SIAM J. NUMER. ANAL.
Vol. 51, No. 1, pp. 445–466

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 a homogeneous time-fractional diffusion equation with an initial condition $v(x)$ and a homogeneous Dirichlet boundary condition in a bounded convex polygonal domain Ω. We study two semidiscrete approximation schemes, i.e., the Galerkin finite element method (FEM) and lumped mass Galerkin FEM, using piecewise linear functions. We establish almost optimal with respect to the data regularity error estimates, including the cases of smooth and nonsmooth initial data, i.e., $v \in H^2(\Omega) \cap H_0^1(\Omega)$ and $v \in L_2(\Omega)$. For the lumped mass method, the optimal L_2-norm error estimate is valid only under an additional assumption on the mesh, which in two dimensions is known to be satisfied for symmetric meshes. Finally, we present some numerical results that give insight into the reliability of the theoretical study.

Key words. finite element method, fractional diffusion, optimal error estimates, semidiscrete Gelerkin method, lumped mass method

AMS subject classifications. 65M60, 65N30, 65N15

DOI. 10.1137/120873984

1. Introduction.

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

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

where Ω is a bounded convex polygonal domain in $\mathbb{R}^d\,(d = 1, 2, 3)$ with a boundary $\partial\Omega$, 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., [10, p. 91] or [22, p. 78])

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

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 $\frac{du}{dt}$ [10], and thus 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 subdiffusion), 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 differential equations arising in dynamical

*Received by the editors April 18, 2012; accepted for publication (in revised form) December 4, 2012; published electronically February 12, 2013. The research of R. Lazarov and Z. Zhou was supported in part by US NSF grant DMS-1016525. The work of all authors has been supported also by award KUS-C1-016-04, made by King Abdullah University of Science and Technology (KAUST).
http://www.siam.org/journals/sinum/51-1/87398.html

†Department of Mathematics and Institute for Applied Mathematics and Computational Science, Texas A&M University, College Station, TX 77843 (btjin@math.tamu.edu, lazarov@math.tamu.edu, zzhou@math.tamu.edu).

systems in control theory, electrical circuits with fractance, generalized voltage divider, viscoelasticity, fractional order multipoles in electromagnetism, electrochemistry, and models of neurons in biology is provided in [4]; see also [22].

The extraordinary modeling capabilities of FPDEs have generated considerable interest in devising, analyzing, and testing numerical methods for such problems. As a result, a number of numerical techniques were developed, and their stability and convergence were investigated [12, 14, 19, 20, 27]. Yuste and Acedo [27] presented a numerical scheme by combining the forward time centered space method and the Grunwald–Letnikov method, and they provided a stability analysis. By exploiting the variational framework due to Ervin and Roop [7], Li and Xu [14] developed a spectral method in both temporal and spatial variable and established various a priori error estimates. Mustapha [20] studied semidiscrete in time and fully discrete schemes and derived error bounds for smooth initial data [20, Theorem 4.3].

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 solution smoothness 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 related inverse problems; see, e.g., [3], [23, Problem (4.12)], and also [8, 9] for its parabolic counterpart.

There are a few papers considering the construction and analysis of numerical methods with optimal (with respect to the solution regularity) error estimates for the following equation with a positive type memory term:

$$\partial_t u - \frac{1}{\Gamma(\alpha)} \int_0^t (t-\tau)^{\alpha-1} \Delta u(\tau) d\tau = f(x,t), \quad t > 0, \quad 0 < \alpha < 1, \tag{1.3}$$

which is closely related to but different from (1.1). For example, McLean and Thomée [15, 16] developed a numerical method based on spatial finite element discretization and Laplace transformation with quadratures in time for (1.3) with a homogeneous Dirichlet boundary data. In [15, Theorem 5.1] the convergence of the method was 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 [16, Theorem 4.2] a maximum-norm error estimate of order $O(h^2 \ell_h^2)$ was shown for initial data $v \in H^2(\Omega) \cap H_0^1(\Omega)$.

The scarcity of optimal (with respect to the solution regularity) error estimates for the numerical schemes for FPDEs with nonsmooth data contrasts sharply with the standard parabolic counterpart. Here the error analysis is complete and various optimal estimates are available [25]. The key ingredient of the analysis is the smoothing property of the parabolic operator and its discrete counterpart [25, Lemmas 3.2 and 2.5]. For FPDE (1.1), such a property has been established recently [23]; see Theorem 2.2 below for details.

The goal of this paper 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 finite element method (FEM) for problem (1.1) on convex polygonal domains.

Now we describe the Galerkin scheme, using the standard notation from [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 problem (1.1) now reads: find $u_h(t) \in X_h$ such that

$$(1.4) \qquad (\partial_t^\alpha u_h, \chi) + a(u_h, \chi) = (f, \chi) \quad \forall \chi \in X_h, \ T \geq t > 0, \quad u_h(0) = v_h,$$

where $a(u, w) = (\nabla u, \nabla w)$ for $u, w \in H_0^1(\Omega)$ and $v_h \in X_h$ is an approximation of the initial data v. The choice of v_h will depend on the smoothness of the initial data v. We shall study the convergence of the semidiscrete Galerkin FEM (1.4) for the case of initial data $v \in \dot{H}^q(\Omega)$, $q = 0, 1, 2$. (See section 2.2 for their definitions.) Following Thomée [25], we shall take $v_h = R_h v$ in case of smooth initial data, i.e., $q = 2$, and $v_h = P_h v$ in case of nonsmooth initial data, i.e., $q = 0$, where R_h and P_h are Ritz and the orthogonal $L_2(\Omega)$-projection on the finite element space X_h, respectively.

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 the solution u_h satisfies the following error bounds [25, Theorems 3.1 and 3.2]:

$$\|u_h(t) - u(t)\|_q \leq C h^{2-q} t^{-1+\frac{p}{2}} \|v\|_p, \quad p = 0, 2, \quad q = 0, 1.$$

In a recent work [17], McLean and Thomée established the following estimates for the Galerkin method for problem (1.1):

$$\|u(t) - u_h(t)\| \leq C h^2 \|v\|_2 \quad \text{and} \quad \|u(t) - u_h(t)\| \leq C t^{-\alpha} h^2 \|v\|.$$

The proof is based on some refined estimates of the Laplace transform in time for the error $u(t) - u_h(t)$, which is rather different from our technique. Further, no estimates on the gradient of the error were provided, and mass lumping was not discussed.

In this paper we establish analogous estimates for the semidiscrete Galerkin method (1.4) for problem (1.1). The main difficulty in the error analysis stems from limited smoothing properties of the FPDE, cf. Theorem 2.2. 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, whereas the exponential function in the standard parabolic case 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. First, in case of smooth initial data, we derived an error bound uniformly in $t \geq 0$ (cf. Theorem 3.5), as is in the case of the standard parabolic problem. Second, 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 3.7). It is similar to the parabolic counterpart but derived for quasi-uniform meshes and with an additional log-factor ℓ_h.

Further, we study the more practical lumped mass scheme. We show the same convergence rate for initial data $v \in \dot{H}^2(\Omega)$ (cf. Theorem 4.4) and an 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 Theorem 4.5). For nonsmooth data $v \in L_2(\Omega)$ for general quasi-uniform meshes, we are only able to establish a suboptimal L_2-error bound of order $O(h \ell_h t^{-\alpha})$; cf. (4.15). For a class of special triangulations satisfying condition (4.16), an almost optimal estimate (4.17) holds, analogous to its parabolic counterpart [2, Theorem 4.1].

Finally, in Theorem 5.1 we establish a superconvergence result for the error of the postprocessed gradient 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 $O(h^2t^{-\frac{\alpha}{2}})$ and $O(h^2\ell_h t^{-\frac{\alpha}{2}})$ for the Galerkin and lumped mass approximation, respectively.

The rest of the paper is organized as follows. In section 2, we recall some basic properties of the Mittag-Leffler function, the smoothing property of (1.1), and 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. Finally, in section 6 we present numerical tests on various one-dimensional examples, including both smooth and nonsmooth data, which confirm our theoretical study. Throughout the paper, 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 FPDE (1.1), and basic estimates for the finite-element projection operators.

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

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

The Mittag-Leffler function $E_{\alpha,\beta}(z)$ is a two-parameter family of entire functions in z of order α^{-1} and type 1 [10, pp. 42]. It generalizes the exponential function in the sense that $E_{1,1}(z) = e^z$. Two most important members of this family are $E_{\alpha,1}(z)$ and $E_{\alpha,\alpha}(z)$, 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 Djrbashian [5, Chapter 1]. The estimate (2.1) below can be found in [10, pp. 43] or [22, Theorem 1.4], while (2.2) is discussed in [10, Lemma 2.33].

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*

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

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

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

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 Dirichlet eigenvalues and eigenfunctions of $-\Delta$ on the domain Ω. 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 $L_2(\Omega)$, $|v|_1$ is 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 $E(t)$:

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

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 $u(t) = E(t)v$. It follows from an eigenfunction expansion and (2.2) [23]. Further, for the nonhomogeneous equation with vanishing initial data $v \equiv 0$, we shall use the operator defined for $\chi \in L_2(\Omega)$ by

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

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 $f(x,t) \in L_\infty((0,T); L_2(\Omega))$ and $v \in L_2(\Omega)$, 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), we have the following stability estimates, essentially established in [23, Theorem 2.1] and slightly extended in the theorem below; see also [18] for related regularity estimates. Since these estimates will play a key role in the error analysis of the FEM approximations, we give some hints of the proof.

THEOREM 2.2. *The solution $u(t) = E(t)v$ to problem* (1.1) *with $f \equiv 0$ satisfies*

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

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

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

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

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

$$\|u(t)\| \le C\|v\|. \tag{2.7}$$

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

$$\begin{aligned}|\partial_t^\alpha u(t)|_p^2 &= \sum_{j=1}^\infty \lambda_j^{2+p} E_{\alpha,1}(-\lambda_j t^\alpha)^2 \,(v,\varphi_j)^2 \\ &= t^{-\alpha(2+p-q)} \sum_{j=1}^\infty (\lambda_j t^\alpha)^{2+p-q} E_{\alpha,1}(-\lambda_j t^\alpha)^2 \lambda_j^q (v,\varphi_j)^2 \\ &\le C t^{-\alpha(2+p-q)} \sum_{j=1}^\infty \frac{(\lambda_j t^\alpha)^{2+p-q}}{(1+\lambda_j t^\alpha)^2} \lambda_j^q (v,\varphi_j)^2 \le C t^{-\alpha(2+p-q)} |v|_q^2,\end{aligned}$$

where we have used the fact that in view of Young's inequality, $\frac{(\lambda_j t^\alpha)^{2+p-q}}{(1+\lambda_j t^\alpha)^2} \le C$ for $p \le q \le p+2$. This completes the proof of the theorem. ∎

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 [25, Lemma 3.2]. This reflects the fact that the FPDE has limited smoothing properties. The limited smoothing is also present 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 $R_h : H_0^1(\Omega) \to X_h$, respectively, defined by

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

It is well known that the operators P_h and R_h have the following approximation properties; cf. [25, Lemma 1.1] or [6, Theorems 3.16 and 3.18].

LEMMA 2.3. *The operators P_h and R_h satisfy*

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

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

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

3. Semidiscrete Galerkin FEM. In this section we derive error estimates for the standard semidiscrete Galerkin FEM. We begin with basic facts of the semidiscrete Galerkin FEM and the smoothing properties of relevant solution operators. The error estimates hinge crucially on the properties of the operator $\bar E_h$.

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

$$-(\Delta_h\psi, \chi) = (\nabla\psi, \nabla\chi) \quad \forall \psi,\, \chi \in X_h \tag{3.1}$$

and $f_h = P_h f$, we may write the spatially discrete problem (1.4) as

$$\partial_t^\alpha u_h(t) - \Delta_h u_h(t) = f_h(t) \quad \text{for} \quad t \ge 0 \quad \text{with} \quad u_h(0) = v_h. \tag{3.2}$$

Now we give a representation of the solution of (3.2) 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 $t > 0$:

$$E_h(t)v_h = \sum_{j=1}^{N} E_{\alpha,1}(-\lambda_j^h t^\alpha)(v_h, \varphi_j^h)\varphi_j^h, \tag{3.3}$$

$$\bar{E}_h(t)f_h = \sum_{j=1}^{N} t^{\alpha-1} E_{\alpha,\alpha}(-\lambda_j^h t^\alpha)\,(f_h, \varphi_j^h)\,\varphi_j^h. \tag{3.4}$$

Then the solution $u_h(x,t)$ of the discrete problem (3.2) can be expressed by

$$u_h(x,t) = E_h(t)v_h + \int_0^t \bar{E}_h(t-s)f_h(s)\,ds. \tag{3.5}$$

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

$$|||\psi|||_p^2 = \sum_{j=1}^{N} (\lambda_j^h)^p(\psi, \varphi_j^h)^2, \quad \psi \in X_h. \tag{3.6}$$

Clearly, the norm $|||\cdot|||_p$ is well defined for all real p. By 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 $\psi \in X_h$.

We shall show some 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* (3.3) *and $v_h \in X_h$. Then*

$$|||(\partial_t^\alpha)^\ell u_h(t)|||_p = |||(\partial_t^\alpha)^\ell E_h(t)v_h|||_p \le Ct^{-\alpha(\ell+\frac{p-q}{2})}|||v_h|||_q, \quad t > 0, \tag{3.7}$$

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

Proof. First, consider the case $\ell = 0$. Using the representation (3.3) of the solution $u_h(t)$ and Lemma 2.1 we get for $q \le p$

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

Here in the last inequality we have used the fact that for $q \le p$ and $p \le q \le p+2$ 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 (3.3) 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 appealing to Lemma 2.1 and Young's inequality, we obtain

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

The desired estimate follows from this immediately. ∎

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

Lemma 3.2. *Let $\bar{E}_h$ be defined by* (3.4) *and $\psi \in X_h$. Then we have for all $t > 0$,*

$$
|||\bar{E}_h(t)\psi|||_p \le \begin{cases} Ct^{-1+\alpha(1+\frac{q-p}{2})}|||\psi|||_q, & p-2 \le q \le p, \\ Ct^{-1+\alpha}|||\psi|||_q, & p < q. \end{cases} \tag{3.8}
$$

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 $q \le p$

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

where the last line follows from $0 \le p-q \le 2$. The assertion for $p < q$ follows from the fact that $\{\lambda_j^h\}$ are bounded away from zero independent of the mesh size h. ∎

Remark 3.1. Lemma 3.2 shows 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. *For any $l > s$, there exists a constant C independent of h such that*

$$
|||\psi|||_l \le Ch^{s-l}|||\psi|||_s \quad \forall \psi \in X_h. \tag{3.9}
$$

Proof. For quasi-uniform triangulations $\mathcal{T}_h$ the inverse inequality $|\psi|_1 \le Ch^{-1}\|\psi\|$ holds for all $\psi \in X_h$. By the definition of $-\Delta_h$ this implies $\max_{1\le j\le N}\lambda_j^h \le C/h^2$. Thus, for the norm $|||\cdot|||_p$ defined in (3.6), there holds for any real $l > s$

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

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

$$
u_h - u = (u_h - R_h u) + (R_h u - u) := \vartheta + \varrho. \tag{3.10}
$$

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

$$\|\varrho(t)\| + h\|\nabla\varrho(t)\| \le Ch^2 t^{-\alpha(1-\frac{q}{2})}|v|_q \quad v \in \dot{H}^q, \tag{3.11}$$

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* (1.4)*, respectively, with $v_h = R_h v$. Then for $\vartheta(t) = u_h(t) - R_h u(t)$ we have*

$$\|\vartheta(t)\| + h\|\nabla\vartheta(t)\| \le Ch^2|v|_2. \tag{3.12}$$

Proof. Using the identity $\Delta_h R_h = P_h \Delta$, we note that ϑ satisfies

$$\partial_t^\alpha \vartheta(t) - \Delta_h \vartheta(t) = -P_h \partial_t^\alpha \varrho(t) \quad \text{for} \quad t > 0. \tag{3.13}$$

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 (3.5),

$$\vartheta(t) = -\int_0^t \bar{E}_h(t-s) P_h \partial_t^\alpha \varrho(s)\, ds. \tag{3.14}$$

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 $q = 1, 2$,

$$\begin{aligned} \|\nabla \bar{E}_h(t-s) P_h \partial_t^\alpha \varrho(s)\| &\le C(t-s)^{\frac{\alpha}{2}-1}\, \|\partial_t^\alpha \varrho(s)\| \\ &\le Ch(t-s)^{\frac{\alpha}{2}-1}\, |\partial_t^\alpha u(s)|_1 \\ &\le Ch(t-s)^{\frac{\alpha}{2}-1} s^{\alpha(-\frac{3}{2}+\frac{q}{2})}|v|_q. \end{aligned} \tag{3.15}$$

By substituting this inequality into (3.14) we obtain that for $q = 1, 2$

$$\|\nabla\vartheta(t)\| \le Ch \int_0^t (t-s)^{\frac{\alpha}{2}-1} s^{\alpha(-\frac{3}{2}+\frac{q}{2})}\, ds\, |v|_q \le Cht^{-\alpha(1-\frac{q}{2})}|v|_q, \tag{3.16}$$

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

$$\begin{aligned} \int_0^t (t-s)^{\frac{\alpha}{2}-1} s^{\alpha(-\frac{3}{2}+\frac{q}{2})}\, ds &= t^{\frac{\alpha}{2}-\frac{3\alpha}{2}+\frac{q\alpha}{2}} \int_0^1 (1-s)^{\frac{\alpha}{2}-1} s^{\alpha(-\frac{3}{2}+\frac{q}{2})} ds \\ &= B(\tfrac{\alpha}{2}, \alpha(-\tfrac{3}{2}+\tfrac{q}{2})+1) t^{-\alpha(1-\frac{q}{2})}, \end{aligned}$$

where $B(\cdot,\cdot)$ is the Beta function. Since for $q = 1, 2$, $\frac{\alpha}{2} > 0$, and $\frac{-3+q}{2}\alpha + 1 > 0$, the value $B(\alpha, \alpha(-\frac{3}{2}+\frac{q}{2})+1)$ is finite. Taking $q = 2$ yields the desired estimate for $\nabla\vartheta$.

Next, by Lemma 3.2 with $p = q = 0$ and that of the operator E in Theorem 2.2 with $\ell = 1$ and $p = q = 2$, we get

$$\begin{aligned} \|\vartheta(t)\| &\le \int_0^t \|\bar{E}_h(t-s) P_h \partial_t^\alpha \varrho(s)\| ds \le C \int_0^t (t-s)^{\alpha-1} \|\partial_t^\alpha \varrho(s)\| ds \\ &\le Ch^2 \int_0^t (t-s)^{\alpha-1} |\partial_t^\alpha u(s)|_2 ds \le Ch^2 \int_0^t (t-s)^{\alpha-1} s^{-\alpha} ds |v|_2 \\ &= CB(\alpha, 1-\alpha) h^2 |v|_2. \end{aligned} \tag{3.17}$$

This completes the proof. ∎

The main result in this part follows directly from estimates (3.11) and (3.12).

THEOREM 3.5. *Let u and u_h be the solutions of* (1.1) *and* (1.4)*, respectively, with $v_h = R_h v$. Then*

$$\|u_h(t) - u(t)\| + h\|\nabla(u_h(t) - u(t))\| \le Ch^2 |v|_2. \tag{3.18}$$

Remark 3.2. As a by-product of estimates (3.11) and (3.16) we also got a bound for the error for $v \in \dot{H}^1(\Omega)$ and $v_h = R_h v$:

$$\|\nabla(u_h(t) - u(t))\| \le Cht^{-\frac{\alpha}{2}} |v|_1. \tag{3.19}$$

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

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

which yields

$$\|\nabla \vartheta\| \le Ch^2 t^{-\frac{\alpha}{2}} |v|_2. \tag{3.20}$$

3.3. Error estimates for nonsmooth 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 $P_h v$ 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 2.3 and Theorem 2.2 we have

$$\|\widetilde{\varrho}(t)\| + h\|\nabla \widetilde{\varrho}(t)\| \le Ch^2 |u(t)|_2 \le Ch^2 t^{-\alpha(1-\frac{q}{2})} \|v\|_q, \quad q = 0, 1. \tag{3.21}$$

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 using the identity $\Delta_h R_h = P_h \Delta$, we get the following problem for $\widetilde{\vartheta}$:

$$\partial_t^\alpha \widetilde{\vartheta}(t) - \Delta_h \widetilde{\vartheta}(t) = -\Delta_h (R_h u - P_h u)(t), \quad t > 0, \quad \widetilde{\vartheta}(0) = 0. \tag{3.22}$$

Then with the help of formula (3.4), $\widetilde{\vartheta}(t)$ can be represented by

$$\widetilde{\vartheta}(t) = -\int_0^t \bar{E}_h(t-s) \Delta_h (R_h u - P_h u)(s) \, ds. \tag{3.23}$$

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

LEMMA 3.6. *Let $\widetilde{\vartheta}(t)$ be defined by* (3.23)*. Then for $p = 0, 1$, $q = 0, 1$, and $\ell_h = |\ln h|$, the following estimate holds:*

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

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

$$\begin{aligned} \|\widetilde{\vartheta}(t)\|_p &\le \int_0^t \|\bar{E}_h(t-s) \Delta_h (R_h u - P_h u)(s)\|_p ds \\ &\le \int_0^t (t-s)^{\frac{\epsilon}{2}\alpha - 1} |||\Delta_h (R_h u - P_h u)|||_{p-2+\epsilon} ds \\ &\le \int_0^t (t-s)^{\frac{\epsilon}{2}\alpha - 1} |||R_h u - P_h u|||_{p+\epsilon} ds := A. \end{aligned}$$

Further, we apply the inverse inequality (3.9) for $R_hu - P_hu$, the bounds (2.9) and (2.10) for $P_hu - u$ and $R_hu - u$, respectively, and the smoothing property (2.5) with $\ell = 0$ and $p = 2$ to get

$$\begin{aligned}
A &\le Ch^{-\epsilon} \int_0^t (t-s)^{\frac{\epsilon}{2}\alpha - 1} \|R_hu - P_hu\|_p ds \\
&\le Ch^{2-p-\epsilon} \int_0^t (t-s)^{\frac{\epsilon}{2}\alpha - 1} \|u(s)\|_2 ds \\
&\le Ch^{2-p-\epsilon} \int_0^t (t-s)^{\frac{\epsilon}{2}\alpha - 1} s^{-\alpha(1-\frac{q}{2})} ds \, \|v\|_q \\
&= CB\left(\tfrac{\epsilon}{2}\alpha, 1-\alpha+\tfrac{q}{2}\alpha\right) h^{2-p-\epsilon} t^{-\alpha(1-\frac{q}{2}-\frac{\epsilon}{2})} \|v\|_q \\
&\le C\epsilon^{-1} h^{2-p-\epsilon} t^{-\alpha(1-\frac{q}{2})} \|v\|_q.
\end{aligned}$$

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 Laurent expansion of the Gamma function. The desired assertion follows by choosing $\epsilon = 1/\ell_h$. ∎

Then Lemma 3.6 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 3.7. *Let u and u_h be the solutions of* (1.1) *and* (1.4) *with $v_h = P_hv$, respectively. Then with $\ell_h = |\ln h|$*

$$\|u_h(t) - u(t)\| + h\|\nabla(u_h(t) - u(t))\| \le Ch^2 \, \ell_h \, t^{-\alpha(1-\frac{q}{2})} \|v\|_q, \quad q = 0, 1. \tag{3.24}$$

Remark 3.4. For $v \in \dot{H}^1(\Omega)$ and $v_h = R_hv$, we have established the estimate (3.19), which is slightly better than (3.24), 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 FEM 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 (1.4) with a bilinear form $a(\cdot,\cdot) : V \times V \mapsto \mathbb{R}$ of the form

$$a(u, \chi) = \int_\Omega (k(x)\nabla u \cdot \nabla \chi + c(x)u\chi) \, dx, \tag{3.25}$$

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

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

where $c_0, c_1 > 0$ are constants and the bilinear form $a(\cdot,\cdot)$ is coercive on $V \subset H^1(\Omega)$. Further, we assume that the problem $a(w, \chi) = (f, \chi)$ for all $\chi \in V$ has a unique solution $w \in V$, which for $f \in L_2(\Omega)$ has full elliptic regularity, $\|w\|_{H^2} \le C\|f\|_{L_2}$.

Under these conditions we can define a positive definite operator $\mathcal{A} : V \to V'$, 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

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

Then all results for problem (1.1) can be easily extended to the case with the elliptic part of this more general form.

4. Lumped mass FEM. Here we consider the more practical lumped mass FEM (see, e.g., [25, Chapter 15, pp. 239–244]) and study the convergence rates for smooth and nonsmooth initial data.

4.1. Derivation of the lumped mass FEM. 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

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

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

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

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

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

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

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

Also, we introduce a projection operator $\bar{P}_h : L_2(\Omega) \to X_h$ 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

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

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 (3.4), we introduce the operator $\bar{F}_h$ by

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

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

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

For our analysis we shall need the following modification of the discrete norm (3.6), $|||\cdot|||_p$, on the space X_h:

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

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 proof is a simple consequence of the definitions and is omitted. ∎

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

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

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. ∎

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

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

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 [2, Lemma 2.4].

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

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

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

THEOREM 4.4. *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. 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. Upon noting the estimate (3.18) for $u_h - u$, it suffices to show

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

It follows from the definitions of $u_h(t)$, $\bar{u}_h(t)$, 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 $\chi \in X_h$

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

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. ∎

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 $\chi \in X_h$

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

Consequently,

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

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

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

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

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

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

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

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

4.3. Error estimates for nonsmooth initial data. Now we consider the case of nonsmooth initial data $v \in L_2(\Omega)$ as well as the intermediate case $v \in \dot{H}^1$. Due to the lower regularity, we take $v_h = P_h v$. Like 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 (3.24) it suffices to establish proper bounds for $\delta(t)$.

THEOREM 4.5. *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 $t > 0$:*

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

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

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

$$\|Q_h\chi\| \le Ch^2\|\chi\| \quad \forall\chi \in X_h, \tag{4.16}$$

then the following almost optimal error estimate is valid:

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

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 $p = 0, 1$

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

Consequently, by Lemmas 4.3 and 3.1 and $\dot{H}^1$- and L_2-stability of the operator P_h from Lemma 2.3, we deduce for $q = 0, 1$

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

Now the estimate (4.14) follows by triangle inequality from this and estimate (3.24) 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

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

This estimate and Lemma 4.3 give

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

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

$$\begin{aligned} \|\delta(t)\| \le Ch^{-\epsilon} \int_0^t (t-s)^{\frac{\epsilon}{2}\alpha-1} \|Q_h \partial_t^\alpha u_h(s)\| \, ds &\le Ch^{2-\epsilon} \int_0^t (t-s)^{\frac{\epsilon}{2}\alpha-1} \|\partial_t^\alpha u_h(s)\| \, ds \\ \le Ch^{2-\epsilon} \int_0^t (t-s)^{\frac{\epsilon}{2}\alpha-1} s^{-\alpha} ds \, |u_h(0)| &\le C\epsilon^{-1} h^{2-\epsilon} t^{-\alpha(1-\frac{\epsilon}{2})} \|v\|. \end{aligned}$$

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

Remark 4.3. By interpolation (4.17) is valid also for $0 < q < 1$.

Remark 4.4. The condition (4.16) on the quadrature error operator Q_h is satisfied for symmetric meshes [2, section 5]. In one dimension, the symmetry requirement can be relaxed to almost symmetry [2, section 6]. In case (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 that in the classical parabolic equation [2, Theorem 4.4].

Remark 4.5. We note that we have used 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 convergence rate for the gradient $\nabla(u_h - u)$ 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 [11]. To establish a superconvergent recovery of the gradient, Křížek and Neittaanmäki [11] introduced an operator G_h which postprocesses the gradient of the Ritz projection $R_h u$ of a function u [11, equation (2.2)] with the following properties:

(a) If $u \in H^3(\Omega)$, then [11, Theorem 4.2]

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

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

$$\|G_h(\chi)\| \le C\|\nabla \chi\|. \tag{5.2}$$

The bound (5.2) follows immediately from [11, 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 ∇u by $G_h(R_h u)$ due to the special postprocessing procedure valid for sufficiently smooth u and special meshes.

This result could be used to establish a higher convergence rate for the semidiscrete 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 a strongly uniform triangulation of Ω, i.e., every two adjacent triangles form a parallelogram. Then the following estimate is valid:*

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

Proof. It follows from the fact that u satisfies (1.1), i.e., $\partial_t^\alpha u(t) = \Delta u$, and from Theorem 2.2 (with $\ell = 1$, $p = 1$, and $q = 2$) that

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

Then using the above superconvergent recovery operator G_h of the gradient with the properties (5.1) and (5.2) and the estimate (3.20) for $\theta(t) = R_h u(t) - u_h(t)$, we get

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

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 lumped mass solution $\bar{u}_h$:

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

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

6. Numerical results. Here we present some numerical tests to verify the theoretical error estimates for the Galerkin and lumped mass FEMs.

6.1. Test problems. We consider problem (1.1) on the unit interval $\Omega = (0, 1)$ and perform 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) $v(x) = x\chi_{[0,\frac{1}{2}]} + (1-x)\chi_{(\frac{1}{2},1]}$, where χ refers to the characteristic function.

(c) Nonsmooth initial data, (c1) $v(x) = 1$, (c2) $v(x) = x$, and (c3) $v(x) = \chi_{[0,\frac{1}{2}]}$. For all three examples v is not compatible with the homogeneous Dirichlet boundary condition, $v \notin H_0^1$, but $v \in H^{\frac{1}{2}-\epsilon}$ for any $\epsilon > 0$.

(d) Initial data v that is a Dirac $\delta_{\frac{1}{2}}(x)$-function concentrated at $x = 0.5$. Such case is not covered by our theory.

(e) Variable coefficient case (cf. (3.25)). 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].

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. In the case of constant $k(x)$ and $q(x) = 0$ (cf. section 3.4) 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 $-\bar{\Delta}_h$, defined by (3.1) and (4.4), respectively, satisfy

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

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 for $j = 1, \ldots, N$

$$\lambda_j^h = \bar{\lambda}_j^h/(1 - \tfrac{h^2}{6}\bar{\lambda}_j^h), \quad \bar{\lambda}_j^h = \frac{4}{h^2}\sin^2\frac{\pi j}{2(N+1)}, \quad \text{and} \quad \varphi_j^h(x_k) = \bar{\varphi}_j^h(x_k) = \sqrt{2}\sin(j\pi x_k)$$

for x_k being a mesh point and φ_j^h and $\bar{\varphi}_j^h$ linear over the finite elements. These are used in computing the finite element solutions of the Galerkin and lumped mass methods through their representations (3.3) and (4.5), respectively.

To compute a reference solution we have used also a direct numerical technique [13] by discretizing the time interval, $t_n = n\tau$, $n = 0, 1, \ldots$, with τ being the time step size. Whenever 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 $\tau = 10^{-6}$, so that the error incurred by temporal discretization is negligible and the numerical results are identical with that by the solution presentation.

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 numerical results for the lumped mass FEM, since that for the Galerkin FEM are almost identical.

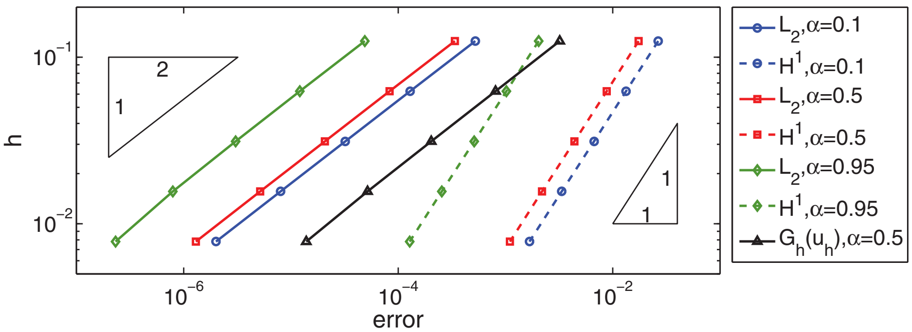


FIG. 6.1. *Numerical results for smooth initial data, example* (a) *with* $\alpha = 0.1, 0.5, 0.95$ *at* $t = 1$.

TABLE 6.1
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

6.2. Smooth case, example (a). In Figure 6.1, we show plots of the numerical results for various α at $t = 1$ 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. We also present the error of the recovered gradient $G_h(u_h)$. Since in one dimension the midpoint 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]. Clearly, the recovered gradient $G_h(u_h)$ exhibits an $O(h^2)$ 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, and 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).

6.3. Intermediate case, 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 6.1 were evaluated at $t = 1$ for $\alpha = 0.5$, where `ratio` in the last column refers to the ratio between the errors as the mesh size h halves. 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 agrees with the theory for the intermediate case.

6.4. Nonsmooth data, examples (c). In Tables 6.2 and 6.3 we present the computational results for problems (c1) and (c2), respectively. 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 6.2 we plot the results shown in Tables 6.2 and 6.3, i.e., for problems (c1) and (c2). These numerical results fully confirm the theoretically predicted rates for the nonsmooth initial data.

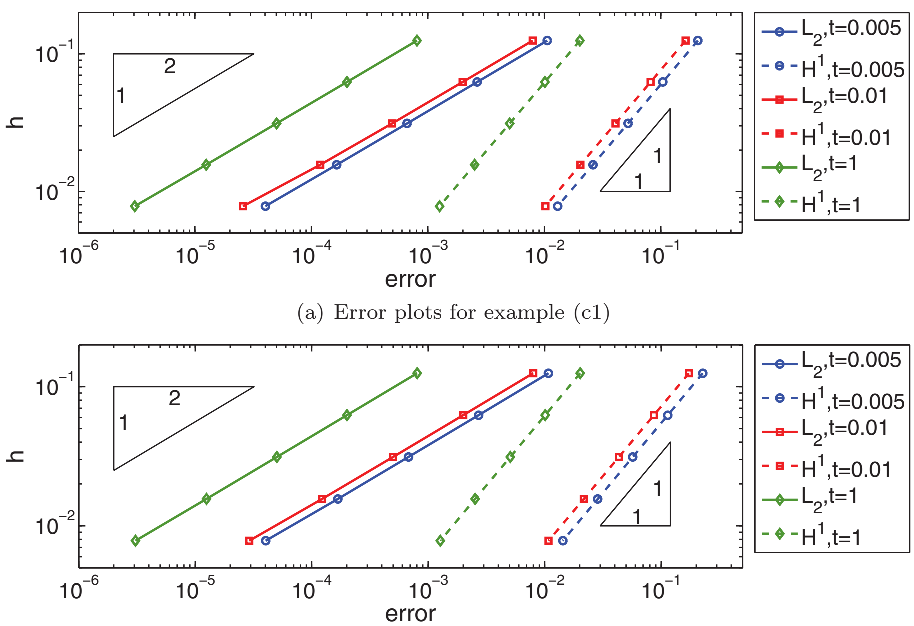
Now we consider the third example of nonsmooth case, the characteristic function of the interval $(0, 0.5)$, namely, $v(x) = \chi_{[0,\frac{1}{2}]}$. Note that if we use the interpolation of v as the initial data for the semidiscrete problem, the L_2-error has only a suboptimal first-order convergence. However, if we choose L_2 projection as is discussed earlier, then the results agree well with our estimates; see Table 6.4. We also discretize this example by the Galerkin method, and the results are presented in Table 6.5. A

TABLE 6.2
Nonsmooth initial data, example (c1) *with* $\alpha = 0.5$.

Time	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 6.3
Nonsmooth initial data, example (c2) *with* $\alpha = 0.5$.

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



(a) Error plots for example (c1)

(b) Error plots for example (c2)

FIG. 6.2. *Numerical results for nonsmooth initial data with* $\alpha = 0.5$.

comparison of Tables 6.4 and 6.5 clearly indicates that the Galerkin method and the lumped 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.

Finally, in Table 6.6 we show the L_2-norm of the error for smooth and nonsmooth data, i.e., examples (a) and (c3), for fixed $h = 2^{-7}$ and t approaching 0. We

TABLE 6.4
Nonsmooth initial data, example (c3) *with* $\alpha = 0.5$.

Time	h	1/8	1/16	1/32	1/64	1/128	Ratio
$t = 0.005$	L_2-norm	8.59e-3	2.16e-3	5.42e-4	1.36e-4	3.39e-5	≈ 4.01
	H^1-norm	2.70e-1	1.29e-1	6.33e-2	3.13e-2	1.55e-2	≈ 2.00
$t = 0.01$	L_2-norm	6.58e-3	1.65e-3	4.13e-4	1.03e-4	2.58e-5	≈ 3.99
	H^1-norm	2.00e-1	9.61e-2	4.71e-2	2.33e-2	1.16e-2	≈ 2.00
$t = 1$	L_2-norm	8.13e-4	2.03e-4	5.08e-5	1.27e-5	3.18e-6	≈ 4.00
	H^1-norm	2.11e-2	1.06e-2	5.22e-3	2.59e-3	1.29e-3	≈ 2.01

TABLE 6.5
Standard Galerkin FEM for nonsmooth initial data, example (c3) *with* $\alpha = 0.5$.

Time	h	1/8	1/16	1/32	1/64	1/128	Ratio
$t = 0.005$	L_2-norm	8.54e-3	2.13e-3	5.33e-4	1.33e-4	3.33e-5	≈ 4.01
	H^1-norm	2.67e-1	1.24e-2	6.18e-2	3.09e-2	1.54e-2	≈ 2.01
$t = 0.01$	L_2-norm	6.51e-3	1.63e-3	4.06e-4	1.02e-4	2.54e-5	≈ 4.00
	H^1-norm	1.84e-1	9.20e-2	4.60e-2	2.30e-2	1.15e-2	≈ 2.00
$t = 1$	L_2-norm	8.00e-4	2.00e-4	5.00e-5	1.25e-5	3.13e-6	≈ 4.00
	H^1-norm	2.05e-2	1.03e-2	5.13e-3	2.56e-3	1.28e-3	≈ 2.00

TABLE 6.6
L_2*-error of the lumped mass FEM with* $\alpha = 0.5$ *and* $h = 2^{-7}$ *for* $t \to 0$ *for smooth, example* (a), *and nonsmooth, example* (c3), *initial data.*

t	1e-2	1e-3	1e-4	1e-5	1e-6	1e-7	Order
Nonsmooth	2.58e-5	6.00e-5	1.32e-4	3.13e-4	7.39e-4	1.74e-3	≈ -0.37
Smooth	1.27e-5	3.07e-5	4.53e-5	5.28e-5	5.65e-5	5.85e-5	≈ -0.02

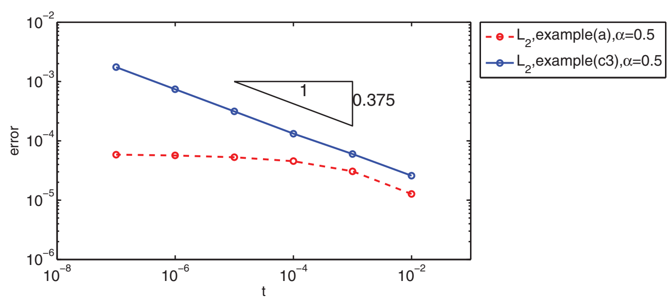


FIG. 6.3. L_2*-error of the lumped mass FEM with* $\alpha = 0.5$ *and* $h = 2^{-7}$ *as* $t \to 0$, *for smooth, example* (a), *and nonsmooth, example* (c3), *initial data.*

observe that in the smooth case the error essentially stays unchanged, whereas in the nonsmooth case it deteriorates as $t \to 0$. In example (c3) the initial data $v \in \dot{H}^{\frac{1}{2}-\epsilon}$ for any $\epsilon > 0$, and it follows from (4.15) and interpolation that the error grows like $O(t^{-\frac{3}{4}\alpha})$ as $t \to 0$. This is fully confirmed in Table 6.6 and Figure 6.3, where, for fixed h, the results are plotted in a log-log scale. Note that the slope agrees well with the one predicted by the theory, i.e., $-\frac{3}{4}\alpha = -0.375$ for $\alpha = 0.5$.

6.5. Initial data a Dirac δ-function, example (d). We note that this case is not covered by the presented theory. Formally, the orthogonal L_2-projection P_h is

TABLE 6.7
Lumped mass FEM with initial data a Dirac δ-function, $\alpha = 0.5$.

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

TABLE 6.8
Numerical results for variable coefficients and nonsmooth 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

not defined for such functions. However, we can regard (v, χ) for $\chi \in X_h \subset H_0^1(\Omega)$ as duality pairing between the spaces $H^{-1}(\Omega)$ 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_j, then we can define $P_h \delta_{\frac{1}{2}}$ appropriately with its finite element expansion given by the jth column of the inverse of the mass matrix. This is taken as the initial data for the semidiscrete problem in our computations. It is observed 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 Table 6.7. It is remarkable that the scheme can actually achieve good convergence rates in L_2- and H^1-norm for such very weak solutions. The theoretical justification of these rates is a subject of our current work.

6.6. Variable coefficients, example (e). Although we do not have an explicit representation of the exact solution, we compare the numerical solution with a reference solution obtained on very fine grids with mesh size $h = 1/512$ and time step size $\tau = 10^{-5}$. The normalized L_2- and H^1-norms of the error are reported in Table 6.8 for $t = 0.01$ and $\alpha = 0.5$. The results confirm the theoretically predicted rates.

In summary, the empirical convergence rates in all numerical experiments confirm the theoretical findings for both smooth and nonsmooth initial data, including the case of the recovered gradient $G_h(u_h)$ discussed in section 5.

Acknowledgment. The authors would like to thank the anonymous reviewers for their constructive comments, which have led to an improvement of the presentation.

REFERENCES

[1] J.-Ph. Bouchaud and A. Georges, *Anomalous diffusion in disordered media: Statistical mechanisms, models and physical applications*, Phys. Rep., 195 (1990), pp. 127–293.
[2] P. Chatzipantelidis, R.D. Lazarov, and V. Thomée, *Some error estimates for the lumped mass finite element method for a parabolic problem*, Math. Comp., 81 (2012), pp. 1–20.
[3] J. Cheng, J. Nakagawa, Y. Yamamoto, and T. Yamazaki, *Uniqueness in an inverse problem for a one-dimensional fractional diffusion equation*, Inverse Problems, 25 (2009), 115002.
[4] L. Debnath, *Recent applications of fractional calculus to science and engineering*, Int. J. Math. Math. Sci., 54 (2003), pp. 3413–3442.
[5] M.M. Djrbashian, *Harmonic Analysis and Boundary Value Problems in the Complex Domain*, Birkhäuser, Basel, 1993.
[6] A. Ern and J.-L. Guermond, *Theory and Practice of Finite Elements*, Springer, New York, 2004.

[7] V.J. Ervin and J.P. Roop, *Variational formulation for the stationary fractional advection dispersion equation*, Numer. Methods Partial Differential Equations, 22 (2006), pp. 558–576.
[8] B. Jin and X. Lu, *Numerical identification of a Robin coefficient in parabolic problems*, Math. Comp., 81 (2012), pp. 1369–1398.
[9] Y.L. Keung and J. Zou, *Numerical identifications of parameters in parabolic systems*, Inverse Problems, 14 (1998), pp. 83–100.
[10] A.A. Kilbas, H.M. Srivastava, and J.J. Trujillo, *Theory and Applications of Fractional Differential Equations*, Elsevier, Amsterdam, 2006.
[11] M. Křížek and P. Neittaanmäki, *On a global superconvergence of the gradient of linear triangular elements*, J. Comput. Appl. Math., 18 (1987), pp. 221–233.
[12] T.A.M. Langlands and B.I. Henry, *The accuracy and stability of an implicit solution method for the fractional diffusion equation*, J. Comput. Phys., 205 (2005), pp. 719–736.
[13] Y. Lin and C. Xu, *Finite difference/spectral approximations for the time-fractional diffusion equation*, J. Comput. Phys., 225 (2007), pp. 1533–1552.
[14] X. Li and C. Xu, *A space-time spectral method for the time fractional diffusion equation*, SIAM J. Numer. Anal., 47 (2009), pp. 2108–2131.
[15] W. McLean and V. Thomée, *Time discretization of an evolution equation via Laplace transforms*, IMA J. Numer. Anal., 24 (2004), pp. 439–463.
[16] 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 (2010), pp. 208–230.
[17] W. McLean and V. Thomée, *Numerical solution via Laplace transforms of a fractional order evolution equation*, J. Integral Equations Appl., 22 (2010), pp. 57–94.
[18] W. McLean, *Regularity of solutions to a time-fractional diffusion equation*, ANZIAM J., 52 (2010), pp. 123–138.
[19] M.M. Meerschaert, H.-P. Scheffler, and C. Tadjeran, *Finite difference methods for two-dimensional fractional dispersion equation*, J. Comput. Phys., 211 (2006), pp. 249–261.
[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 (2011), pp. 719–739.
[21] R.R. Nigmatulin, *The realization of the generalized transfer equation in a medium with fractal geometry*, Phys. Stat. Sol. B, 133 (1986), pp. 425–430.
[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 (2011), pp. 426–447.
[24] H. Seybold and R. Hilfer, *Numerical algorithm for calculating the generalized Mittag-Leffler function*, SIAM J. Numer. Anal., 47 (2008/09), pp. 69–88.
[25] V. Thomée, *Galerkin Finite Element Methods for Parabolic Problems*, Springer, Berlin, 1997.
[26] L.B. Wahlbin, *Superconvergence in Galerkin Finite Element Methods*, Springer, Berlin, 1995.
[27] S.B. Yuste and L. Acedo, *An explicit finite difference method and a new Neumann-type stability analysis for fractional diffusion equations*, SIAM J. Numer. Anal., 42 (2005), pp. 1862–1874.