressed through the problem data, i.e., the right hand side f and the initial data v, was not considered. Thus, these studies do not cover the interesting case of solutions with limited regularity due to low regularity of the data, a typical case for inverse problems related to this equation; see e.g., [3], [23, Problem (4.12)], and also [10, 11] for its parabolic counterpart.

There are a few papers considering construction and analysis of numerical methods with optimal with respect to the regularity of the solution error estimates for the following equation with a positive type memory terms [17, 18, 20]):

$$(1.2) \qquad \partial_t u - \frac{1}{\Gamma(\alpha)} \int_0^t (t-\tau)^{\alpha-1} \Delta u(\tau) d\tau = f(x,t),\ t > 0,\ 0 < \alpha < 1,$$

This equation is closely related, but different from (1.1). For example, McLean and Thomée in [17, 18] developed a numerical method based on spatial finite element discretization and Laplace transformation with quadratures in time for (1.2) with a homogeneous Dirichlet boundary data. In [17, Theorem 5.1] the convergence of the proposed method has been studied and maximum-norm error estimates of order $O(t^{-1-\alpha}h^2\ell_h^2)$, $\ell_h = |\ln h|$, were established for initial data $v \in L_\infty(\Omega)$. Further, in [18, Theorem 4.2] a maximum-norm error estimate of order $O(h^2\ell_h^2)$ was shown for smooth initial data $v \in \dot{H}^2$. Mustafa [20] studied a semidiscrete in time and and fully

discrete schemes, Crank-Nicolson in time and finite elements in space, and derived error bounds for smooth initial data; see, e.g. [20, Theorem 4.3].

The lack of optimal with respect to the regularity error estimates for the numerical schemes for FPDEs with nonsmooth data is in sharp contrast with the finite element method (FEM) for standard parabolic problems, $\alpha = 1$. Here the error analysis is complete and various optimal with respect to the regularity of the solution estimates are available [25]. The key inngredient of the analysis is the smoothing property of the parabolic operator and its discrete counterpart [25, Lemmas 3.2 and 2.5]. For the FPDE (1.1), such property has been established recently by Sakamoto and Yamamoto [23]; see Theorem 2.1 below for details.

The goal of this note is to develop an error analysis with optimal with respect to the regularity of the initial data estimates for the semidiscrete Galerkin and the lumped mass Galerkin FEMs for the problem (1.1) on convex polygonal domains.

Now we describe our main results. We shall use the standard notations in the finite element method [25]. Let $\{\mathcal{T}_h\}_{0<h<1}$ be a family of regular partitions of the domain Ω into d-simplexes, called finite elements, with h denoting the maximum diameter. Throughout, we assume that the triangulation $\mathcal{T}_h$ is quasi-uniform, that is the diameter of the inscribed disk in the finite element $\tau \in \mathcal{T}_h$ is bounded from below by h, uniformly on $\mathcal{T}_h$. The approximate solution u_h will be sought in the finite element space $X_h \equiv X_h(\Omega)$ of continuous piecewise linear functions over the triangulation $\mathcal{T}_h$

$$X_h = \left\{ \chi \in H_0^1(\Omega) : \ \chi \text{ is a linear function over } \tau, \ \ \forall \tau \in \mathcal{T}_h \right\}.$$

The semidiscrete Galerkin FEM for the problem (1.1) is: find $u_h(t) \in X_h$ such that

(1.3)
$$(\partial_t^\alpha u_h, \chi) + a(u_h, \chi) = (f, \chi), \quad \forall \chi \in X_h, \ T \geq t > 0,$$
$$u_h(0) = v_h,$$

where $a(u, w) = (\nabla u, \nabla w)$ for $u, w \in H_0^1(\Omega)$, and $v_h \in X_h$ is a given approximation of the initial data v. The choice of v_h will depend on the smoothness of the initial data v. Following Thomée [25], we shall take $v_h = R_h v$ in case of smooth initial data and $v_h = P_h v$ in case of nonsmooth initial data, where R_h and P_h are Ritz and the orthogonal $L_2(\Omega)$-projection on the finite element space X_h, respectively.

We shall study the convergence of the semidiscrete Galerkin FEM (1.3) for the case of initial data $v \in \dot{H}^q(\Omega)$, $q = 0, 1, 2$ (for the definition of these spaces, see Section 2.2). The case $q = 2$ is referred to as smooth initial data, while the case $q = 0$ is known as nonsmooth initial data.

In the past, the initial value problem for a standard parabolic equation, i.e. $\alpha = 1$, has been thoroughly studied in all these cases. It is well known that, for smooth initial data, the solution u_h satisfies an error bound uniformly in $t \geq 0$ [25, Theorem 3.1]:

(1.4) $\|u_h(t) - u(t)\| + h\|\nabla(u_h(t) - u(t))\| \leq Ch^2 \|v\|_2, \quad \text{for} \quad t \geq 0.$

We also have a nonsmooth data error estimate, for v assumed to be only in $L_2(\Omega)$, but which deteriorates for t approaching 0 [25, Theorem 3.2], namely,

$$(1.5) \quad \|u_h(t) - u(t)\| + h\|\nabla(u_h(t) - u(t))\| \le Ch^2t^{-1}\|v\|, \quad \text{for} \quad t > 0.$$

The proof of all these results exploits the smoothing property of the parabolic problem via its representation through the solution operator $E(t) = e^{t\Delta}$, namely,

$$u(t) = E(t)v + \int_0^t E(t-s)f(s)\,ds, \ t > 0.$$

In this paper we establish analogous results for the semidiscrete Galerkin FEM (1.3) for the model problem (1.1). The main difficulty in the error analysis stems from limited smoothing properties of the FPDE, cf. Theorem 2.1. Note that the solution operator for the FPDE is defined through the Mittag-Leffler function, which decays only linearly at infinity, cf. Lemma 2.1, in contrast with the standard parabolic equation whose solution decays exponentially for $t \to \infty$. The difficulty is overcome by exploiting the mapping property of the discrete solution operators.

Our main results are as follows. Firstly, in case of smooth initial data, we derived the same error bound (1.4) uniformly in $t \ge 0$ (cf. Theorem 3.1), as is in the case of the standard parabolic problem. Secondly, for quasi-uniform meshes we derived a nonsmooth data error estimate, for $v \in L_2(\Omega)$ only, which deteriorates for t approaching 0 (cf. Theorem 3.2)

$$(1.6) \ \ \|u_h(t) - u(t)\| + h\|\nabla(u_h(t) - u(t))\| \le Ch^2\ell_h t^{-\alpha}\|v\|, \ \ell_h = |\ln h|, \ t > 0.$$

This result is similar to the counterpart of standard parabolic problem but derived for quasi-uniform meshes and with an additional log-factor, ℓ_h.

Further, we study the more practical lumped mass semidiscrete Galerkin FEM. We have shown the same rate of convergence for the case of smooth initial data (cf. Theorem 4.1), and also the almost optimal error estimate for the gradient in the case of data $v \in \dot{H}^1(\Omega)$ and $v \in L_2(\Omega)$ (see estimate (4.2)). For the case of nonsmooth data, $v \in L_2(\Omega)$, for general quasi-uniform meshes, we were only able to establish a suboptimal L_2-error bound of order $O(h\ell_h t^{-\alpha})$, see (4.15). Further, inspired by the study in [2], we also consider special meshes. Namely, for a class of special triangulations satisfying the condition (4.16), which holds for meshes that are symmetric with respect to each internal vertex [2, Section 5], we show an almost optimal convergence estimate (4.17):

$$\|\bar{u}_h(t) - u(t)\| \le Ch^2\ell_h t^{-\alpha}\|v\|,$$

where $\bar{u}_h(t)$ is the solution of the lumped mass FEM. This estimate is similar to the one derived for the lumped mass semidiscrete Galerkin method for the standard parabolic equation [2, Theorem 4.1].

Finally, in Theorem 5.1 we establish a superconvergence result for the postprocessed gradient of the error in case of smooth initial data and a planar domain for special meshes. This improves the convergence order in

H^1-norm from $O(h)$ to $O(h^2\ell_h t^{-\frac{\alpha}{2}})$ for both Galerkin and the lumped mass finite element approximation.

The paper is organized as follows. In Section 2, we state basic properties of the Mittag-Leffler function, the smoothing property of the equation (1.1), and some basic estimates for finite element projection operators. In Sections 3 and 4, we derive error estimates for the standard Galerkin FEM and lumped mass FEM, respectively. In Section 5 we give a superconvergence result for the gradient of the error in case of smooth initial data. Finally, in Section 6 we present some numerical tests for a number of one-dimensional examples, including both smooth and non-smooth data. The numerical tests confirm our theoretical study.

We assume that the mesh size h of the triangulation $\mathcal{T}_h$ satisfies $0 < h < 1$. Throughout we shall denote by C a generic constant, which may differ at different occurrences, but is always independent of the mesh size h, the solution u and the initial data v.

2. Preliminaries

In this section, we collect useful facts on the Mittag-Leffler function, regularity results for the fractional diffusion equation (1.1), and basic estimates for the finite-element projection operators.

2.1. **Mittag-Leffler function.** We shall use extensively the Mittag-Leffler function $E_{\alpha,\beta}(z)$ defined below

$$E_{\alpha,\beta}(z) = \sum_{k=0}^{\infty} \frac{z^k}{\Gamma(k\alpha + \beta)} \quad z \in \mathbb{C},$$

where $\Gamma(\cdot)$ is the standard Gamma function defined as

$$\Gamma(z) = \int_0^{\infty} t^{z-1} e^{-t} dt \ \Re(z) > 0.$$

The Mittag-Leffler function is a two-parameter family of entire functions of z of order α^{-1} and type 1 [12, pp. 42]. The exponential function is a particular case of the Mittag-Leffler function, namely $E_{1,1}(z) = e^z$, [12, pp. 42]. Two most important members of this family are $E_{\alpha,1}(-\lambda t^{\alpha})$ and $t^{\alpha-1} E_{\alpha,\alpha}(-\lambda t^{\alpha})$, which occur in the solution operators for the initial value problem and the nonhomogeneous problem (1.1), respectively. There are several important properties of the Mittag-Leffler function $E_{\alpha,\beta}(z)$, mostly derived by M. Djrbashian (cf. [7, Chapter 1]).

Lemma 2.1. *Let* $0 < \alpha < 2$ *and* $\beta \in \mathbb{R}$ *be arbitrary, and* $\frac{\alpha\pi}{2} < \mu < \min(\pi, \alpha\pi)$. *Then there exists a constant* $C = C(\alpha, \beta, \mu) > 0$ *such that*

(2.1) $$|E_{\alpha,\beta}(z)| \leq \frac{C}{1 + |z|} \qquad \mu \leq |\arg(z)| \leq \pi.$$

Moreover, for $\lambda > 0$, $\alpha > 0$, *and* $t > 0$ *we have*

(2.2) $$\partial_t^\alpha E_{\alpha,1}(-\lambda t^\alpha) = -\lambda E_{\alpha,1}(-\lambda t^\alpha).$$

Proof. The estimate (2.1) can be found in [12, pp. 43, equation (1.8.28)] or [22, Theorem 1.4], while (2.2) is discussed in [12, Lemma 2.33, equation (2.4.58)]. □

2.2. **Solution representation.**
To discuss the regularity of the solution of (1.1), we shall need some notation. For $q \geq 0$, we denote by $\dot{H}^q(\Omega) \subset L_2(\Omega)$ the Hilbert space induced by the norm

$$|v|_q^2 = \sum_{j=1}^{\infty} \lambda_j^q (v, \varphi_j)^2,$$

with $(\cdot, \cdot)$ denoting the inner product in $L_2(\Omega)$ and $\{\lambda_j\}_{j=1}^{\infty}$ and $\{\varphi_j\}_{j=1}^{\infty}$ being respectively the eigenvalues and eigenfunctions of $-\Delta$ with homogeneous Dirichlet boundary data on $\partial\Omega$. The set $\{\varphi_j\}_{j=1}^{\infty}$ forms an orthonormal basis in $L_2(\Omega)$. Thus $|v|_0 = \|v\| = (v, v)^{1/2}$ is the norm in $L_2(\Omega)$, $|v|_1$ the norm in $H_0^1 = H_0^1(\Omega)$ and $|v|_2 = \|\Delta v\|$ is equivalent to the norm in $H^2(\Omega)$ when $v = 0$ on $\partial\Omega$ [25]. We set $\dot{H}^{-q} = (\dot{H}^q)'$, the set of all bounded linear functionals on the space $\dot{H}^q$.

Now we give a representation of the solution of problem (1.1) using the Dirichlet eigenpairs $\{(\lambda_j, \varphi_j)\}$. First, we introduce the operator $E(t)$:

$$(2.3) \qquad E(t)v = \sum_{j=1}^{\infty} E_{\alpha,1}(-\lambda_j t^\alpha) (v, \varphi_j) \varphi_j(x).$$

This is the solution operator to problem (1.1) with a homogeneous right hand side, so that for $f(x, t) \equiv 0$ we have $u(t) = E(t)v$. This representation follows from an eigenfunction expansion and (2.2) [23]. Further, for the non-homogeneous equation with a homogeneous initial data $v \equiv 0$, we shall use the operator defined for $\chi \in L^2(\Omega)$ as

$$(2.4) \qquad \bar{E}(t)\chi = \sum_{j=0}^{\infty} t^{\alpha-1} E_{\alpha,\alpha}(-\lambda_j t^\alpha) (\chi, \varphi_j) \varphi_j(x).$$

The operators $E(t)$ and $\bar{E}(t)$ are used to represent the solution $u(x, t)$ of (1.1):

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

It was shown in [23, Theorem 2.2] that if $f(x, t) \in L_2((0, T); L_2(\Omega))$, then there is a unique solution $u(x, t) \in L_2((0, T); \dot{H}^2(\Omega))$.

For the solution of the homogeneous equation (1.1), which is the object of our study, we have the following stability and smoothing estimates, essentially established in [23, Theorem 2.1], and slightly extended in the theorem below. Since these estimates will play a key role in the error analysis of the FEM approximations, we give some simple hints of the proof.

Theorem 2.1. *The solution $u(t) = E(t)v$ to problem (1.1) with $f \equiv 0$ satisfies the following estimates*

$$(2.5) \qquad |(\partial_t^\alpha)^\ell u(t)|_p \le Ct^{-\alpha(\ell + \frac{p-q}{2})}|v|_q, \quad t > 0,$$

where for $\ell = 0$, $0 \le q \le p \le 2$ and for $\ell = 1$, $0 \le p \le q \le 2$ and $q \le p + 2$.

Proof. First we discuss the case $\ell = 0$. According to parts (i) and (iii) of [23, Theorem 2.1], we have

$$(2.6) \qquad |u(t)|_2 + \|\partial_t^\alpha u(t)\| \le Ct^{-\alpha(1-\frac{q}{2})}|v|_q, \quad q = 0, 2.$$

By means of interpolation of estimates (2.6) for $q = 0$ and $q = 2$, we get the desired estimate (2.5) for the the case $p = 2$, $0 \le q \le 2$.

Further, applying part (i) of [23, Theorem 2.1], we have

$$(2.7) \qquad \|u(t)\| \le C\|v\|.$$

Thus, interpolation of (2.6) for $q = 2$ and (2.7) yields (2.5) for $0 \le p = q \le 2$. The remaining cases, $0 \le q < p < 2$, follow from the interpolation of (2.6) with $q = 0$ and (2.7). This shows the assertion for $\ell = 0$.

Now we consider the case $\ell = 1$. It follows from the representation formula (2.3) and Lemma 2.1 that

$$|\partial_t^\alpha u(t)|_p^2 = \sum_{j=1}^\infty \lambda_j^{2+p} E_{\alpha,1}(-\lambda_j t^\alpha)^2 (v, \varphi_j)^2$$

$$= t^{-\alpha(2+p-q)} \sum_{j=1}^\infty (\lambda_j t^\alpha)^{2+p-q} E_{\alpha,1}(-\lambda_j t^\alpha)^2 \lambda_j^q (v, \varphi_j)^2$$

$$\le Ct^{-\alpha(2+p-q)} \sum_{j=1}^\infty \frac{(\lambda_j t^\alpha)^{2+p-q}}{(1 + \lambda_j t^\alpha)^2} \lambda_j^q (v, \varphi_j)^2$$

$$\le Ct^{-\alpha(2+p-q)} \sup_{j \in \mathbb{N}} \frac{(\lambda_j t^\alpha)^{2+p-q}}{(1 + \lambda_j t^\alpha)^2} \sum_{j=1}^\infty \lambda_j^q (v, \varphi_j)^2 \le Ct^{-\alpha(2+p-q)}|v|_q^2,$$

where we have used the fact that, in view of Young's inequality,

$$\sup_{j \in \mathbb{N}} \frac{(\lambda_j t^\alpha)^{2+p-q}}{(1 + \lambda_j t^\alpha)^2} \le C \quad \text{for} \quad p \le q \le p + 2.$$

Thus, we get

$$|\partial_t^\alpha u(t)|_p \le Ct^{-\alpha(1 + \frac{p-q}{2})}|v|_q.$$

This completes the proof of the theorem. $\qquad\qquad\qquad\qquad\qquad$ □

Remark 2.1. *Note that for $\ell = 1$ we have the restriction $p \le q$, which is not present in the similar result for the standard parabolic problem, see, e.g. [25, Lemma 3.2]. This reflects the fact that FPDE has limited smoothing properties. The limited smoothing is also valid for the semidiscrete Galerkin approximation (see Lemma 3.1), which will influence the error estimates for the finite element solution.*

2.3. **Properties of Ritz and L_2-projections on X_h.** In our analysis we shall also use the orthogonal L_2-projection $P_h : L_2(\Omega) \to X_h$ and the Ritz projection $R_h : H_0^1(\Omega) \to X_h$ defined by

$$
\begin{aligned}
(P_h\psi, \chi) &= (\psi, \chi) \quad \forall \chi \in X_h, \\
(\nabla R_h\psi, \nabla\chi) &= (\nabla\psi, \nabla\chi) \quad \forall \chi \in X_h,
\end{aligned}
\tag{2.8}
$$

respectively. It is well-known that the operators P_h and R_h have the following approximation properties.

Lemma 2.2. *The operators P_h and R_h satisfy*

$$(2.9) \quad \|P_h\psi - \psi\| + h\|\nabla(P_h\psi - \psi)\| \le Ch^q|\psi|_q, \quad for \quad \psi \in \dot{H}^q, \ q = 1, 2.$$

$$(2.10) \quad \|R_h\psi - \psi\| + h\|\nabla(R_h\psi - \psi)\| \le Ch^q|\psi|_q, \quad for \quad \psi \in \dot{H}^q, \ q = 1, 2.$$

In particular, (2.9) indicates that P_h is stable in $\dot{H}^1$.

Proof. The estimates (2.10) are well known, cf. e.g. [25, Lemma 1.1] or [8, Theorem 3.16 and Theorem 3.18]. For globally uniform meshes, the case considered in this paper, the $\dot{H}^1$ stability of P_h directly follows from the error bound (2.9) and the inverse inequality. However, for more general meshes such stability is valid only under some mild assumptions on the mesh; see, e.g. [4]. $\qquad\square$

3. SEMIDISCRETE GALERKIN FEM

In this section we derive error estimates for the standard semidiscrete Galerkin FEM. First we recall some basic known facts for the spatially semidiscrete standard Galerkin FEM. We begin with the smoothing properties of the solution operators for the semidiscrete method as well as other preliminary results needed in the sequel. The error estimates hinge crucially on the smoothing properties of the discrete operator $\bar{E}_h$, cf. (3.3).

3.1. **Semidiscrete Galerkin FEM and its properties.** Upon introducing the discrete Laplacian $\Delta_h : X_h \to X_h$ defined by

$$-(\Delta_h\psi, \chi) = (\nabla\psi, \nabla\chi) \quad \forall \psi, \chi \in X_h,$$

and $f_h = P_h f$ we may write the spatially discrete problem (1.3) as

$$(3.1) \qquad \partial_t^\alpha u_h(t) - \Delta_h u_h(t) = f_h(t) \quad \text{for} \quad t \ge 0 \quad \text{with} \quad u_h(0) = v_h.$$

Now we give a representation of the solution of (3.1) using the eigenvalues and eigenfunctions $\{\lambda_j^h\}_{j=1}^N$ and $\{\varphi_j^h\}_{j=1}^N$ of the discrete Laplacian $-\Delta_h$. First we introduce the discrete analogues of (2.3) and (2.4) for $t > 0$:

$$(3.2) \qquad E_h(t)v_h = \sum_{j=1}^N E_{\alpha,1}(-\lambda_j^h t^\alpha)(v_h, \varphi_j^h)\varphi_j^h$$

and

$$\text{(3.3)} \qquad \bar{E}_h(t)f_h = \sum_{j=1}^{N} t^{\alpha-1} E_{\alpha,\alpha}(-\lambda_j^h t^\alpha)\,(f_h, \varphi_j^h)\,\varphi_j^h.$$

Then the solution $u_h(x,t)$ of the discrete problem (3.1) can be expressed by:

$$\text{(3.4)} \qquad u_h(x,t) = E_h(t)v_h + \int_0^t \bar{E}_h(t-s)f_h(s)\,ds.$$

Also, on the finite element space X_h, we introduce the discrete norm $\||\cdot\||_p$ for any $p \in \mathbb{R}$ defined by

$$\text{(3.5)} \qquad \||\psi\||_p^2 = \sum_{j=1}^{N} (\lambda_j^h)^p (\psi, \varphi_j^h)^2 \quad \psi \in X_h.$$

Since we are dealing with finite dimensional spaces, the above norm is well defined for all real p. From the very definition of the discrete Laplacian $-\Delta_h$ we have $\||\psi\||_1 = |\psi|_1$ and also $\||\psi\||_0 = \|\psi\|$ for any $\psi \in X_h$. So there will be no confusion in using $|\psi|_p$ instead of $\||\psi\||_p$ for $p = 0, 1$ and $\psi \in X_h$.

To analyze the convergence of the semidiscrete Galerkin method, we shall need various smoothing properties of the operator $E_h(t)$, which are discrete analogues of those formulated in (2.5). The estimates will be used for analyzing the convergence of the lumped mass FEM in Section 4.

Lemma 3.1. *Let $E_h(t)$ be defined by (3.2) and $v_h \in X_h$. Then*

$$\text{(3.6)} \quad \||(\partial_t^\alpha)^\ell u_h(t)\||_p = \||(\partial_t^\alpha)^\ell E_h(t)v_h\||_p \leq Ct^{-\alpha(\ell + \frac{p-q}{2})}\||v_h\||_q, \quad t > 0,$$

where for $\ell = 0$, $q \leq p$ and $0 \leq p - q \leq 2$ and for $\ell = 1$, $p \leq q \leq p + 2$.

Proof. First, consider the case $\ell = 0$. Then using the representation (3.2) of the solution $u_h(t)$ and the bound for the Mittag-Leffler function $E_{\alpha,\beta}(z)$ in Lemma 2.1 we get for $q \leq p$

$$\||u_h(t)\||_p^2 = \sum_{j=1}^{N} (\lambda_j^h)^p |(u_h(t), \varphi_j^h)|^2 = \sum_{j=1}^{N} (\lambda_j^h)^p |E_{\alpha,1}(-\lambda_j^h t^\alpha)|^2 |(v_h, \varphi_j^h)|^2$$

$$\leq Ct^{-\alpha(p-q)} \sum_{j=1}^{N} \frac{(\lambda_j^h t^\alpha)^{p-q}}{(1 + \lambda_j^h t^\alpha)^2} (\lambda_j^h)^q (v_h, \varphi_j^h)^2$$

$$\leq Ct^{-\alpha(p-q)} \sum_{j=1}^{N} (\lambda_j^h)^q |(v_h, \varphi_j^h)|^2 = Ct^{-\alpha(p-q)} \||v_h\||_q^2.$$

Here in the last inequality we have used the fact that for $q \leq p$ and $p \leq q \leq p + 2$ the expression $\max_j (\lambda_j^h t^\alpha)^{p-q}/(1 + \lambda_j^h t^\alpha)^2$ is bounded.

The estimates for the case $\ell = 1$ are obtained analogously using the representation (3.2) of the solution $u_h(t)$ for $p \le q \le p+2$ and Lemma 2.1:

$$
\begin{aligned}
\||\partial_t^\alpha u_h(t)\||_p^2 &= \sum_{j=1}^N (\lambda_j^h)^p |(\partial_t^\alpha u_h(t), \varphi_j^h)|^2 \\
&= \sum_{j=1}^N (\lambda_j^h)^{2+p} |E_{\alpha,1}(-\lambda_j^h t^\alpha)|^2 |(v_h, \varphi_j^h)|^2.
\end{aligned}
$$

Now using the bound of Mittag-Leffler function in Lemma 2.1 and Young's inequality, we obtain

$$
\begin{aligned}
\||\partial_t^\alpha u_h(t)\||_p^2 &\le C t^{-(2\alpha+\alpha(p-q))} \sum_{j=1}^N \frac{(\lambda_j^h t^\alpha)^{2+p-q}}{(1+\lambda_j^h t^\alpha)^2} (\lambda_j^h)^q |(v_h, \varphi_j^h)|^2 \\
&\le C t^{-(2\alpha+\alpha(p-q))} \sum_{j=1}^N (\lambda_j^h)^q |(v_h, \varphi_j^h)|^2 \\
&= C t^{-(2\alpha+\alpha(p-q))} \||v_h\||_q^2.
\end{aligned}
$$

The desired estimate follows from this immediately. $\qquad\square$

The following estimates are crucial for the a priori error analysis in the sequel.

Lemma 3.2. *Let $\bar{E}_h$ be defined by (3.3) and $\psi \in X_h$. Then we have for all $t > 0$,*

$$
(3.7) \qquad \||\bar{E}_h(t)\psi\||_p \le \begin{cases} C t^{-1+\alpha(1+\frac{q-p}{2})} \||\psi\||_q, & p-2 \le q \le p, \\ C t^{-1+\alpha} \||\psi\||_q, & p < q. \end{cases}
$$

Proof. By the definition of the operator $\bar{E}_h(t)$ and using Lemma 2.1 for $E_{\alpha,\alpha}(z)$, we have for any $p \in \mathbb{R}$ and $q \le p$

$$
\begin{aligned}
\||\bar{E}_h(t)\psi\||_p^2 &= t^{-2+2\alpha} \sum_{j=1}^N E_{\alpha,\alpha}^2(-\lambda_j^h t^\alpha)(\lambda_j^h)^p (\psi, \varphi_j^h)^2 \\
&\le C t^{-2+\alpha(2+q-p)} \max_j \frac{(\lambda_j^h t^\alpha)^{p-q}}{(1+\lambda_j^h t^\alpha)^2} \sum_{j=1}^N (\lambda_j^h)^q (\psi, \varphi_j^h)^2 \\
&= C t^{-2+\alpha(2+q-p)} \||\psi\||_q^2,
\end{aligned}
$$

where for getting the last inequality we took into account $0 \le p-q \le 2$. The desired assertion for $p < q$ follows from the fact that the eigenvalues $\{\lambda_j^h\}$ are bounded away from zero independently of the mesh size h. $\quad\square$

Remark 3.1. *Lemma 3.2 expresses the smoothing properties of the operator $\bar{E}_h$. While $p = 0, 1$, the parameter q can be arbitrary as long as it complies with the condition $p-2 \le q \le p$. This flexibility in the choice of q is essential for deriving error estimates for problems with initial data of low regularity.*

Further, we shall need the following inverse inequality.

Lemma 3.3. *There exists a constant C independent of h such that for all $\psi \in X_h$ we have for any real $l > s$*

$$\||\psi\||_l \leq Ch^{s-l}\||\psi\||_s. \tag{3.8}$$

Proof. For quasi-uniform triangulations $\mathcal{T}_h$ the inverse inequality $|\psi|_1 \leq Ch^{-1}\|\psi\|$ holds for all $\psi \in X_h$. By the definition of $-\Delta_h$ this implies $\max_{1\leq j \leq N} \lambda_j^h \leq C/h^2$. Thus, for the norm $\||\cdot\||_p$ defined in (3.5), we obtain that for any real $l > s$

$$\||\psi\||_l^2 \leq C \max_j (\lambda_j^h)^{l-s} \sum_{j=1}^N (\lambda_j^h)^s (\psi, \varphi_j^h)^2 \leq Ch^{2(s-l)}\||\psi\||_s^2.$$

That completes the proof. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

3.2. Error estimates for smooth initial data.

Here we establish error estimates for the semidiscrete Galerkin method for initial data $v \in \dot{H}^2(\Omega)$. In a customary way we split the error $u_h(t) - u(t)$ into two terms as

$$u_h - u = (u_h - R_h u) + (R_h u - u) := \vartheta + \varrho. \tag{3.9}$$

By (2.10) and (2.5) we have for any $t > 0$ and $q = 1, 2$,

$$\|\varrho(t)\| + h\|\nabla\varrho(t)\| \leq Ch^2 t^{-\alpha(1-\frac{q}{2})}|v|_q \quad v \in \dot{H}^q, \tag{3.10}$$

so it suffices to get proper estimates for $\vartheta(t)$, which is done in the following lemma.

Lemma 3.4. *Let u and u_h be the solutions of (1.1) and (1.3), respectively, with $v_h = R_h v$. Then for $\vartheta(t) = u_h(t) - R_h u(t)$ we have*

$$\|\vartheta(t)\| + h\|\nabla\vartheta(t)\| \leq Ch^2|v|_2. \tag{3.11}$$

Proof. We note that ϑ satisfies

$$\partial_t^\alpha \vartheta(t) - \Delta_h \vartheta(t) = -P_h \partial_t^\alpha \varrho(t) \quad \text{for} \quad t > 0. \tag{3.12}$$

For $v \in \dot{H}^q$, $q = 1, 2$ the Ritz projection $R_h v$ is well defined, so that $\vartheta(0) = 0$ and hence, by Duhamel's principle (3.4),

$$\vartheta(t) = -\int_0^t \bar{E}_h(t-s)P_h\partial_t^\alpha \varrho(s)\, ds. \tag{3.13}$$

By using Lemma 3.2 with $p = 1$ and $q = 0$, the stability of P_h, (2.10), and the estimate (2.5) with $\ell = 1$, $p = 1$ and we find, for $q = 1, 2$,

$$\begin{aligned}
\|\nabla\bar{E}_h(t-s)P_h\partial_t^\alpha \varrho(s)\| &\leq C(t-s)^{\frac{\alpha}{2}-1}\|\partial_t^\alpha \varrho(s)\| \\
&\leq Ch(t-s)^{\frac{\alpha}{2}-1}|\partial_t^\alpha u(s)|_1 \\
&\leq Ch(t-s)^{\frac{\alpha}{2}-1}s^{\alpha(-\frac{3}{2}+\frac{q}{2})}|v|_q.
\end{aligned} \tag{3.14}$$

By substituting this inequality into (3.13) we obtain that for $q = 1, 2$

$$(3.15) \quad \|\nabla\vartheta(t)\| \le Ch \int_0^t (t-s)^{\frac{\alpha}{2}-1} s^{\alpha(-\frac{3}{2}+\frac{q}{2})}\,ds\,|v|_q \le Cht^{-\alpha(1-\frac{q}{2})}|v|_q,$$

where we have used that for $\alpha < 1$

$$\int_0^t (t-s)^{\frac{\alpha}{2}-1} s^{\alpha(-\frac{3}{2}+\frac{q}{2})}\,ds = t^{\frac{\alpha}{2}-\frac{3\alpha}{2}+\frac{q\alpha}{2}} \int_0^1 (1-s)^{\frac{\alpha}{2}-1} s^{\alpha(-\frac{3}{2}+\frac{q}{2})}\,ds$$

$$= B(\tfrac{\alpha}{2}, \alpha(-\tfrac{3}{2}+\tfrac{q}{2})+1)t^{-\alpha(1-\frac{q}{2})},$$

with $B(\cdot,\cdot)$ being the standard Beta function. Since both arguments, $\frac{\alpha}{2} > 0$ and $\frac{-3+q}{2}\alpha + 1 > 0$ for $q = 1, 2$, the value $B(\alpha, \alpha(-\frac{3}{2}+\frac{q}{2})+1)$ is finite. Taking $q = 2$ we get the desired estimate for $\nabla\vartheta$.

Next, by using the smoothing property of the operator $\bar{E}_h$ in Lemma 3.2 with $p = q = 0$ and that of the operator E in Theorem 2.1 with $\ell = 1$ and $p = q = 2$, we get

$$
\begin{aligned}
\|\vartheta(t)\| &\le \int_0^t \|\bar{E}_h(t-s)P_h\partial_t^\alpha \varrho(s)\|ds \\
&\le C\int_0^t (t-s)^{\alpha-1}\|\partial_t^\alpha \varrho(s)\|ds \\
&\le Ch^2 \int_0^t (t-s)^{\alpha-1}|\partial_t^\alpha u(s)|_2 ds \\
&\le Ch^2 \int_0^t (t-s)^{\alpha-1}s^{-\alpha}ds|v|_2 = CB(\alpha, 1-\alpha)h^2|v|_2.
\end{aligned}
$$
(3.16)

This completes the proof. $\square$

Using the triangle inequality and the estimates (3.10) and (3.11) we get the main result in the subsection.

Theorem 3.1. *Let u and u_h be the solutions of (1.1) and (1.3), respectively, with $v_h = R_h v$. Then*

$$(3.17) \qquad \|u_h(t) - u(t)\| + h\|\nabla(u_h(t) - u(t))\| \le Ch^2|v|_2.$$

Remark 3.2. *As a byproduct of estimates (2.10) and (3.15) we also got a bound for the error for $v \in \dot{H}^1(\Omega)$ and $v_h = R_h v$:*

$$(3.18) \qquad \|\nabla(u_h(t) - u(t))\| \le Cht^{-\frac{\alpha}{2}}|v|_1.$$

Remark 3.3. *In view of the smoothing property of the operator $\bar{E}_h$ established in Lemma 3.2, we can improve the estimate of $\vartheta(t)$ for $q = 2$ to $O(h^2)$ at the expense of slightly increasing the factor by $O(t^{-\frac{\alpha}{2}})$:*

$$
\begin{aligned}
\|\nabla\bar{E}_h(t-s)P_h\partial_t^\alpha \varrho(s)\| &\le Ch^2(t-s)^{\frac{\alpha}{2}-1}|\partial_t^\alpha u(s)|_2 \\
&\le Ch^2(t-s)^{\frac{\alpha}{2}-1}s^{-\alpha}|v|_2,
\end{aligned}
$$

which yields

$$(3.19) \qquad \|\nabla\vartheta\| \le Ch^2 t^{-\frac{\alpha}{2}}|v|_2.$$

3.3. **Error estimates for non-smooth initial data.** Now we prove an error estimate for nonsmooth initial data, $v \in L_2(\Omega)$, and the intermediate case, $v \in \dot{H}^1(\Omega)$. Since the Ritz projection $R_h v$ is not defined for $v \in L_2(\Omega)$, we shall use instead the L_2-projection P_h onto the finite element space X_h, and split the error $u_h - u$ into:

$$u_h - u = (u_h - P_h u) + (P_h u - u) := \widetilde{\vartheta} + \widetilde{\varrho}.$$

By Lemma 2.2 and Theorem 2.1 we have

$$(3.20) \quad \|\widetilde{\varrho}(t)\| + h\|\nabla\widetilde{\varrho}(t)\| \leq Ch^2|u(t)|_2 \leq Ch^2 t^{-\alpha(1-\frac{q}{2})}\|v\|_q, \quad q = 0, 1.$$

Thus, we only need to estimate the term $\widetilde{\vartheta}$. Obviously, $P_h \partial_t^\alpha \widetilde{\varrho} = \partial_t^\alpha P_h(P_h u - u) = 0$ and we get the following problem for $\widetilde{\vartheta}$:

$$(3.21) \quad \partial_t^\alpha \widetilde{\vartheta}(t) - \Delta_h \widetilde{\vartheta}(t) = -\Delta_h(R_h u - P_h u)(t), \quad t > 0, \quad \widetilde{\vartheta}(0) = 0.$$

Then with the help of formula (3.3), $\widetilde{\vartheta}(t)$ can be represented by

$$(3.22) \qquad \widetilde{\vartheta}(t) = -\int_0^t \bar{E}_h(t - s)\Delta_h(R_h u - P_h u)(s)\, ds.$$

Next, we show the following estimate for $\widetilde{\vartheta}(t)$:

Lemma 3.5. *Let* $\widetilde{\vartheta}(t)$ *be defined by* (3.22). *Then for* $p = 0, 1$, $q = 0, 1$, *and* $\ell_h = |\ln h|$, *the following estimate holds*

$$\|\widetilde{\vartheta}(t)\|_p \leq Ch^{2-p}\ell_h t^{-\alpha(1-\frac{q}{2})}\|v\|_q.$$

Proof. By Lemma 3.2 with $p = 0, 1$ and $q = p - 2 + \epsilon$, for any $\epsilon > 0$ we have

$$\|\widetilde{\vartheta}(t)\|_p \leq \int_0^t \|\bar{E}_h(t-s)\Delta_h(R_h u - P_h u)(s)\|_p ds$$

$$\leq \int_0^t (t-s)^{\frac{\epsilon}{2}\alpha - 1}\||\Delta_h(R_h u - P_h u)\||_{p-2+\epsilon} ds$$

$$\leq \int_0^t (t-s)^{\frac{\epsilon}{2}\alpha - 1}\||R_h u - P_h u\||_{p+\epsilon} ds := A.$$

Further, we apply the inverse inequality (3.8) for $R_h u - P_h u$, the bounds (2.9) and (2.10), for $P_h u - u$ and $R_h u - u$, respectively, and the smoothing

property (2.5) with $\ell = 0$ and $p = 2$ to get

$$A \leq Ch^{-\epsilon} \int_0^t (t-s)^{\frac{\epsilon}{2}\alpha-1} \|R_h u - P_h u\|_p ds$$

$$\leq Ch^{2-p-\epsilon} \int_0^t (t-s)^{\frac{\epsilon}{2}\alpha-1} \|u(s)\|_2 ds$$

$$\leq Ch^{2-p-\epsilon} \int_0^t (t-s)^{\frac{\epsilon}{2}\alpha-1} s^{-\alpha(1-\frac{q}{2})} ds \, \|v\|_q$$

$$= CB\left(\frac{\epsilon}{2}\alpha, 1-\alpha+\frac{q}{2}\alpha\right) h^{2-p-\epsilon} t^{-\alpha(1-\frac{q}{2}-\frac{\epsilon}{2})} \|v\|_q$$

$$\leq \frac{C}{\epsilon} h^{2-p-\epsilon} t^{-\alpha(1-\frac{q}{2})} \|v\|_q.$$

The last inequality follows from the fact $B(\frac{\epsilon}{2}\alpha, 1-\alpha+\frac{q}{2}\alpha) = \frac{\Gamma(\frac{\epsilon}{2}\alpha)\Gamma(1-\alpha+\frac{q}{2}\alpha)}{\Gamma(1-\alpha+\frac{q+\epsilon}{2}\alpha)}$
and $\Gamma(\frac{\epsilon}{2}\alpha) \sim \frac{2}{\alpha\epsilon}$ as $\epsilon \to 0^+$, e.g., by means of Laurenz expansion of the Gamma function. The desired assertion follows by choosing $\epsilon = 1/\ell_h$. □

Then Lemma 3.5 and the triangle inequality yield the following almost optimal error estimate for the semidiscrete Galerkin method for initial data $v \in \dot{H}^q$, $q = 0, 1$:

Theorem 3.2. *Let u and u_h be the solutions of (1.1) and (1.3) with $v_h = P_h v$, respectively. Then with $\ell_h = |\ln h|$*

$$(3.23) \quad \|u_h(t) - u(t)\| + h\|\nabla(u_h(t) - u(t))\| \leq Ch^2 \ell_h \, t^{-\alpha(1-\frac{q}{2})} \|v\|_q, \ q = 0, 1.$$

Remark 3.4. *For $v \in \dot{H}^1(\Omega)$ and $v_h = R_h v$, we have established the estimate (3.18), which is slightly better than (3.23), since it does not have the factor ℓ_h.*

3.4. **Problems with more general elliptic operators.** The preceding analysis could be straightforwardly extended to problems with more general boundary conditions/spatially varying coefficients. In fact this is the strength of the finite element method and the advantages of the direct numerical methods for treating such problems in comparison with some analytical techniques that are limited to constant coefficients and canonical domains. More precisely, we can study problem (1.3) with a bilinear form $a(\cdot, \cdot) : V \times V \mapsto \mathbb{R}$ of the form:

$$(3.24) \quad a(u, \chi) = \int_\Omega (k(x)\nabla u \cdot \nabla\chi + c(x)u\chi) \, dx,$$

where $k(x)$ is a symmetric $d \times d$ matrix-valued measurable function on the domain Ω with smooth entries and $c(x)$ is an L_∞-function. We assume that

$$c_0|\xi|^2 \leq \xi^T k(x)\xi \leq c_1|\xi|^2, \quad \text{for} \quad \xi \in \mathbb{R}^d, \ x \in \Omega,$$

where $c_0, c_1 > 0$ are constants, and the bilinear form $a(\cdot, \cdot)$ is coercive on $V \equiv H^1(\Omega)$. Further, we assume that the problem $a(w, \chi) = (f, \chi), \forall\chi \in V$

has unique solution $w \in V$, which for $f \in L_2(\Omega)$ has full elliptic regularity, $\|w\|_{H^2} \leq C\|f\|_{L_2}$.

Under these conditions we can define a positive definite operator $\mathcal{A}$: $H_0^1 \to H^{-1}$, which has a complete set of eigenfunctions $\varphi_j(x)$ and respective eigenvalues $\lambda_j(\mathcal{A}) > 0$. Then we can define the spaces $\dot{H}^q$ as in Section 2.2. Further, we define the discrete operator $\mathcal{A}_h : X_h \to X_h$ by

$$(\mathcal{A}_h \psi, \chi) = a(\psi, \chi), \ \forall \psi, \chi \in V_h.$$

Then all results for problem (1.1) can be easily extended to fractional-order problems with elliptic equations of this more general form.

4. Lumped mass finite element method

Now we consider the lumped mass FEM in planar domains (see, e.g. [25, Chapter 15, pp. 239–244]). For completeness we shall introduce this approximation. Let z_j^τ, $j = 1, \ldots, d+1$ be the vertices of the d-simplex $\tau \in \mathcal{T}_h$. Consider the quadrature formula

$$(4.1) \qquad Q_{\tau,h}(f) = \frac{|\tau|}{d+1} \sum_{j=1}^{d+1} f(z_j^\tau) \approx \int_\tau f \, dx.$$

We may then define an approximation of the L_2-inner product in X_h by

$$(4.2) \qquad (w, \chi)_h = \sum_{\tau \in \mathcal{T}_h} Q_{\tau,h}(w\chi).$$

Then lumped mass Galerkin FEM is: find $\bar{u}_h(t) \in X_h$ such that

$$(4.3) \qquad \begin{aligned} (\partial_t^\alpha \bar{u}_h, \chi)_h + a(\bar{u}_h, \chi) &= (f, \chi) \quad \forall \chi \in X_h, \ t > 0, \\ \bar{u}_h(0) &= v_h. \end{aligned}$$

We now introduce the discrete Laplacian $-\bar{\Delta}_h : X_h \to X_h$, corresponding to the inner product $(\cdot, \cdot)_h$, by

$$(4.4) \qquad -(\bar{\Delta}_h \psi, \chi)_h = (\nabla \psi, \nabla \chi) \quad \forall \psi, \chi \in X_h.$$

Also, we introduce the L_2-projection, $\bar{P}_h : L_2(\Omega) \to X_h$ by

$$(\bar{P}_h f, \chi)_h = (f, \chi), \quad \forall \chi \in X_h.$$

The lumped mass method can then be written with $f_h = \bar{P}_h f$ in operator form as

$$\partial_t^\alpha \bar{u}_h(t) - \bar{\Delta}_h \bar{u}_h(t) = f_h(t) \quad \text{for } t \geq 0 \quad \text{with } \bar{u}_h(0) = v_h.$$

Similarly as in Section 3, we define the discrete operator F_h by

$$(4.5) \qquad F_h(t)v_h = \sum_{j=1}^N E_{\alpha,1}(-\bar{\lambda}_j^h t^\alpha)(v_h, \bar{\varphi}_j^h)_h \bar{\varphi}_j^h,$$

where $\{\bar{\lambda}_j^h\}_{j=1}^N$ and $\{\bar{\varphi}_j^h\}_{j=1}^N$ are respectively the eigenvalues and the orthonormal eigenfunctions of $-\bar{\Delta}_h$ with respect to $(\cdot, \cdot)_h$.

Analogously to (3.3), we introduce the operator $\bar{F}_h$ by

$$(4.6) \qquad \bar{F}_h f_h(t) = \sum_{j=1}^{N} t^{\alpha-1} E_{\alpha,\alpha}(-\bar{\lambda}_j^h t^\alpha)(f_h, \bar{\varphi}_j^h)_h \bar{\varphi}_j^h.$$

Then the solution $\bar{u}_h$ to problem (4.3) can be represented as follows

$$\bar{u}_h(t) = F_h(t)v_h + \int_0^t \bar{F}_h(t-s)f_h(s)ds.$$

For our analysis we shall need the following modification of the discrete norm (3.5), $||| \cdot |||_p$, on the space X_h

$$(4.7) \qquad |||\psi|||_p^2 = \sum_{j=1}^{N} (\bar{\lambda}_j^h)^p (\psi, \bar{\varphi}_j^h)_h^2 \quad \forall p \in \mathbb{R}.$$

The following norm equivalence result is useful.

Lemma 4.1. *The norm* $||| \cdot |||_p$ *defined in* (4.7) *is equivalent to the norm* $| \cdot |_p$ *on the space* X_h *for* $p = 0, 1$.

Proof. The equivalence the the two norms for the case of $p = 0$ is well known:

$$\frac{1}{2}|||\psi|||_0 \le \|\psi\| \le |||\psi|||_0, \quad \forall \psi \in X_h.$$

From the definitions of the discrete Laplacian $-\bar{\Delta}_h$ and the eigenpairs $\{(\bar{\lambda}_j^h, \bar{\varphi}_j^h)\}$, we deduce

$$\|\nabla\psi\|^2 = ((-\bar{\Delta}_h)\psi, \psi)_h = \sum_{j=1}^{N} \bar{\lambda}_j^h (\psi, \bar{\varphi}_j^h)_h^2 = |||\psi|||_1 \quad \forall \psi \in X_h.$$

This completes the proof of the lemma. $\square$

We shall also need the following inverse inequality, whose proof is identical with that of Lemma 3.3:

$$(4.8) \qquad |||\psi|||_l \le Ch^{s-l}|||\psi|||_s \quad l > s.$$

We show the following analogue of Lemma 3.2:

Lemma 4.2. *Let* $\bar{F}_h$ *be defined by* (4.6). *Then we have for* $\psi \in X_h$ *and all* $t > 0$,

$$|||\bar{F}_h(t)\psi|||_p \le \begin{cases} Ct^{-1+\alpha(1+\frac{q-p}{2})}|||\psi|||_q, & p-2 \le q \le p, \\ Ct^{-1+\alpha}|||\psi|||_q, & p < q. \end{cases}$$

Proof. The proof essentially follows the steps of the proof of Lemma 3.2 by replacing the eigenpairs $(\lambda_j^h, \varphi_j^h)$ by $(\bar{\lambda}_j^h, \bar{\varphi}_j^h)$, and the L_2-inner product $(\cdot, \cdot)$ by the approximate L_2-inner product $(\cdot, \cdot)_h$ and thus it is omitted. $\square$

We need the quadrature error operator $Q_h : X_h \to X_h$ defined by

(4.9) $(\nabla Q_h \chi, \nabla \psi) = \epsilon_h(\chi, \psi) := (\chi, \psi)_h - (\chi, \psi) \quad \forall \chi, \psi \in X_h.$

The operator Q_h, introduced in [2], represents the quadrature error (due to mass lumping) in a special way. It satisfies the following error estimate:

Lemma 4.3. Let $\bar{\Delta}_h$ and Q_h be the operators defined by (4.4) and (4.9), respectively. Then

$$\|\nabla Q_h \chi\| + h\|\bar{\Delta}_h Q_h \chi\| \le Ch^{p+1}\|\nabla^p \chi\| \quad \forall \chi \in X_h, \ p = 0, 1.$$

Proof. See [2, Lemma 2.4]. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ □

4.1. **Error estimate for smooth initial data.** We shall now establish error estimates for the lumped mass FEM for smooth initial data, i.e., $v \in \dot{H}^2(\Omega)$.

Theorem 4.1. Let u and $\bar{u}_h$ be the solutions of (1.1) and (4.3), respectively, with $v_h = R_h v$. Then

$$\|\bar{u}_h(t) - u(t)\| + h\|\nabla(\bar{u}_h(t) - u(t))\| \le Ch^2|v|_2.$$

Proof. Now we split the error into $\bar{u}_h(t) - u(t) = u_h(t) - u(t) + \delta(t)$ with $\delta(t) = \bar{u}_h(t) - u_h(t)$ and $u_h(t)$ being the solution by the standard Galerkin FEM. Since we have already established the estimate (3.17) for $u_h - u$, it suffices to get the following estimate for $\delta(t)$:

(4.10) $\|\delta(t)\| + h\|\nabla\delta(t)\| \le Ch^2|v|_2.$

It follows from the definitions of the $u_h(t)$, $\bar{u}_h(t)$, and Q_h that

$$\partial_t^\alpha \delta(t) - \bar{\Delta}_h \delta(t) = \bar{\Delta}_h Q_h \partial_t^\alpha u_h(t) \quad \text{for } t > 0, \quad \delta(0) = 0$$

and by Duhamel's principle we have

$$\delta(t) = \int_0^t \bar{F}_h(t-s)\bar{\Delta}_h Q_h \partial_t^\alpha u_h(s)ds.$$

Using Lemmas 4.1, 4.2, and 4.3 we get for $\chi \in X_h$:

$$\|\nabla \bar{F}_h(t)\bar{\Delta}_h Q_h \chi\| \le Ct^{\frac{\alpha}{2}-1}\|\bar{\Delta}_h Q_h \chi\| \quad \le Ct^{\frac{\alpha}{2}-1}h\|\nabla\chi\|.$$

Similarly, for $\chi \in X_h$

$$\|\bar{F}_h(t)\bar{\Delta}_h Q_h \chi\| \le Ct^{\frac{\alpha}{2}-1}\|\|\bar{\Delta}_h Q_h \chi\|\|_{-1} \le Ct^{\frac{\alpha}{2}-1}\|\nabla Q_h \chi\| \le Ct^{\frac{\alpha}{2}-1}h^2\|\nabla\chi\|.$$

Consequently, using Lemma 3.1 with $l = 1$, $p = 1$ and $q = 2$ we get

$$\|\delta(t)\| + h\|\nabla\delta(t)\| \le Ch^2\int_0^t (t-s)^{\frac{\alpha}{2}-1}\|\|\partial_t^\alpha u_h(s)\|\|_1 ds$$

$$\le Ch^2\int_0^t (t-s)^{\frac{\alpha}{2}-1}s^{-\frac{\alpha}{2}}ds\ \|\|u_h(0)\|\|_2.$$

Since $\Delta_h R_h = P_h \Delta$, we deduce

$$\|\|u_h(0)\|\|_2 = \|\Delta_h R_h u(0)\| = \|P_h \Delta u(0)\| \le |u(0)|_2 \le C\|v\|_2,$$

which yields (4.10) and concludes the proof. $\qquad\qquad\qquad\qquad\qquad\qquad$ □

An improved bound for $\|\nabla\delta(t)\|$ can be obtained as follows. In view of Lemmas 4.1 and 4.3 and (4.8), we observe that for any $\epsilon > 0$ and $\chi \in X_h$

$$\|\nabla\bar{F}_h(t)\bar{\Delta}_h Q_h\chi\| \leq Ct^{\frac{\epsilon}{2}\alpha-1}\||\bar{\Delta}_h Q_h\chi\||_{-1+\epsilon} \leq Ct^{\frac{\epsilon}{2}\alpha-1}h^{2-\epsilon}\|\nabla\chi\|.$$

Consequently,

$$(4.11) \qquad \|\nabla\delta(t)\| \leq Ch^{2-\epsilon}\int_0^t (t-s)^{\frac{\epsilon}{2}\alpha-1}\||\partial_t^\alpha u_h(s)\||_1 ds.$$

Now, to (4.11) we apply Lemma 3.1 with $l = 1$, $p = 1$ and $q = 2$ to get

$$\|\nabla\delta(t)\| \leq Ch^{2-\epsilon}\int_0^t (t-s)^{\frac{\epsilon}{2}\alpha-1}s^{-\frac{\alpha}{2}}ds\||u_h(0)\||_2 \leq C\frac{1}{\epsilon}h^{2-\epsilon}t^{-\alpha\frac{1-\epsilon}{2}}|v|_2.$$

Remark 4.1. *In the above estimate, by choosing $\epsilon = 1/\ell_h$, $\ell_h = |\ln h|$, we get*

$$(4.12) \qquad\qquad\qquad \|\nabla\delta(t)\| \leq Ch^2\ell_h t^{-\frac{\alpha}{2}}|v|_2,$$

which improves the bound of $\|\nabla\delta(t)\|$ for fixed $t > 0$ by almost one order.

Remark 4.2. *Instead, if we apply to (4.11) Lemma 3.1 with $l = 1$, $p = 1$ and $q = 1$ we shall get an improved estimate for $\delta(t)$ in the case of initial data $v \in \dot{H}^1$:*

$$(4.13) \qquad\qquad\qquad \|\nabla\delta(t)\| \leq Ch^2\ell_h t^{-\alpha}|v|_1.$$

4.2. **Error estimates for nonsmooth initial data.** Now we consider the case of nonsmooth initial data, i.e., $v \in L_2(\Omega)$ as well as the intermediate case $v \in \dot{H}^1$. Due to the lower regularity, we take $v_h = P_h v$. As before, the idea is to split the error into $\bar{u}_h(t) - u(t) = u_h(t) - u(t) + \delta(t)$ with $\delta(t) = \bar{u}_h(t) - u_h(t)$ and $u_h(t)$ being the solution of the standard Galerkin FEM. Thus, in view of estimate (3.23) it suffices to establish proper bound for $\delta(t)$.

Theorem 4.2. *Let u and $\bar{u}_h$ be the solutions of (1.1) and (4.3), respectively, with $v_h = P_h v$. Then with $\ell_h = |\ln h|$, the following estimates are valid for $t > 0$:*

$$(4.14) \qquad \|\nabla(\bar{u}_h(t) - u(t))\| \leq Ch\ell_h t^{-\alpha(1-\frac{q}{2})}|v|_q \quad q = 0, 1,$$

and

$$(4.15) \qquad \|\bar{u}_h(t) - u(t)\| \leq Ch^{q+1}\ell_h t^{-\alpha(1-\frac{q}{2})}|v|_q \quad q = 0, 1.$$

Furthermore, if the quadrature error operator Q_h defined by (4.9) satisfies

$$(4.16) \qquad\qquad \|Q_h\chi\| \leq Ch^2\|\chi\| \quad \forall\chi \in X_h,$$

then the following almost optimal error estimate is valid:

$$(4.17) \qquad\qquad\qquad \|\bar{u}_h(t) - u(t)\| \leq Ch^2\ell_h t^{-\alpha}\|v\|.$$

Proof. By Duhamel's principle

$$\delta(t) = \int_0^t \bar{F}_h(t-s)\bar{\Delta}_h Q_h \partial_t^\alpha u_h(s)ds.$$

Then by appealing to the smoothing property of the operator $\bar{F}_h$ in Lemma 4.2 and the inverse inequality (4.8), we get for $\chi \in X_h$, $\epsilon > 0$, and $p = 0, 1$

$$
\begin{aligned}
\||\bar{F}_h(t)\bar{\Delta}_h Q_h\chi\||_p &\leq Ct^{\frac{\epsilon}{2}\alpha-1}\||\bar{\Delta}_h Q_h\chi\||_{p-2+\epsilon}\\
&= Ct^{\frac{\epsilon}{2}\alpha-1}\||Q_h\chi\||_{p+\epsilon}\\
&\leq Ct^{\frac{\epsilon}{2}\alpha-1}h^{-\epsilon}\||Q_h\chi\||_p\\
&\leq Ct^{\frac{\epsilon}{2}\alpha-1}h^{-\epsilon}\|Q_h\chi\|_p.
\end{aligned}
$$
(4.18)

Consequently, by Lemmas 4.3, 3.1 and $\dot{H}^1$- and L_2-stability of the operator P_h from Lemma 2.2, we deduce for $q = 0, 1$

$$
\begin{aligned}
\|\nabla\delta(t)\| &\leq Ch^{q+1-\epsilon}\int_0^t (t-s)^{\frac{\epsilon}{2}\alpha-1}\|\partial_t^\alpha u_h(s)\|_q ds\\
&\leq Ch^{q+1-\epsilon}\int_0^t (t-s)^{\frac{\epsilon}{2}\alpha-1}s^{-\alpha}ds\|u_h(0)\|_q\\
&= Ch^{q+1-\epsilon}t^{-\alpha(1-\frac{\epsilon}{2})}B\left(\frac{\epsilon}{2}\alpha, 1-\alpha\right)\|P_h v\|_q\\
&\leq C\frac{1}{\epsilon}h^{q+1-\epsilon}t^{-\alpha(1-\frac{\epsilon}{2})}|v|_q.
\end{aligned}
$$

Now the desired estimate (4.14) follows by triangle inequality from the estimate (3.23) and the above estimate by taking $\epsilon = 1$ and $\epsilon = 1/\ell_h$ for the cases $q = 1$ and 0, respectively.

Next we derive an L_2- error estimate. First, note that for $\chi \in X_h$ we have

$$\|\bar{F}_h(t)\bar{\Delta}_h Q_h\chi\| \leq Ct^{\frac{\alpha}{2}-1}\||\bar{\Delta}_h Q_h\chi\||_{-1} \leq Ct^{\frac{\alpha}{2}-1}\|\nabla Q_h\chi\|.$$

This estimate and Lemma 4.3 give

$$
\begin{aligned}
\|\delta(t)\| &\leq Ch^{q+1}\int_0^t (t-s)^{\frac{\alpha}{2}-1}\|\partial_t^\alpha u_h(s)\|_q ds\\
&\leq Ch^{q+1}\int_0^t (t-s)^{\frac{\alpha}{2}-1}s^{-\alpha}ds\,|u_h(0)|_q\\
&\leq Ch^{q+1}t^{-\frac{\alpha}{2}}B\left(\frac{\alpha}{2}, 1-\alpha\right)|P_h v|_q\\
&\leq Ch^{q+1}t^{-\frac{\alpha}{2}}|v|_q, \quad q = 0, 1,
\end{aligned}
$$

which shows the desired estimate (4.15).

Finally, if (4.16) holds, by applying (4.18) with $p = 0$ and $\epsilon \in (0, \frac{1}{2})$, we get

$$\|\delta(t)\| \leq Ch^{-\epsilon} \int_0^t (t-s)^{\frac{\epsilon}{2}\alpha-1} \|Q_h \partial_t^\alpha u_h(s)\| \, ds$$

$$\leq Ch^{2-\epsilon} \int_0^t (t-s)^{\frac{\epsilon}{2}\alpha-1} \|\partial_t^\alpha u_h(s)\| \, ds$$

$$\leq Ch^{2-\epsilon} \int_0^t (t-s)^{\frac{\epsilon}{2}\alpha-1} s^{-\alpha} ds |u_h(0)| \, ds$$

$$\leq C\frac{1}{\epsilon} h^{2-\epsilon} t^{-\alpha(1-\frac{\epsilon}{2})} \|v\|.$$

Then (4.17) follows immediately by choosing $\epsilon = 1/\ell_h$. $\square$

Remark 4.3. *The condition* (4.16) *on the quadrature error operator* Q_h *is satisfied for symmetric meshes; see* [2, Sections 5]. *In case the condition* (4.16) *does not hold, we were able to show only a suboptimal* $O(h)$-*convergence rate for* L_2-*norm of the error, which is reminiscent of the situation in the classical parabolic equation (see, e.g.* [2, Theorem 4.4]).

Remark 4.4. *As we mentioned before, assumption* (4.16) *is valid for symmetric meshes. In fact, in one dimension, the symmetry requirement can be relaxed to almost symmetry* [2, Section 6], *and* (4.17) *can be proven as well.*

Remark 4.5. *We note that we have used a globally quasi-uniform meshes, while the results in* [2] *are valid for meshes that satisfy the inverse inequality only locally.*

5. SPECIAL MESHES

Remark 3.3 (as well as Remark 4.1) suggests that one can achieve a higher order convergence rate for $\nabla(u_h - u)$, if one can get an estimate of the error $\nabla(R_h u - u)$ in some special norm. This could be achieved using the superconvergence property of the gradient available for special meshes and solutions in $H^3(\Omega)$. Examples of special meshes exhibiting superconvergence property include triangulations in which every two adjacent triangles form a parallelogram [13]. To establish a super-convergent recovery of the gradient, Křižek and Neittaanmäki in [13] introduced an operator of the averaged (recovered, postprocessed) gradient $G_h(R_h u)$ of the Ritz projection $R_h u$ of a function u (see [13, equation (2.2)]) with the following properties:

(a) If $u \in H^3(\Omega)$ then, [13, Theorem 4.2]

(5.1) $$\|\nabla u - G_h(R_h u)\| \leq Ch^2 \|u\|_{H^3(\Omega)}.$$

(b) For $\chi \in X_h$ the following bound is valid:

(5.2) $$\|G_h(\chi)\| \leq C\|\nabla\chi\|.$$

The bound (5.2) follows immediately from [13, inequality (3.4)] established for a reference finite element by rescaling and using the fact that $\chi \in X_h$. We point out that one can get a higher order approximation of the ∇u by

$G_h(R_h u)$ due to the special post-processing procedure valid for sufficiently smooth solutions and special meshes.

   This result could be used to establish a higher convergence rate for the semi-discrete Galerkin method (and similarly for the lumped mass method) for smooth initial data. Specifically, we have the following result.

Theorem 5.1. *Let $\mathcal{T}_h$ be strongly uniform triangulation of Ω, that is, every two adjacent triangles form a parallelogram. Then the following estimate is valid*

(5.3) $$\|\nabla u(t) - G_h(u_h(t))\| \le Ch^2 t^{-\alpha/2}\|v\|_2.$$

Proof. It follows from the fact that u satisfies equation (1.1), i.e., $\partial_t^\alpha u(t) = \Delta u$ and from Theorem 2.1 (with $\ell = 1$, $p = 1$ and $q = 2$) that

$$|u(t)|_3 \le Ct^{-\alpha/2}|v|_2.$$

Then using the above super-convergent recovery operator G_h of the gradient with the properties (5.1), (5.2) and the estimate (3.19) for $\theta(t) = R_h u(t) - u_h(t)$, we get

$$\begin{aligned}
\|\nabla u(t) - G_h(u_h(t))\| &= \|\nabla u(t) - G_h(R_h u(t))\| + \|G_h(R_h u(t) - u_h(t))\| \\
&\le Ch^2\|u(t)\|_{H^3(\Omega)} + C\|\nabla\theta(t)\| \\
&\le Ch^2 t^{-\alpha/2}\|v\|_2
\end{aligned}$$

which shows the desired estimate.   □

Remark 5.1. *By repeating the proof of Theorem 5.1 and appealing to Remark 4.1, we can derive the following error estimate for the solution of the lumped mass Galerkin FEM*

$$\|\nabla u(t) - G_h(\bar{u}_h(t))\| \le Ch^2 \ell_h t^{-\alpha/2}\|v\|_2, \quad \ell_h = |\ln h|.$$

Remark 5.2. *Obviously any strongly regular triangulation is also symmetric at each internal vertex and therefore for such meshes we have as well optimal convergence in L_2-norm for nonsmooth data; see (4.17).*

6. Numerical results

   In this section, we present some numerical results to verify the error estimates. We consider the following one-dimensional problem on the unit interval $(0, 1)$

(6.1)
$$\begin{aligned}
\partial_t^\alpha u(x,t) - \partial_{xx}^2 u(x,t) &= 0, \quad 0 < x < 1 \quad 0 \le t \le T, \\
u(0,t) = u(1,t) &= 0, \quad 0 \le t \le T, \\
u(x,0) &= v(x), \quad 0 \le x \le 1.
\end{aligned}$$

We performed numerical tests on five different data:

(a) Smooth initial data: $v(x) = -4x^2 + 4x$; in this case the initial data v is in $H^2(\Omega) \cap H_0^1(\Omega)$, and the exact solution $u(x,t)$ can be represented by a rapidly converging Fourier series:

$$u(x,t) = \frac{16}{\pi^3} \sum_{n=1}^{\infty} \frac{1}{n^3} E_{\alpha,1}(-n^2\pi^2 t^\alpha)(1 - (-1)^n) \sin n\pi x.$$

(b) Initial data in $\dot{H}^1$ (intermediate smoothness):

(6.2)
$$v(x) = \begin{cases} x, & x \in [0, \frac{1}{2}], \\ 1 - x, & x \in (\frac{1}{2}, 1]. \end{cases}$$

(c) Nonsmooth initial data: (1) $v(x) = 1$, (2) $v(x) = x$, and (3) $v(x) = \chi_{[0,\frac{1}{2}]}$, the characteristic function of the interval $(0, \frac{1}{2})$. Since this choice of v is not compatible with the homogeneous Dirichlet boundary data, obviously, in all three cases $v \notin H_0^1$. However, in all these examples $v \in H^s$, $0 < s < \frac{1}{2}$.

(d) We also consider initial data v that is a Dirac $\delta_{\frac{1}{2}}(x)$-function concentrated at $x = \frac{1}{2}$. Such weak data is not covered by our theory. However, it is interesting to see how the method performs for such highly nonsmooth initial data. We note that the choice of the Dirac delta function as initial data is common for certain parameter identification problems in fractional diffusion problems [3].

(e) Variable coefficient case (cf. (3.24)): we take $k(x) = 3 + \sin(2\pi x)$ and initial condition $v(x) = 1$. This class of problems was discussed in Section 3.4.

The exact solution for each example from (a) to (d) can be expressed by an infinite series involving the Mittag-Leffler function $E_{\alpha,1}(z)$. To accurately evaluate the Mittag-Leffler functions, we employ the algorithm developed in [24], which is based on three different approximations of the function, i.e., Taylor series, integral representations and exponential asymptotics, in different regions of the domain.

We divide the unit interval $(0,1)$ into $N + 1$ equally spaced subintervals, with a mesh size $h = 1/(N+1)$. The space X_h consists of continuous piecewise linear functions on the partition. In the case of a constant coefficient $k(x)$ (cf. Section 3.4) we can represent the exact solution to the semidiscrete problem by (3.2) for the standard semidiscrete Galerkin method and by (4.5) for the lumped mass method using the eigenpairs $(\lambda_j^h, \varphi_j^h(x))$ and $(\bar{\lambda}_j^h, \bar{\varphi}_j^h(x))$ of the respective one-dimensional discrete Laplacian $-\Delta_h$ and $-\bar{\Delta}_h$, i.e.,

$$(-\Delta_h \varphi_j^h, v) = \lambda_j^h (\varphi_j^h, v) \quad \text{and} \quad (-\bar{\Delta}_h \bar{\varphi}_j^h, v)_h = \bar{\lambda}_j^h (\bar{\varphi}_j^h, v)_h \quad \forall v \in X_h.$$

Here (w, v) and $(w, v)_h$ refer to the standard L_2-inner product and the approximate L_2-inner product (4.2) on the space X_h, respectively. Then

$$\lambda_j^h = \bar{\lambda}_j^h = \frac{4}{h^2} \sin^2 \frac{\pi j}{2(N+1)} \text{ and } \varphi_j^h(x_k) = \bar{\varphi}_j^h(x_k) = \sqrt{2}\sin(j\pi x_k), \ j = 1, 2, \ldots, N$$

for x_k being a mesh point and linear over the finite elements. These will be used in computing the finite element approximations by the Galerkin and lumped mass methods.

We also have used a direct numerical solution technique by first discretizing the time interval, $t_n = n\tau$, $n = 0, 1, \ldots$, with τ being the time step size, and then using a weighted finite difference approximation of the fractional derivative $\partial_t^\alpha u(x, t_n)$ developed in [16]:

$$\partial_t^\alpha u(x, t_n) = \frac{1}{\Gamma(1-\alpha)} \sum_{j=0}^{n-1} \int_{t_j}^{t_{j+1}} \frac{\partial u(x,s)}{\partial s}(t_n - s)^{-\alpha} \, ds$$

$$\approx \frac{1}{\Gamma(1-\alpha)} \sum_{j=0}^{n-1} \frac{u(x, t_{j+1}) - u(x, t_j)}{\tau} \int_{t_j}^{t_{j+1}} (t_n - s)^{-\alpha} ds$$

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

where the weights $b_j = (j+1)^{1-\alpha} - j^{1-\alpha}$, $j = 0, 1, \ldots, n-1$. It has been shown that if the solution $u(x, t)$ is sufficiently smooth and the time step τ is a constant, then local truncation error of the approximation is bounded by $C\tau^{2-\alpha}$ for some C depending only on u [16, equation (3.3)]. When needed, we have used this approximation on very fine meshes in both space and time to compute a reference solution. Unless otherwise specified, we have set $\tau = 1.0 \times 10^{-6}$, so that the error incurred by temporal discretization can be ignored. Whenever possible, we have compared the accuracy of this reference solution with the exact representation. Our experiments show that with a very small time step size, these two produce the same numerical results.

For each example, we measure the accuracy of the approximation $u_h(t)$ by the normalized error $\|u(t) - u_h(t)\|/\|v\|$ and $\|\partial_x(u(t) - u_h(t))\|/\|v\|$. The normalization enables us to observe the behavior of the error with respect to time in case of nonsmooth initial data. We shall present only numerical results for the lumped mass FEM, since that for the Galerkin FEM is almost identical.

6.1. **Numerical experiments for the smooth initial data: example (a).** In Table 1 we report the numerical results for $t = 1$ and $\alpha = 0.1$, 0.5, 0.95. In Figure 1, we show plots of the results from Table 1 in a log-log scale. We see that the slopes of the error curves are 2 and 1, respectively, for L_2- and H^1-norm of the error. In the last column of Table

1 we also present the error of the recovered gradient $G_h(u_h)$. Since in one-dimension the mid-point of each interval has the desired superconvergence property, the recovered gradient in the case is very simple, just sampled at these points [26, Theorem 1.5.1]. It is clear that the recovered gradient $G_h(u_h)$ exhibits a $O(h^2)$ convergence rate, concurring with the estimates in Theorem 5.1 and Remark 5.1. It is worth noting that the smaller is the α value, the larger is the error (in either the L_2- or H^1-norm). This is attributed to the property of the Mittag-Leffler function $E_{\alpha,1}(-\lambda t^\alpha)$, which, asymptotically, decays faster as α approaches unity, cf. Lemma 2.1 and the representation (2.3).

TABLE 1. Numerical results for smooth data, example (a): $t = 1$ and $h = 2^{-k}$.

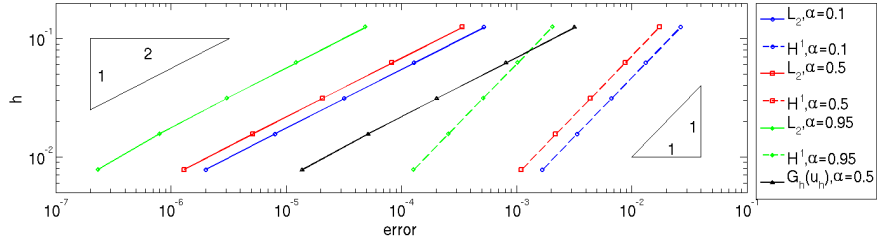|	L_2-error			H^1-error			$G_h(u_h)$-error
k	$\alpha = 0.1$	$\alpha = 0.5$	$\alpha = 0.95$	$\alpha = 0.1$	$\alpha = 0.5$	$\alpha = 0.95$	$\alpha = 0.5$
3	5.23e-4	3.37e-4	4.84e-5	2.65e-2	1.74e-2	2.04e-3	3.20e-3
4	1.29e-4	8.31e-5	1.21e-5	1.33e-2	8.77e-3	1.02e-3	8.07e-4
5	3.21e-5	2.07e-5	3.05e-6	6.69e-3	4.39e-3	5.11e-4	2.03e-4
6	8.01e-6	5.17e-6	7.93e-7	3.34e-3	2.19e-3	2.55e-4	5.17e-5
7	2.00e-6	1.30e-6	2.32e-7	1.67e-3	1.10e-3	1.28e-4	1.39e-5

FIGURE 1. Numerical results for smooth initial data, example (a) with $\alpha = 0.1, 0.5, 0.95$ at $t = 1$.

6.2. **Numerical experiments for the intermediate case of smoothness of the data, example (b).** In this example the initial data $v(x)$ is in $H_0^1(\Omega) \cap H^{\frac{3}{2}-\epsilon}(\Omega)$ with $\epsilon > 0$, and thus it represents an intermediate case. All the numerical results reported in Table 2 were evaluated at $t = 1$ for $\alpha = 0.5$. The slopes of the error curves in a log-log scale are 2 and 1 respectively for L_2 and H^1 norm of the errors, which is in agreement with the theory for the intermediate case; `ratio` in the last column of Table 2, refers to the ratio between the errors as the mesh size h halves.
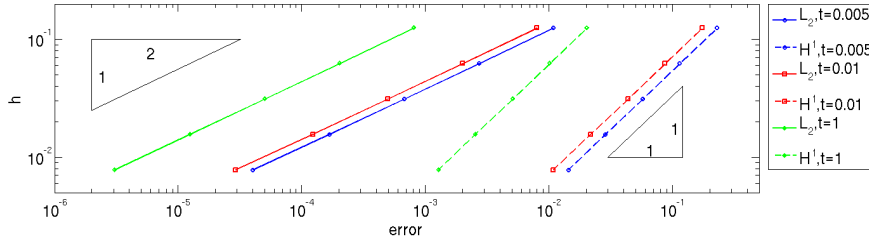
TABLE 2. Numerical results for the intermediate case (b) with $\alpha = 0.5$ at $t = 1$.

h	1/8	1/16	1/32	1/64	1/128	ratio
L_2-error	8.08e-4	2.00e-4	5.00e-5	1.26e-5	3.24e-6	≈ 3.97
H^1-error	1.80e-2	8.84e-3	4.39e-3	2.19e-3	1.10e-3	≈ 2.00

6.3. **Numerical experiments for nonsmooth initial data: example (c).** In Tables 3 and 4 we present the computational results for problem (c), cases (1) and (2). For nonsmooth initial data, we are particularly interested in errors for t close to zero, and thus we also present the error at $t = 0.005$ and $t = 0.01$. In Figure 2 we plot the results shown in Tables 3 and 4, i.e., for problem (c), cases (1) and (2). These numerical results fully confirm the theoretically predicted rates for the nonsmooth initial data.
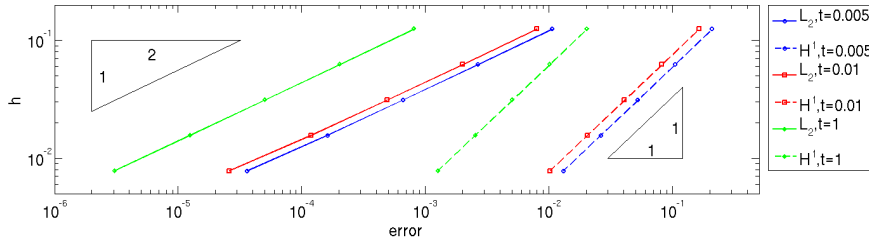
(a) Error plots for Example (c) (1)

(b) Error plots for Example (c) (2)

FIGURE 2. Numerical results for nonsmooth initial data with $\alpha = 0.5$ at $t = 0.005, \ 0.01, \ 1.0$.

Now we consider the third example of nonsmooth case, the characteristic function of the interval $(0, 0.5)$, namely, $v(x) = \chi_{[0,\frac{1}{2}]}$. Note that if we use the interpolation of v as the initial data for the semidiscrete problem, the L_2-error has only a suboptimal first-order convergence. However, if we choose L_2 projection as is discussed in previous sections, then the results agree well with our estimates; see Table 5. We also discretize this example by the Galerkin method, and the results are presented in Table 6. A comparison of Tables 5 and 6 clearly indicates that the Galerkin method and the lumped

TABLE 3. Nonsmooth initial data, example (c) (1): $\alpha = 0.5$ at $t = 0.005, 0.01, 1$.

	h	1/8	1/16	1/32	1/64	1/128	ratio
$t = 0.005$	L_2-norm	1.06e-2	2.65e-3	6.63e-4	1.65e-4	4.02e-5	≈ 4.05
	H^1-norm	2.08e-1	1.04e-1	5.22e-2	2.61e-2	1.30e-2	≈ 2.00
$t = 0.01$	L_2-norm	7.94e-3	1.99e-3	4.93e-4	1.19e-4	2.59e-5	≈ 4.08
	H^1-norm	1.63e-1	8.16e-2	4.08e-2	2.04e-2	1.02e-2	≈ 2.00
$t = 1$	L_2-norm	8.07e-4	2.02e-4	5.03e-5	1.25e-5	3.05e-6	≈ 4.02
	H^1-norm	2.02e-2	1.01e-2	5.04e-3	2.52e-3	1.26e-3	≈ 2.00

TABLE 4. Nonsmooth initial data, example (c)(2): $\alpha = 0.5$ at $t = 0.005, 0.01, 1$.

Time	h	1/8	1/16	1/32	1/64	1/128	ratio
$t = 0.005$	L_2-norm	1.08e-2	2.71e-3	6.79e-4	1.69e-4	4.13e-5	≈ 4.03
	H^1-norm	2.28e-1	1.14e-2	5.71e-2	2.86e-2	1.43e-2	≈ 2.00
$t = 0.01$	L_2-norm	7.98e-3	2.00e-3	4.99e-4	1.23e-4	2.91e-5	≈ 4.02
	H^1-norm	1.73e-1	8.67e-2	4.34e-2	2.17e-2	1.08e-2	≈ 2.00
$t = 1$	L_2-norm	8.05e-4	2.01e-4	5.03e-5	1.25e-5	3.07e-6	≈ 4.01
	H^1-norm	2.02e-2	1.01e-2	5.07e-3	2.53e-3	1.27e-3	≈ 2.00

mass method yield almost identical results for this example. Although not presented, we note that similar observations hold for other examples as well. Hence, we have focused our presentation on the lumped mass method.

TABLE 5. Nonsmooth initial data, example (c)(3): $\alpha = 0.5$ at $t = 0.005, 0.01, 1$

time	h	1/8	1/16	1/32	1/64	1/128	ratio
$t = 0.005$	L_2-norm	8.54e-3	2.16e-3	5.45e-4	1.31e-4	3.17e-5	≈ 4.06
	H^1-norm	2.18e-1	1.08e-1	5.38e-2	2.68e-2	1.34e-2	≈ 2.00
$t = 0.01$	L_2-norm	6.54e-3	1.64e-3	4.14e-4	1.06e-4	2.90e-5	≈ 3.96
	H^1-norm	1.63e-1	8.04e-2	4.00e-2	2.00e-2	9.96e-3	≈ 2.00
$t = 1$	L_2-norm	8.10e-4	2.03e-4	5.07e-5	1.27e-5	3.23e-6	≈ 3.99
	H^1-norm	1.82e-2	9.02e-3	4.46e-3	2.22e-3	1.11e-3	≈ 2.01

6.4. **Numerical experiments for initial data a Dirac δ-function.** We note that this case is not covered by our theory. Formally, the orthogonal L_2-projection P_h is not well defined for such functions. However, we can look at (v, χ) for $\chi \in X_h \subset H_0^1(\Omega)$ as a duality pairing between the spaces $H^{-1}(\Omega)$

TABLE 6. Nonsmooth initial data, example (c)(3) with $\alpha = 0.5$ at $t = 0.005, 0.01, 1$ by the Galerkin method.

time	h	1/8	1/16	1/32	1/64	1/128	ratio
$t = 0.005$	L_2-norm	8.60e-3	2.14e-3	5.30e-4	1.28e-4	2.85e-5	≈ 4.11
	H^1-norm	1.78e-1	9.78e-2	5.11e-2	2.61e-2	1.32e-2	≈ 1.96
$t = 0.01$	L_2-norm	6.56e-3	1.64e-3	4.06e-4	9.94e-5	2.29e-5	≈ 4.11
	H^1-norm	1.34e-1	7.34e-2	3.82e-2	1.95e-2	9.85e-3	≈ 1.95
$t = 1$	L_2-norm	8.07e-4	2.02e-4	5.04e-5	1.25e-5	3.09e-6	≈ 4.03
	H^1-norm	1.54e-2	8.30e-3	4.29e-3	2.18e-3	1.10e-3	≈ 1.96

and $H_0^1(\Omega)$ and therefore $(\delta_{\frac{1}{2}}, \chi) = \chi(\frac{1}{2})$. If $x = \frac{1}{2}$ is a mesh point, say x_L, then we can define $P_h \delta_{\frac{1}{2}}$ appropriately with its finite element expansion given by the Lth column of the inverse of the mass matrix. This was the initial data for the semidiscrete problem that we used in our computations. In Table 7 we show the L_2- and H^1-norm of the error for this case. It is noted that the H^1-norm of the error converges as $O(h^{\frac{1}{2}})$, while the error in the L_2-norm converges as $O(h^{\frac{3}{2}})$; see the last column of Table 7. It is quite remarkable that we can practically have good convergence rates in L_2- and H^1-norm for such very weak solutions. A theoretical justification of these rates is a subject of our current work.

TABLE 7. Lumped mass FEM with initial data a Dirac δ-function, $\alpha = 0.5$, $t = 0.005, 0.01, 1$.

time	h	1/8	1/16	1/32	1/64	1/128	ratio
$t = 0.005$	L_2-norm	7.24e-2	2.66e-2	9.54e-3	3.40e-3	1.21e-3	≈ 2.79
	H^1-norm	1.51e0	1.07e0	7.60e-1	5.40e-1	3.81e-1	≈ 1.41
$t = 0.01$	L_2-norm	5.20e-2	1.89e-2	6.77e-3	2.40e-3	8.54e-4	≈ 2.79
	H^1-norm	1.07e0	7.59e-1	5.37e-1	3.80e-1	2.70e-1	≈ 1.41
$t = 1$	L_2-norm	5.47e-3	1.93e-3	6.84e-4	2.42e-4	8.56e-5	≈ 2.79
	H^1-norm	1.07e-1	7.58e-2	5.37e-2	3.80e-2	2.70e-2	≈ 1.41

6.5. **Numerical experiments for variable coefficients, example (e).** Although we do not have an explicit representation of the exact solution, we compare the numerical solution with the approximate solution obtained on very fine meshes, namely, with mesh-size $h = 1/512$ and time-step size $\tau = 1.0 \times 10^{-5}$. Normalized L_2- and H^1-norms of the error are reported in Table 8 for $t = 0.01$ for $\alpha = 0.5$. The results confirm the theoretically predicted rate.

In summary, the convergence rates observed for all the numerical experiments are in excellent agreement with the theoretical findings for both

TABLE 8. Numerical results for variable coefficients and non-smooth initial data with $\alpha = 0.5$ at $t = 0.01$.

h	1/8	1/16	1/32	1/64	1/128	ratio
L_2-error	3.24e-3	8.21e-4	2.05e-4	5.09e-5	1.23e-6	≈ 4.02
H^1-error	7.15e-2	3.60e-2	1.80e-2	8.94e-3	4.36e-3	≈ 2.01

smooth and nonsmooth initial data, including also the case of the recovered gradient $G_h(u_h)$ discussed in Section 5, see also Remark 5.1.

Acknowledgments

References

[1] J.-P. Bouchaud and A. Georges. Anomalous diffusion in disordered media: statistical mechanisms, models and physical applications. *Phys. Rep.*, 195(4-5):127–293, 1990.

[2] P. Chatzipantelidis, R. Lazarov, and V. Thomée. Some error estimates for the lumped mass finite element method for a parabolic problem. *Math. Comp.*, 81(277):1–20, 2012.

[3] J. Cheng, J. Nakagawa, M. Yamamoto, and T. Yamazaki. Uniqueness in an inverse problem for a one-dimensional fractional diffusion equation. *Inverse Problems*, 25(11):115002, 1–16, 2009.

[4] M. Crouzeix and V. Thomée. The stability in L_p and W_p^1 of the L_2-projection onto finite element function spaces. *Math. Comp.*, 48(178):521–532, 1987.

[5] L. Debnath. Recent applications of fractional calculus to science and engineering. *Int. J. Math. Math. Sci.*, 54:3413–3442, 2003.

[6] W. Deng. Finite element method for the space and time fractional Fokker-Planck equation. *SIAM J. Numer. Anal.*, 47(1):204–226, 2008/09.

[7] M. Djrbashian. *Harmonic Analysis and Boundary Value Problems in the Complex Domain*. Birkhäuser, Basel, 1993.

[8] A. Ern and J.-L. Guermond. *Theory and Practice of Finite Elements*, volume 159 of *Applied Mathematical Sciences*. Springer-Verlag, New York, 2004.

[9] V. Ervin and J. Roop. Variational formulation for the stationary fractional advection dispersion equation. *Numer. Methods Partial Diff. Eq.*, 22(3):558–576, 2006.

[10] B. Jin and X. Lu. Numerical identification of a Robin coefficient in parabolic problems. *Math. Comput.*, 81(279):1369–1398, 2012.

[11] Y. L. Keung and J. Zou. Numerical identifications of parameters in parabolic systems. *Inverse Problems*, 14(1):83–100, 1998.

[12] A. Kilbas, H. Srivastava, and J. Trujillo. *Theory and Applications of Fractional Differential Equations*. Elsevier, Amsterdam, 2006.

[13] M. Křížek and P. Neittaanmäki. On a global superconvergence of the gradient of linear triangular elements. *J. Comput. Appl. Math.*, 18(2):221–233, 1987.

[14] T. Langlands and B. Henry. The accuracy and stability of an implicit solution method for the fractional diffusion equation. *J. Comput. Phys.*, 205(2):719–736, 2005.

[15] X. Li and C. Xu. A space-time spectral method for the time fractional diffusion equation. *SIAM J. Numer. Anal.*, 47(3):2108–2131, 2009.

[16] Y. Lin and C. Xu. Finite difference/spectral approximations for the time-fractional diffusion equation. *J. Comput. Phys.*, 225(2):1533–1552, 2007.

[17] W. McLean and V. Thomée. Time discretization of an evolution equation via Laplace transforms. *IMA J. Numer. Anal.*, 24(3):439–463, 2004.

[18] W. McLean and V. Thomée. Maximum-norm error analysis of a numerical solution via Laplace transformation and quadrature of a fractional-order evolution equation. *IMA J. Numer. Anal.*, 30(1):208–230, 2010.

[19] M. Meerschaert, H.-P. Scheffler, and C. Tadjeran. Finite difference methods for two-dimensional fractional dispersion equation. *J. Comput. Phys.*, 211(1):249–261, 2006.

[20] K. Mustapha. An implicit finite-difference time-stepping method for a sub-diffusion equation, with spatial discretization by finite elements. *IMA J. Numer. Anal.*, 31(2):719–739, 2011.

[21] R. Nigmatulin. The realization of the generalized transfer equation in a medium with fractal geometry. *Phys. Stat. Sol. B*, 133:425–430, 1986.

[22] I. Podlubny. *Fractional Differential Equations*. Academic Press, San Diego, CA, 1999.

[23] 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.

[24] H. Seybold and R. Hilfer. Numerical algorithm for calculating the generalized Mittag-Leffler function. *SIAM J. Numer. Anal.*, 47(1):69–88, 2008/09.

[25] V. Thomée. *Galerkin Finite Element Methods for Parabolic Problems*, volume 25 of *Springer Series in Computational Mathematics*. Springer-Verlag, Berlin, 1997.

[26] L. Wahlbin. *Superconvergence in Galerkin Finite Element Methods*, volume 1605 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, 1995.

[27] S. Yuste and L. Acedo. An explicit finite difference method and a new von Neumann-type stability analysis for fractional diffusion equations. *SIAM J. Numer. Anal.*, 42(5):1862–1874, 2005.

Department of Mathematics and Institute for Applied Mathematics and Computational Science, Texas A&M University, College Station, TX, 77843, USA

E-mail address: `btjin@math.tamu.edu`

Department of Mathematics, Texas A&M University, College Station, TX, 77843, USA

E-mail address: `lazarov@math.tamu.edu`

Department of Mathematics, Texas A&M University, College Station, TX, 77843, USA

E-mail address: `zzhou@math.tamu.edu`