Journal of Computational Physics 281 (2015) 825–843

Contents lists available at ScienceDirect

Journal of Computational Physics

www.elsevier.com/locate/jcp

The Galerkin finite element method for a multi-term time-fractional diffusion equation

Bangti Jin [a,*], Raytcho Lazarov [b], Yikan Liu [c], Zhi Zhou [b]

[a] Department of Computer Science, University College London, Gower Street, London, WC1E 6BT, UK
[b] Department of Mathematics, Texas A&M University, College Station, TX 77843-3368, USA
[c] Graduate School of Mathematical Sciences, The University of Tokyo, 3-8-1 Komaba, Meguro, Tokyo 153, Japan

ARTICLE INFO

Article history:
Received 27 January 2014
Received in revised form 22 October 2014
Accepted 26 October 2014
Available online 30 October 2014

Keywords:
Multi-term time-fractional diffusion equation
Finite element method
Error estimate
Semidiscrete scheme
Caputo derivative

ABSTRACT

We consider the initial/boundary value problem for a diffusion equation involving multiple time-fractional derivatives on a bounded convex polyhedral domain. We analyze a space semidiscrete scheme based on the standard Galerkin finite element method using continuous piecewise linear functions. Nearly optimal error estimates for both cases of initial data and inhomogeneous term are derived, which cover both smooth and nonsmooth data. Further we develop a fully discrete scheme based on a finite difference discretization of the time-fractional derivatives, and discuss its stability and error estimate. Extensive numerical experiments for one- and two-dimensional problems confirm the theoretical convergence rates.
© 2014 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/3.0/).

1. Introduction

We consider the following initial/boundary value problem for a multi-term time fractional diffusion equation in $u(x,t)$:

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

where Ω denotes a bounded convex polygonal domain in $\mathbb{R}^d$ $(d = 1, 2, 3)$ with a boundary $\partial\Omega$, f is the source term, and the initial data v is a given function on Ω and $T > 0$ is a fixed value. Here the differential operator $P(\partial_t)$ is defined by

$$
P(\partial_t) = \partial_t^{\alpha} + \sum_{i=1}^{m} b_i \partial_t^{\alpha_i},
$$

where $0 < \alpha_m \le \ldots \le \alpha_1 < \alpha < 1$ are the orders of the fractional derivatives, $b_i > 0$, $i = 1, 2, \ldots, m$, with the left-sided Caputo fractional derivative $\partial_t^{\beta} u$, $0 < \beta < 1$, being defined by (cf. [17, p. 91])

* Corresponding author.
E-mail addresses: bangti.jin@gmail.com (B. Jin), lazarov@math.tamu.edu (R. Lazarov), ykliu@ms.u-tokyo.ac.jp (Y. Liu), zzhou@math.tamu.edu (Z. Zhou).

http://dx.doi.org/10.1016/j.jcp.2014.10.051
0021-9991/© 2014 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/3.0/).

$$\partial_t^\beta u(t) = \frac{1}{\Gamma(1-\beta)} \int_0^t (t-s)^{-\beta} u'(s)ds, \tag{1.2}$$

where $\Gamma(\cdot)$ denotes the Gamma function.

In the case of $m = 0$, the model (1.1) reduces to its single-term counterpart

$$\partial_t^\alpha u - \Delta u = f \quad \text{in } \Omega \times (0, T]. \tag{1.3}$$

This model has been studied extensively from different aspects due to its extraordinary capability of modeling anomalous diffusion phenomena in highly heterogeneous aquifers and complex viscoelastic materials [1,31]. It is the fractional analogue of the classical diffusion equation: with $\alpha = 1$, it recovers the latter, and thus inherits some of its analytical properties. However, it differs considerably from the latter in the sense that, due to the presence of the nonlocal fractional derivative term, it has limited smoothing property in space and slow asymptotic decay in time [32], which in turn also impacts related numerical analysis [12] and inverse problems [14,32].

The model (1.1) was developed to improve the modeling accuracy of the single-term model (1.3) for describing anomalous diffusion. For example, in [33], a two-term fractional-order diffusion model was proposed for the total concentration in solute transport, in order to distinguish explicitly the mobile and immobile status of the solute using fractional dynamics. The kinetic equation with two fractional derivatives of different orders appears also quite naturally when describing subdiffusive motion in velocity fields [28]; see also [16] for discussions on the model for wave-type phenomena.

There are very few mathematical studies on the model (1.1). Luchko [25] established a maximum principle for problem (1.1), and constructed a generalized solution for the case $f \equiv 0$ using the multinomial Mittag-Leffler function. Jiang et al. [9] derived formal analytical solutions for the diffusion equation with fractional derivatives in both time and space. Li and Yamamoto [22] established the existence, uniqueness, and the Hölder regularity of the solution using a fixed point argument for problem (1.1) with variable coefficients $\{b_i\}$. Very recently, Li et al. [21] showed the uniqueness and continuous dependence of the solution on the initial value v and the source term f, by exploiting refined properties of the multinomial Mittag-Leffler function.

The potential applications of the model (1.1) motivate the design and analysis of numerical schemes that have optimal (with respect to data regularity) convergence rates. Such schemes are especially valuable for problems where the solution has low regularity. The case $m = 0$, i.e., the single-term model (1.3), has been extensively studied, and stability and error estimates were provided; see [23,34,37] for the finite difference method, [19,20,36] for the spectral method, [10–12,18,27,29,30] for the finite element method, and [4,7] for meshfree methods based on radial basis functions, to name a few. In particular, in [10–12], the authors established almost optimal error estimates with respect to the regularity of the initial data v and the right hand side f for a semidiscrete Galerkin scheme. These studies include the interesting case of very weak data, i.e., $v \in \dot H^q(\Omega)$ and $f \in L^\infty(0, T; \dot H^q(\Omega))$ for $-1 < q \le 0$.

Numerical methods for the multi-term ordinary differential equation were considered in [6,15]. In [38], a scheme based on the finite element method in space and a specialized finite difference method in time was proposed for (1.1), and error estimates were derived. We also refer to [24] for a numerical scheme based on a fractional predictor–corrector method for the multi-term time fractional wave-diffusion equation. The error analysis in these works is done under the assumption that the solution is sufficiently smooth and therefore it excludes the case of low regularity solutions. This is the main goal of the present study. However, the derivation of optimal with respect to the regularity error estimates requires additional analysis of the properties of problem (1.1), e.g., stability, asymptotic behavior for $t \to 0^+$. Relevant results of this type have recently been obtained in [21], which, however, are not enough for the analysis of the semidiscrete Galerkin scheme, and hence in Section 2, we make the necessary extensions.

Now we describe the semidiscrete Galerkin scheme. Let $\{\mathcal{T}_h\}_{0<h<1}$ be a family of shape regular and quasi-uniform partitions of the domain Ω into d-simplexes, called finite elements, with a maximum diameter h. The approximate solution u_h is sought in the finite element space X_h of continuous piecewise linear functions over the triangulation $\mathcal{T}_h$

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

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

$$\big(P(\partial_t)u_h, \chi\big) + a(u_h, \chi) = (f, \chi), \quad \forall \chi \in X_h,\ T \ge t > 0,\ u_h(0) = v_h, \tag{1.4}$$

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 scheme (1.4) for the case of initial data $v \in \dot H^q(\Omega)$, $-1 < q \le 2$, and right hand side $f \in L^\infty(0, T; \dot H^q(\Omega))$, $-1 < q < 1$. The case of nonsmooth data, i.e., $-1 < q \le 0$, is very common in inverse problems and optimal control [14,32]; see also [5,13] for the parabolic counterpart.

The goal of this work is to develop a numerical scheme based on the finite element approximation for the model (1.1), and provide a complete error analysis. We derive error estimates optimal with respect to the data regularity for the semidiscrete scheme, and a convergence rate $O(h^2 + \tau^{2-\alpha})$ for the fully discrete scheme in case of a smooth solution. Specifically, our essential contributions are as follows. First, we obtain an improved regularity result for the inhomogeneous problem, by allowing less regular source term, cf. Theorem 2.3. This is achieved by exploiting the complete monotonicity of the multinomial Mittag-Leffler function, cf. Lemma 2.4. Second, we derive nearly optimal error estimates for a semidiscrete Galerkin scheme for both homogeneous and inhomogeneous problems, cf. Theorems 3.1–3.4, which cover both smooth and nonsmooth data. Third, we develop a fully discrete scheme based on a finite difference method in time, and establish its stability and error estimates, cf. Theorem 4.1. We note that the derived error estimate for the fully discrete scheme holds only for smooth solutions.

The rest of the paper is organized as follows. In Section 2, we recall the solution theory for the model (1.1) for both homogeneous and inhomogeneous problems, using properties of the multinomial Mittag-Leffler function. The readers not interested in the analysis may proceed directly to Section 3. Almost optimal error estimates for their semidiscrete Galerkin finite element approximations are given in Section 3. Then a fully discrete scheme based on a finite difference approximation of the Caputo fractional derivatives is given in Section 4, and an error analysis is also provided. Finally, extensive numerical experiments are presented to illustrate the accuracy and efficiency of the Galerkin scheme, and to verify the convergence theory. Throughout, we denote by C a generic constant, which may differ at different occurrences, but always independent of the mesh size h and time step size τ.

2. Solution theory

In this part, we recall the solution theory for problem (1.1). We shall describe the solution representation using the multinomial Mittag-Leffler function, and derive optimal solution regularity for the homogeneous and inhomogeneous problems.

2.1. Multinomial Mittag-Leffler function

First we recall the multinomial Mittag-Leffler function, introduced in [8]. For $0 < \beta < 2$, $0 < \beta_i < 1$ and $z_i \in \mathbb{C}$, $i = 1, ..., m$, the multinomial Mittag-Leffler function $E_{(\beta_1,...,\beta_m),\beta}(z_1, ..., z_m)$ is defined by

$$E_{(\beta_1,...,\beta_m),\beta}(z_1, ..., z_m) = \sum_{k=0}^{\infty} \sum_{\substack{l_1+...+l_m=k \\ l_1\geq 0,...,l_m\geq 0}} (k; l_1, ..., l_m) \frac{\prod_{i=1}^m z_i^{l_i}}{\Gamma(\beta + \sum_{i=1}^m \beta_i l_i)},$$

where the notation $(k; l_1, ..., l_m)$, $l_1 \geq 0$, $i = 1, \ldots, m$, denotes the multinomial coefficient, i.e.,

$$(k; l_1, ..., l_m) = \frac{k!}{l_1!...l_m!} \quad \text{with } k = \sum_{i=1}^m l_i.$$

It generalizes the exponential function e^z: with $m = 1$ and $\beta = \beta_1 = 1$, it reproduces the exponential function e^z. It appears in the solution representation of problem (1.1), cf. (2.4) below. We shall need the following two important lemmas on the function $E_{(\beta_1,...,\beta_m),\beta}(z_1, ..., z_m)$, recently obtained in [21].

Lemma 2.1. *Let $0 < \beta < 2$, $0 < \beta_i < 1$, $\beta_1 > \max\{\beta_2, ..., \beta_m\}$ and $\frac{\beta_1\pi}{2} < \mu < \beta_1\pi$. Assume that there is $K > 0$ such that $-K \leq z_i < 0$, $i = 2, ..., m$. Then there exists a constant $C = C(\beta_1, ..., \beta_m, \beta, K, \mu) > 0$ such that*

$$E_{(\beta_1,...,\beta_m),\beta}(z_1, ..., z_m) \leq \frac{C}{1 + |z_1|}, \qquad \mu \leq \left|\arg(z_1)\right| \leq \pi.$$

Lemma 2.2. *Let $0 < \beta < 2$, $0 < \beta_i < 1$ and $z_i \in \mathbb{C}$, $i = 1, ..., m$. Then we have*

$$\frac{1}{\Gamma(\beta_0)} + \sum_{i=1}^m z_i E_{(\beta_1,...,\beta_m),\beta_0+\beta_i}(z_1, ..., z_m) = E_{(\beta_1,...,\beta_m),\beta_0}(z_1, ..., z_m).$$

2.2. Solution representation

For $s \geq -1$, we denote by $\dot{H}^s(\Omega) \subset H^{-1}(\Omega)$ the Hilbert space induced by the norm:

$$\|v\|^2_{\dot{H}^s(\Omega)} = \sum_{j=1}^{\infty} \lambda_j^s \langle v, \varphi_j\rangle^2$$

with $\{\lambda_j\}_{j=1}^\infty$ and $\{\varphi_j\}_{j=1}^\infty$ being respectively the eigenvalues and the $L^2(\Omega)$-orthonormal eigenfunctions of the Laplace operator $-\Delta$ on the domain Ω with a homogeneous Dirichlet boundary condition. Then $\{\varphi_j\}_{j=1}^\infty$ and $\{\lambda_j^{1/2}\varphi_j\}_{j=1}^\infty$, form an orthonormal basis in $L^2(\Omega)$ and $H^{-1}(\Omega)$, respectively. Further, $\|v\|_{\dot{H}^0(\Omega)} = \|v\|_{L^2(\Omega)} = (v, v)^{1/2}$ is the norm in $L^2(\Omega)$ and $\|v\|_{\dot{H}^{-1}(\Omega)} = \|v\|_{H^{-1}(\Omega)}$ is the norm in $H^{-1}(\Omega)$. It is easy to verify that $\|v\|_{\dot{H}^1(\Omega)} = \|\nabla v\|_{L^2(\Omega)}$ is also the norm in $H_0^1(\Omega)$

and $\|v\|_{\dot H^2(\Omega)} = \|\Delta v\|_{L^2(\Omega)}$ is equivalent to the norm in $H^2(\Omega)\cap H_0^1(\Omega)$ [35, Lemma 3.1]. Note that $\dot H^s(\Omega)$, $s \ge -1$ form a Hilbert scale of interpolation spaces. Hence, we denote $\|\cdot\|_{H^s(\Omega)}$ to be the norm on the interpolation scale between $H_0^1(\Omega)$ and $L^2(\Omega)$ for $s \in [0,1]$ and $\|\cdot\|_{H^s(\Omega)}$ to be the norm on the interpolation scale between $L^2(\Omega)$ and $H^{-1}(\Omega)$ for $s\in[-1,0]$. Then, $\|\cdot\|_{H^s(\Omega)}$ and $\|\cdot\|_{\dot H^s(\Omega)}$ are equivalent for $s \in [-1,1]$. Further, for a Banach space B and any $r \ge 1$, we define the space

$$L^r(0,T;B) = \{u(t)\in B \text{ for a.e. } t\in(0,T) \text{ and } \|u\|_{L^r(0,T;B)} < \infty\},$$

and the norm $\|\cdot\|_{L^r(0,T;B)}$ is defined by

$$\|u\|_{L^r(0,T;B)} = \begin{cases} (\int_0^T \|u(t)\|_B^r dt)^{1/r}, & r\in[1,\infty), \\ \operatorname{ess\,sup}_{t\in(0,T)}\|u(t)\|_B, & r=\infty. \end{cases}$$

Upon denoting $\vec\alpha = (\alpha, \alpha-\alpha_1, ..., \alpha-\alpha_m)$, we introduce the following solution operator

$$E(t)v = \sum_{j=1}^{\infty}\big(1-\lambda_j t^\alpha E_{\vec\alpha,1+\alpha}\big(-\lambda_j t^\alpha, -b_1 t^{\alpha-\alpha_1}, ..., -b_m t^{\alpha-\alpha_m}\big)\big)(v,\varphi_j)\varphi_j. \tag{2.1}$$

This operator is motivated by a separation of variables [25,26]. Then for problem (1.1) with a homogeneous right hand side, i.e., $f\equiv 0$, we have $u(x,t)=E(t)v$. However, the representation (2.1) is not always very convenient for analyzing its smoothing property. We derive an alternative representation of the solution operator E using Lemma 2.2:

$$\begin{aligned} E(t)v &= \sum_{j=1}^{\infty} E_{\vec\alpha,1}\big(-\lambda_j t^\alpha, -b_1 t^{\alpha-\alpha_1}, ..., -b_m t^{\alpha-\alpha_m}\big)(v,\varphi_j)\varphi_j \\ &\quad + \sum_{i=1}^{m} b_i t^{\alpha-\alpha_i}\sum_{j=1}^{\infty} E_{\vec\alpha,1+\alpha-\alpha_i}\big(-\lambda_j t^\alpha, -b_1 t^{\alpha-\alpha_1}, ..., -b_m t^{\alpha-\alpha_m}\big)(v,\varphi_j)\varphi_j. \end{aligned} \tag{2.2}$$

Besides, we define the following operator $\bar E$ for $\chi\in L^2(\Omega)$ by

$$\bar E(t)\chi = \sum_{j=1}^{\infty} t^{\alpha-1} E_{\vec\alpha,\alpha}\big(-\lambda_j t^\alpha, -b_1 t^{\alpha-\alpha_1}, ..., -b_m t^{\alpha-\alpha_m}\big)(\chi,\varphi_j)\varphi_j. \tag{2.3}$$

The operators $E(t)$ and $\bar E(t)$ can be used to represent the solution u of (1.1) as:

$$u(t) = E(t)v + \int_0^t \bar E(t-s) f(s) ds. \tag{2.4}$$

The operator $\bar E$ has the following smoothing property.

Lemma 2.3. *For any $t>0$ and $\chi\in\dot H^q(\Omega)$, $q\in(-1,2]$, there holds for $0\le p-q\le 2$*

$$\|\bar E(t)\chi\|_{\dot H^p(\Omega)} \le C t^{-1+\alpha(1+(q-p)/2)}\|\chi\|_{\dot H^q(\Omega)}.$$

Proof. The definition of the operator $\bar E$ in (2.3) and Lemma 2.1 yield

$$\begin{aligned} \|\bar E(t)\chi\|_{\dot H^p(\Omega)}^2 &= t^{-2+(2+q-p)\alpha}\sum_{j=1}^{\infty}(\lambda_j t^\alpha)^{p-q}\big|E_{\vec\alpha,\alpha}\big(-\lambda_j t^\alpha, -b_1 t^{\alpha-\alpha_1}, ..., -b_m t^{\alpha-\alpha_m}\big)\big|^2 \lambda_j^q \big|(\chi,\varphi_j)\big|^2 \\ &\le C t^{-2+(2+q-p)\alpha}\sum_{j=1}^{\infty}\frac{(\lambda_j t^\alpha)^{p-q}}{(1+\lambda_j t^\alpha)^2}\lambda_j^q\big|(\chi,\varphi_j)\big|^2 \\ &\le C t^{-2+(2+q-p)\alpha}\sum_{j=1}^{\infty}\lambda_j^q\big|(\chi,\varphi_j)\big|^2 \le C t^{-2+(2+q-p)\alpha}\|\chi\|_{\dot H^q(\Omega)}, \end{aligned}$$

where the last line follows by the inequality $\sup_{j\in\mathbb N}\frac{(\lambda_j t^\alpha)^{p-q}}{(1+\lambda_j t^\alpha)^2}\le C$, for $0\le p-q\le 2$. □

2.3. Solution regularity

First we recall known regularity results. In [22], Li and Yamamoto showed that in the case of variable coefficients $\{b_i(x)\}$, there exists a unique mild solution $u \in C((0,T]; \dot{H}^\gamma(\Omega)) \cap C([0,T]; L^2(\Omega))$ and $u \in C([0,T]; \dot{H}^\gamma(\Omega)) \cap L^\infty(0,T; \dot{H}^2(\Omega))$ when $v \in L^2(\Omega)$, $f = 0$ and $v = 0$, $f \in L^\infty(0,T]; L^2(\Omega))$, respectively, with $\gamma \in [0,1)$. These results were recently refined in [21] for the case of constant coefficients, i.e., problem (1.1). In particular, it was shown that for $v \in \dot{H}^q(\Omega)$, $0 \le q \le 1$, and $f = 0$, $u \in L^{1/(1-q/2)}(0,T; H^2(\Omega) \cap H_0^1(\Omega))$; and for $v = 0$ and $f \in L^r(0,T; \dot{H}^q(\Omega))$, $0 \le q \le 2$, $r \ge 1$, $u \in L^r(0,T; \dot{H}^{q+2-\gamma}(\Omega))$ for some $\gamma \in (0,1]$. Here we follow the approach in [21], and extend these results to a slightly more general setting of $v \in \dot{H}^q(\Omega)$, $-1 < q \le 2$, and $f \in L^2(0,T; \dot{H}^q(\Omega))$, $-1 < q \le 1$. The nonsmooth case, i.e., $-1 < q \le 0$, arises commonly in related inverse problems and optimal control problems.

We shall derive the solution regularity to the homogeneous problem, i.e., $f \equiv 0$, and the inhomogeneous problem, i.e., $v \equiv 0$, separately. These results will be essential for the error analysis of the space semidiscrete Galerkin scheme (1.4) in Section 3. First we consider the homogeneous problem with initial data $v \in \dot{H}^q(\Omega)$, $-1 < q \le 2$.

Theorem 2.1. *Let $u(t) = E(t)v$ be the solution to problem (1.1) with $f \equiv 0$ and $v \in \dot{H}^q(\Omega)$, $q \in (-1,2]$. Then there holds*

$$\|P(\partial_t)^\ell u(t)\|_{\dot{H}^p(\Omega)} \le Ct^{-\alpha(\ell+(p-q)/2)}\|v\|_{\dot{H}^q(\Omega)}, \quad t > 0,$$

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

Proof. We show that (2.2) represents indeed the weak solution to problem (1.1) with $f \equiv 0$ and further it satisfies the desired estimate. We first discuss the case $\ell = 0$. By Lemma 2.1 and (2.2) we have for $0 \le p - q \le 2$

$$\begin{aligned}\|E(t)v\|^2_{\dot{H}^p(\Omega)} &= \sum_{j=1}^\infty \lambda_j^p \Bigg(E_{\vec{\alpha},1}\big(-\lambda_j t^\alpha, -b_1 t^{\alpha-\alpha_1}, \dots, -b_m t^{\alpha-\alpha_m}\big) \\ &\qquad + \sum_{i=1}^m b_i t^{\alpha-\alpha_i} E_{\vec{\alpha},1+\alpha-\alpha_i}\big(-\lambda_j t^\alpha, -b_1 t^{\alpha-\alpha_1}, \dots, -b_m t^{\alpha-\alpha_m}\big)\Bigg)^2 (v,\varphi_j)^2 \\ &\le Ct^{-(p-q)\alpha} \sum_{j=1}^\infty \frac{(\lambda_j t^\alpha)^{p-q}}{(1+\lambda_j t^\alpha)^2} \lambda_j^q \big|(v,\varphi_j)\big|^2 \le Ct^{-(p-q)\alpha}\|v\|^2_{\dot{H}^q(\Omega)},\end{aligned}$$

where the last line follows from the inequality $\sup_{j\in\mathbb{N}} \frac{(\lambda_j t^\alpha)^{p-q}}{(1+\lambda_j t^\alpha)^2} \le C$ for $0 \le p - q \le 2$. The estimate for the case $\ell = 1$ follows from the identity $\|P(\partial_t)E(t)v\|_{\dot{H}^p(\Omega)} = \|E(t)v\|_{\dot{H}^{p+2}(\Omega)}$. It remains to show that (2.2) satisfies also the initial condition in the sense that $\lim_{t\to 0^+} \|E(t)v - v\|_{\dot{H}^q(\Omega)} = 0$. By identity (2.1) and Lemma 2.1, we deduce

$$\|E(t)v - v\|^2_{\dot{H}^q(\Omega)} = \sum_{j=1}^\infty \lambda_j^2 t^{2\alpha} \big|E_{\vec{\alpha},1+\alpha}\big(-\lambda_j t^\alpha, -b_1 t^{\alpha-\alpha_1}, \dots, -b_m t^{\alpha-\alpha_m}\big)\big|^2 \lambda_j^q \big|(v,\varphi_j)\big|^2 \le C\|v\|^2_{\dot{H}^q(\Omega)} < \infty.$$

Using Lemma 2.2, we rewrite the term $\lambda_j t^\alpha E_{\vec{\alpha},1+\alpha}(-\lambda_j t^\alpha, -b_1 t^{\alpha-\alpha_1}, \dots, -b_m t^{\alpha-\alpha_m})$ as

$$\begin{aligned}\lambda_j t^\alpha E_{\vec{\alpha},1+\alpha}\big(-\lambda_j t^\alpha, -b_1 t^{\alpha-\alpha_1}, \dots, -b_m t^{\alpha-\alpha_m}\big) &= \big(1 - E_{\vec{\alpha},1}\big(-\lambda_j t^\alpha, -b_1 t^{\alpha-\alpha_1}, \dots, -b_m t^{\alpha-\alpha_m}\big)\big) \\ &\quad - \sum_{i=1}^m b_i t^{\alpha-\alpha_i} E_{\vec{\alpha},1+\alpha-\alpha_i}\big(-\lambda_j t^\alpha, -b_1 t^{\alpha-\alpha_1}, \dots, -b_m t^{\alpha-\alpha_m}\big).\end{aligned}$$

Upon noting the identity $\lim_{t\to 0^+}(1 - E_{\vec{\alpha},1}(-\lambda_j t^\alpha, -b_1 t^{\alpha-\alpha_1}, \dots, -b_m t^{\alpha-\alpha_m})) = 0$, and the boundedness of $E_{\vec{\alpha},1+\alpha-\alpha_i}(-\lambda_j t^\alpha, -b_1 t^{\alpha-\alpha_1}, \dots, -b_m t^{\alpha-\alpha_m})$ from Lemma 2.1, we deduce that for all j

$$\lim_{t\to 0^+} \lambda_j t^\alpha E_{\vec{\alpha},1+\alpha}\big(-\lambda_j t^\alpha, -b_1 t^{\alpha-\alpha_1}, \dots, -b_m t^{\alpha-\alpha_m}\big) = 0.$$

Hence, the desired assertion follows by Lebesgue's dominated convergence theorem. $\square$

Now we turn to the inhomogeneous problem with a nonsmooth right hand side, i.e., $f \in L^\infty(0,T; \dot{H}^q(\Omega))$, $-1 < q \le 1$, and a zero initial data $v \equiv 0$.

Theorem 2.2. *For $f \in L^\infty(0,T; \dot{H}^q(\Omega))$, $-1 < q \le 1$, and $v \equiv 0$, the representation (2.4) belongs to $L^\infty(0,T; \dot{H}^{q+2-\epsilon}(\Omega))$ for any $\epsilon \in (0,1/2)$ and satisfies*

$$\|u(\cdot,t)\|_{\dot{H}^{q+2-\epsilon}(\Omega)} \le C\epsilon^{-1} t^{\epsilon\alpha/2}\|f\|_{L^\infty(0,t;\dot{H}^q(\Omega))}. \tag{2.5}$$

Hence, it is a solution to problem (1.1) with a homogeneous initial data $v = 0$.

Proof. By construction, it satisfies the governing equation. By Lemma 2.3, we have

$$\begin{aligned}\|u(\cdot,t)\|_{\dot H^{q+2-\epsilon}(\Omega)} &= \left\| \int_0^t \bar E(t-s) f(s) ds \right\|_{\dot H^{q+2-\epsilon}(\Omega)} \le \int_0^t \|\bar E(t-s) f(s)\|_{\dot H^{q+2-\epsilon}(\Omega)} ds \\ &\le C \int_0^t (t-s)^{\epsilon\alpha/2-1} \|f(s)\|_{\dot H^q(\Omega)} ds \le C\epsilon^{-1} t^{\epsilon\alpha/2} \|f\|_{L^\infty(0,t;\dot H^q(\Omega))}\end{aligned}$$

which shows the desired estimate. Further, it satisfies the initial condition $u(0)=0$, i.e., for any $\epsilon>0$, $\lim_{t\to 0^+} \|u(\cdot,t)\|_{\dot H^{q+2-\epsilon}(\Omega)} = 0$, and thus it is indeed a solution of (1.1). □

Next we extend Theorem 2.2 to allow less regular right hand sides $f \in L^2(0,T;\dot H^q(\Omega))$, $-1<q\le 1$. Then the function $u(x,t)$ satisfies also the differential equation as an element in the space $L^2(0,T;\dot H^{q+2}(\Omega))$. However, it may not satisfy the homogeneous initial condition $u(x,0)=0$. In Remark 2.1 below, we argue that a weaker class of source term that produces a legitimate weak solution of (1.1) is $f \in L^r(0,T;\dot H^q(\Omega))$ with $r>1/\alpha$ and $-1<q\le 1$. Obviously, for $1/2<\alpha<1$, it does give a solution $u(x,t)\in L^2(0,T;\dot H^{q+2}(\Omega))$. To this end, we introduce the shorthand notation

$$\bar E^j_{\vec\alpha}(t) = t^{\alpha-1} E_{\vec\alpha,\alpha}\left(-\lambda_j t^\alpha, -b_1 t^{\alpha-\alpha_1}, \dots, -b_m t^{\alpha-\alpha_m}\right).$$

The function $\bar E^j_{\vec\alpha}(t)$ is completely monotone [3].

Lemma 2.4. *The function $\bar E^j_{\vec\alpha}(t)$ for $j\in\mathbb N$ has the following properties:*

$$\bar E^j_{\vec\alpha}(t) \quad \textit{is completely monotone} \quad \textit{and} \quad \int_0^T \left|\bar E^j_{\vec\alpha}(t)\right| dt < \frac{1}{\lambda_j}.$$

Theorem 2.3. *For $f\in L^2(0,T;\dot H^q(\Omega))$, $-1<q\le 1$, the representation (2.4) belongs to $L^2(0,T;\dot H^{q+2}(\Omega))$ and satisfies the a priori estimate*

$$\|u\|_{L^2(0,t;\dot H^{q+2}(\Omega))} + \left\|P(\partial_t)u\right\|_{L^2(0,t;\dot H^q(\Omega))} \le C\|f\|_{L^2(0,t;\dot H^q(\Omega))}. \tag{2.6}$$

Proof. By Young's inequality for the convolution $\|k*f\|_{L^p} \le \|k\|_{L_1}\|f\|_{L^p}$, $k\in L^1$, $f\in L^p$, $p\ge 1$, and Lemma 2.4, we deduce

$$\left\| \int_0^t \bar E^n_{\vec\alpha}(t-\tau) f_n(\tau)\, d\tau \right\|^2_{L^2(0,T)} \le \left(\int_0^T \left|\bar E^n_{\vec\alpha}(t)\right| dt \right)^2 \left(\int_0^T \left|f_n(t)\right|^2 dt \right) \le \frac{1}{\lambda_n^2} \int_0^T \left|f_n(t)\right|^2 dt.$$

Hence,

$$\|u\|^2_{L^2(0,T;\dot H^{q+2}(\Omega))} = \sum_{n=1}^\infty \lambda_n^{q+2} \left\| \int_0^t \bar E^n_{\vec\alpha}(t-\tau) f_n(\tau)\, d\tau \right\|^2_{L^2(0,T)} \le \sum_{n=1}^\infty \lambda_n^q \int_0^T \left|f_n(t)\right|^2 dt = \|f\|^2_{L^2(0,T;\dot H^q(\Omega))}.$$

The estimate on $\|P(\partial_t)u\|_{L^2(0,t;\dot H^q(\Omega))}$ follows analogously. This completes the proof. □

Remark 2.1. The condition $f\in L^\infty(0,T;\dot H^q(\Omega))$ in Theorem 2.2 can be weakened to $f\in L^r(0,T;\dot H^q(\Omega))$ with $r>1/\alpha$. This follows from Lemma 2.3 and Hölder's inequality with r', $1/r'+1/r=1$

$$\|u(\cdot,t)\|_{\dot H^q(\Omega)} \le \int_0^t \|\bar E(t-s)f(s)\|_{\dot H^q(\Omega)} ds \le C\int_0^t (t-s)^{\alpha-1}\|f(s)\|_{\dot H^q(\Omega)} ds \le C\left(\frac{t^{1+r'(\alpha-1)}}{1+r'(\alpha-1)}\right)^{1/r'} \|f\|_{L^r(0,t;\dot H^q(\Omega))},$$

where $1+r'(\alpha-1)>0$ by the condition $r>1/\alpha$. It follows from this that the initial condition $u(\cdot,0)=0$ holds in the following sense: $\lim_{t\to 0^+}\|u(\cdot,t)\|_{\dot H^q(\Omega)}=0$. Hence for any $\alpha\in(1/2,1)$ the representation (2.4) remains a legitimate solution under the weaker condition $f\in L^2(0,T;\dot H^q(\Omega))$.

3. Error estimates for semidiscrete Galerkin scheme

Now we derive and analyze the space semidiscrete Galerkin FEM scheme (1.4). First we describe the semidiscrete scheme, and then derive almost optimal error estimates for the homogeneous and inhomogeneous problems separately. In the analysis we essentially use the technique developed in [12] and improved in [10,11].

3.1. Semidiscrete scheme

To describe the scheme, we need the $L^2(\Omega)$ projection $P_h : L^2(\Omega) \to X_h$ and Ritz projection $R_h : H_0^1(\Omega) \to X_h$, respectively, defined by

$$(P_h\psi, \chi) = (\psi, \chi) \quad \forall \chi \in X_h,$$
$$(\nabla R_h\psi, \nabla\chi) = (\nabla\psi, \nabla\chi) \quad \forall \chi \in X_h.$$

The operators R_h and P_h satisfy the following approximation property [35].

Lemma 3.1. *For any $\psi \in \dot{H}^q(\Omega)$, $1 \le q \le 2$, the operator R_h satisfies:*

$$\|R_h\psi - \psi\|_{L^2(\Omega)} + h\big\|\nabla(R_h\psi - \psi)\big\|_{L^2(\Omega)} \le Ch^q\|\psi\|_{\dot{H}^q(\Omega)}.$$

Further, for $s \in [0, 1]$ we have

$$\big\|(I - P_h)\psi\big\|_{H^s(\Omega)} \le Ch^{2-s}\|\psi\|_{\dot{H}^2(\Omega)} \quad \forall\psi \in H^2(\Omega) \cap H_0^1(\Omega),$$
$$\big\|(I - P_h)\psi\big\|_{H^s(\Omega)} \le Ch^{1-s}\|\psi\|_{\dot{H}^1(\Omega)} \quad \forall\psi \in H_0^1(\Omega).$$

By interpolation, the operator P_h is also bounded on $\dot{H}^q(\Omega)$, $-1 \le q \le 0$.

Now we can describe the semidiscrete Galerkin scheme. 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 (1.4) as

$$P(\partial_t)u_h(t) - \Delta_h u_h(t) = f_h(t) \quad \text{with } u_h(0) = v_h, \tag{3.1}$$

where $v_h \in X_h$ is an approximation to the initial data v. Next we give a solution representation 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 operators E_h and $\bar{E}_h$, the discrete analogues of (2.2) and (2.3), for $t > 0$, defined respectively by

$$\begin{aligned} E_h(t)v_h &= \sum_{j=1}^N E_{\vec{\alpha},1}\big(-\lambda_j^h t^\alpha, -b_1 t^{\alpha-\alpha_1}, \ldots, -b_m t^{\alpha-\alpha_m}\big)\big(v, \varphi_j^h\big)\varphi_j^h \\ &\quad + \sum_{i=1}^m b_i t^{\alpha-\alpha_i} \sum_{j=1}^N E_{\vec{\alpha},1+\alpha-\alpha_i}\big(-\lambda_j^h t^\alpha, -b_1 t^{\alpha-\alpha_1}, \ldots, -b_m t^{\alpha-\alpha_m}\big)\big(v, \varphi_j^h\big)\varphi_j^h, \end{aligned} \tag{3.2}$$

and

$$\bar{E}_h(t) f_h = \sum_{j=1}^N t^{\alpha-1} E_{\vec{\alpha},\alpha}\big(-\lambda_j^h t^\alpha, -b_1 t^{\alpha-\alpha_1}, \ldots, -b_m t^{\alpha-\alpha_m}\big)\big(f_h, \varphi_j^h\big)\varphi_j^h. \tag{3.3}$$

Then the solution u_h of the discrete problem (3.1) can be expressed by:

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

On the finite element space X_h, we introduce the discrete norm $|||\cdot|||_{\dot{H}^p(\Omega)}$ defined by

$$|||\psi|||_{\dot{H}^p(\Omega)}^2 = \sum_{j=1}^N \big(\lambda_j^h\big)^p \big(\psi, \varphi_j^h\big)^2 \quad \psi \in X_h.$$

The norm $|||\cdot|||_{\dot{H}^p(\Omega)}$ is well defined for all real number p. Clearly, $|||\psi|||_{\dot{H}^1(\Omega)} = \|\psi\|_{\dot{H}^1(\Omega)}$ and $|||\psi|||_{\dot{H}^0(\Omega)} = \|\psi\|_{L^2(\Omega)}$ for any $\psi \in X_h$. Further, the following inverse inequality holds [12]: if the mesh $\mathcal{T}_h$ is quasi-uniform, then for any $l > s$

$$|||\psi|||_{\dot{H}^l(\Omega)} \le Ch^{s-l}|||\psi|||_{\dot{H}^s(\Omega)} \quad \forall\psi \in X_h. \tag{3.5}$$

Lemma 3.2. *Assume that the mesh $\mathcal{T}_h$ is quasi-uniform. Then for any $v_h \in X_h$ the function $u_h(t) = E_h(t)v_h$ satisfies*

$$\left\| P(\partial_t)^\ell u_h(t) \right\|_{\dot{H}^p(\Omega)} \le Ct^{-\alpha(\ell+(p-q)/2)} \| v_h \|_{\dot{H}^q(\Omega)}, \quad t > 0,$$

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

Proof. Upon noting $\| P(\partial_t)E_h(t)v_h \|_{\dot{H}^p(\Omega)} = \| E_h(t)v_h \|_{\dot{H}^{p+2}(\Omega)}$, it suffices to show the case $\ell = 0$. Using the representation (3.4) and Lemma 2.1, we have for $0 \le p - q \le 2$

$$\begin{aligned}
\left\| E_h(t)v_h \right\|^2_{\dot{H}^p(\Omega)} &\le C \sum_{j=1}^N \frac{(\lambda_j^h)^p}{(1+\lambda_j^h t^\alpha)^2} \left| (v_h, \varphi_j^h) \right|^2 \\
&\le Ct^{-(p-q)\alpha} \sum_{j=1}^N \frac{(\lambda_j^h t^\alpha)^{p-q}}{(1+\lambda_j^h t^\alpha)^2} (\lambda_j^h)^q \left| (v_h, \varphi_j^h) \right|^2 \\
&\le Ct^{-(p-q)\alpha} \| v_h \|^2_{\dot{H}^q(\Omega)},
\end{aligned}$$

where the last inequality follows from $\sup_{1\le j\le N} \frac{(\lambda_j^h t^\alpha)^{p-q}}{(1+\lambda_j^h t^\alpha)^2} \le C$ for $0 \le p - q \le 2$. □

The next result is a discrete analogue to Lemma 2.3.

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

$$\left\| \bar{E}_h(t)\chi \right\|_{\dot{H}^p(\Omega)} \le \begin{cases} Ct^{-1+\alpha(1+(q-p)/2)} \| \chi \|_{\dot{H}^q(\Omega)}, & 0 \le p - q \le 2, \\ Ct^{-1+\alpha} \| \chi \|_{\dot{H}^q(\Omega)}, & p < q. \end{cases}$$

Proof. The proof for the case $0 \le p - q \le 2$ is similar to Lemma 2.3. The other assertion follows from the fact that $\{\lambda_j^h\}_{j=1}^N$ are bounded from zero independent of h. □

3.2. Error estimates for the homogeneous problem

To derive error estimates, first we consider the case of smooth initial data, i.e., $v \in \dot{H}^2(\Omega)$. To this end, we split the error $u_h(t) - u(t)$ into two terms:

$$u_h - u = (u_h - R_h u) + (R_h u - u) = \vartheta + \varrho.$$

By Lemma 3.1 and Theorem 2.1, we have for any $t > 0$

$$\left\| \varrho(t) \right\|_{L^2(\Omega)} + h \left\| \nabla \varrho(t) \right\|_{L^2(\Omega)} \le Ch^2 \| v \|_{\dot{H}^2(\Omega)}. \tag{3.6}$$

So it suffices to get proper estimates for $\vartheta(t)$, which is given below.

Lemma 3.4. *The function $\vartheta(t) := u_h(t) - R_h u(t)$ satisfies for $p = 0, 1$*

$$\left\| \vartheta(t) \right\|_{\dot{H}^p(\Omega)} \le Ch^{2-p} \| v \|_{\dot{H}^2(\Omega)}.$$

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

$$P(\partial_t)\vartheta(t) - \Delta_h \vartheta(t) = -P_h P(\partial_t)\varrho(t),$$

with $\vartheta(0) = 0$. By the representation (3.4),

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

Then by Lemmas 3.3 and 3.1, and Theorem 2.1 we have for $p = 0, 1$

$$\left\| \vartheta(t) \right\|_{\dot{H}^p(\Omega)} \le \int_0^t \left\| \bar{E}_h(t-s) P_h P(\partial_t)\varrho(s) \right\|_{\dot{H}^p(\Omega)} ds$$

$$\le C\int_0^t (t-s)^{(1-p/2)\alpha-1}\left\|P(\partial_t)\varrho(s)\right\|_{L^2(\Omega)}ds$$

$$\le Ch^{2-p}\int_0^t (t-s)^{(1-p/2)\alpha-1}\left\|P(\partial_t)u(s)\right\|_{\dot H^{2-p}(\Omega)}ds$$

$$\le Ch^{2-p}\int_0^t (t-s)^{(1-p/2)\alpha-1}s^{-(1-p/2)\alpha}\,ds\|v\|_{\dot H^2(\Omega)}\le Ch^{2-p}\|v\|_{\dot H^2(\Omega)},$$

which is the desired result. □

Using (3.6), Lemma 3.4 and the triangle inequality, we arrive at our first estimate, which is formulated in the following theorem:

Theorem 3.1. *Let $v\in\dot H^2(\Omega)$ and $f\equiv 0$, and u and u_h be the solutions of (1.1) and (1.4) with $v_h=R_hv$, respectively. Then*

$$\left\|u_h(t)-u(t)\right\|_{L^2(\Omega)}+h\left\|\nabla\big(u_h(t)-u(t)\big)\right\|_{L^2(\Omega)}\le Ch^2\|v\|_{\dot H^2(\Omega)}.$$

Now we turn to the nonsmooth case, i.e., $v\in\dot H^q(\Omega)$ with $-1<q\le 1$. Since the Ritz projection R_h is not well-defined for nonsmooth data, we use instead the $L^2(\Omega)$-projection $v_h=P_hv$ and split the error u_h-u into:

$$u_h-u=(u_h-P_hu)+(P_hu-u)=:\widetilde\vartheta+\widetilde\varrho. \tag{3.7}$$

By Lemma 3.1 and Theorem 2.1 we have for $-1\le q\le 1$

$$\begin{aligned}\left\|\widetilde\varrho(t)\right\|_{L^2(\Omega)}+h\left\|\nabla\widetilde\varrho(t)\right\|_{L^2(\Omega)}&\le Ch^{2+\min(0,q)}\left\|u(t)\right\|_{\dot H^{2+\min(0,q)}(\Omega)}\\&\le Ch^{2+\min(0,q)}t^{-\alpha(1-\max(q/2,0))}\|v\|_{\dot H^q(\Omega)}.\end{aligned}$$

Thus, we only need to estimate the term $\widetilde\vartheta(t)$, which is stated in the following lemma.

Lemma 3.5. *Let $\widetilde\vartheta(t)=u_h(t)-P_hu(t)$. Then for $p=0,1$, $-1<q\le 1$, there holds (with $\ell_h=|\ln h|$)*

$$\left\|\widetilde\vartheta(t)\right\|_{\dot H^p(\Omega)}\le Ch^{\min(q,0)+2-p}\ell_h t^{-\alpha(1-\max(q/2,0))}\|v\|_{\dot H^q(\Omega)}.$$

Proof. Obviously, $P_hP(\partial_t)\widetilde\varrho=P(\partial_t)P_h(P_hu-u)=0$ and using the identity $\Delta_hR_h=P_h\Delta$, we get the following problem for $\widetilde\vartheta$:

$$P(\partial_t)\widetilde\vartheta(t)-\Delta_h\widetilde\vartheta(t)=-\Delta_h(R_hu-P_hu)(t),\quad t>0,\ \widetilde\vartheta(0)=0. \tag{3.8}$$

Using (3.4), $\widetilde\vartheta(t)$ can be represented by

$$\widetilde\vartheta(t)=-\int_0^t\bar E_h(t-s)\Delta_h(R_hu-P_hu)(s)\,ds. \tag{3.9}$$

Let $A=\bar E_h(t-s)\Delta_h(R_hu-P_hu)(s)$. Then by Lemma 3.3, there holds for $p=0,1$:

$$\begin{aligned}\|A\|_{\dot H^p(\Omega)}&\le C(t-s)^{\epsilon\alpha/2-1}\left\|\Delta_h(R_hu-P_hu)(s)\right\|_{\dot H^{p-2+\epsilon}(\Omega)}\\&\le C(t-s)^{\epsilon\alpha/2-1}\left\|(R_hu-P_hu)(s)\right\|_{\dot H^{p+\epsilon}(\Omega)}.\end{aligned}$$

Then by (3.5), Theorem 2.1 and Lemma 3.1 we have for $p=0,1$, $-1\le q\le 1$ and $\epsilon\in(0,1)$

$$\begin{aligned}\|A\|_{\dot H^p(\Omega)}&\le Ch^{\min(q,0)+2-p-\epsilon}(t-s)^{\epsilon\alpha/2-1}\left\|u(s)\right\|_{\dot H^{\min(q,0)+2}(\Omega)}\\&\le Ch^{\min(q,0)+2-p-\epsilon}(t-s)^{\epsilon\alpha/2-1}s^{-(1-\max(q/2,0))\alpha}\|v\|_{\dot H^q(\Omega)}.\end{aligned}$$

Then plugging the estimate back into (3.9) yields

$$\begin{aligned}\|\widetilde\vartheta\|_{\dot H^p(\Omega)}&\le Ch^{\min(q,0)+2-p-\epsilon}\int_0^t(t-s)^{\epsilon\alpha/2-1}s^{-(1-\max(q/2,0))\alpha}\,ds\|v\|_{\dot H^q(\Omega)}\\&\le C\epsilon^{-1}h^{\min(q,0)+2-p-\epsilon}t^{-\alpha(1-\max(q/2,0))}\|v\|_{\dot H^q(\Omega)},\end{aligned}$$

where the last inequality follows from the fact

$$\int_0^t (t-s)^{\epsilon\alpha/2-1} s^{-(1-\max(q/2,0))\alpha}\, ds = t^{\epsilon\alpha/2-(1-\max(q/2,0))\alpha} B\big(\epsilon\alpha/2, 1-\big(1-\max(q/2,0)\big)\alpha\big),$$

the identity

$$B\big(\epsilon\alpha/2, 1-\big(1-\max(q/2,0)\big)\alpha\big) = \frac{\Gamma(\epsilon\alpha/2)\Gamma(1-(1-\max(q/2,0))\alpha)}{\Gamma(\epsilon\alpha/2+1-(1-\max(q/2,0))\alpha)},$$

and the estimate $\Gamma(\epsilon\alpha/2) = \Gamma(1+\epsilon\alpha/2)/(\epsilon\alpha/2) \le C\epsilon^{-1}$. Now with the choice $\epsilon = 1/\ell_h$, and the fact that $h^{-1/|\ln h|}$ is uniformly bounded for $h > 0$, we obtain the desired estimate. □

Now the triangle inequality yields an error estimate for nonsmooth initial data.

Theorem 3.2. *Let $f \equiv 0$, u and u_h be the solutions of (1.1) with $v \in \dot H^q(\Omega)$, $-1 < q \le 1$, and (1.4) with $v_h = P_h v$, respectively. Then with $\ell_h = |\ln h|$, there holds*

$$\big\|u_h(t) - u(t)\big\|_{L^2(\Omega)} + h\big\|\nabla\big(u_h(t) - u(t)\big)\big\|_{L^2(\Omega)} \le Ch^{\min(q,0)+2}\,\ell_h\, t^{-\alpha(1-\max(q/2,0))}\|v\|_{\dot H^q(\Omega)}.$$

3.3. Error estimates for the inhomogeneous problem

Now we derive error estimates for the semidiscrete Galerkin approximation of the inhomogeneous problem with $f \in L^\infty(0,T;\dot H^q(\Omega))$, $-1 < q \le 0$, and $v \equiv 0$, in both L^2- and L^∞-norm in time. The case $f \in L^\infty(0,T;\dot H^q(\Omega))$, $0 < q \le 1$, is simpler and can be treated analogously. To this end, we appeal again to the splitting (3.7). By Theorem 2.2 and Lemma 3.1, the following estimate holds for $\widetilde\varrho$:

$$\big\|\widetilde\varrho(t)\big\|_{L^2(\Omega)} + h\big\|\nabla\widetilde\varrho(t)\big\|_{L^2(\Omega)} \le Ch^{2+q-\epsilon}t^{-\epsilon\alpha}\big\|u(t)\big\|_{\dot H^{2+q-\epsilon}(\Omega)} \le C\epsilon^{-1}h^{2+q-\epsilon}\|f\|_{L^\infty(0,t;\dot H^q(\Omega))},$$

where the last inequality follows from the fact $t \le T$, and $t^{\epsilon\alpha}$ is bounded. Now the choice $\ell_h = |\ln h|$, $\epsilon = 1/\ell_h$, yields

$$\big\|\widetilde\varrho(t)\big\|_{L^2(\Omega)} + h\big\|\nabla\widetilde\varrho(t)\big\|_{L^2(\Omega)} \le C\ell_h h^{2+q}\|f\|_{L^\infty(0,t;\dot H^q(\Omega))}. \tag{3.10}$$

Thus, it suffices to bound the term $\widetilde\vartheta$; see the lemma below.

Lemma 3.6. *Let $\widetilde\vartheta(t)$ be defined by (3.9), and $f \in L^\infty(0,T;\dot H^q(\Omega))$, $-1 < q \le 0$. Then with $\ell_h = |\ln h|$, there holds*

$$\big\|\widetilde\vartheta(t)\big\|_{L^2(\Omega)} + h\big\|\nabla\widetilde\vartheta(t)\big\|_{L^2(\Omega)} \le Ch^{2+q}\ell_h^2\|f\|_{L^\infty(0,t;\dot H^q(\Omega))}.$$

Proof. By (3.4) and Lemma 3.3, we deduce that for $p = 0, 1$

$$\begin{aligned}\big\|\widetilde\vartheta(t)\big\|_{\dot H^p(\Omega)} &\le \int_0^t \big\|\bar E_h(t-s)\Delta_h(R_h u - P_h u)(s)\big\|_{\dot H^p(\Omega)} ds \\ &\le C\int_0^t (t-s)^{\epsilon\alpha/2-1}\big\|\Delta_h(R_h u - P_h u)(s)\big\|_{\dot H^{p-2+\epsilon}(\Omega)} ds \\ &\le C\int_0^t (t-s)^{\epsilon\alpha/2-1}\big|\!\big|\!\big|R_h u(s) - P_h u(s)\big|\!\big|\!\big|_{\dot H^{p+\epsilon}(\Omega)} ds.\end{aligned}$$

Further, using (3.5) and Lemma 3.1, we deduce for $p = 0, 1$

$$\begin{aligned}\big\|\widetilde\vartheta(t)\big\|_{\dot H^p(\Omega)} &\le Ch^{-\epsilon}\int_0^t (t-s)^{\epsilon\alpha/2-1}\big\|R_h u(s) - P_h u(s)\big\|_{\dot H^p(\Omega)} ds \\ &\le Ch^{2+q-p-2\epsilon}\int_0^t (t-s)^{\epsilon\alpha/2-1}\big\|u(s)\big\|_{\dot H^{2+q-\epsilon}(\Omega)} ds.\end{aligned}$$

Now by (2.5) and the choice $\epsilon = 1/\ell_h$ we get for $p = 0, 1$:

$$\|\widetilde{\vartheta}(t)\|_{\dot H^p(\Omega)} \le C\epsilon^{-1}h^{2+q-p-2\epsilon}\|f\|_{L^\infty(0,t;\dot H^q(\Omega))}\int_0^t (t-s)^{\epsilon\alpha/2-1}s^{\epsilon\alpha/2}ds$$
$$\le C\epsilon^{-2}h^{2+q-p-2\epsilon}\|f\|_{L^\infty(0,t;\dot H^q(\Omega))} \le Ch^{2+q-p}\ell_h^2\|f\|_{L^\infty(0,t;\dot H^q(\Omega))},$$

where the first inequality follows directly from (2.5) and the second inequality follows from

$$\int_0^t (t-s)^{\epsilon\alpha/2-1}s^{\epsilon\alpha/2}ds \le t^{\epsilon\alpha/2}\int_0^t (t-s)^{\epsilon\alpha/2-1}ds = \frac{2}{\epsilon\alpha}t^{\epsilon\alpha} \le C\epsilon^{-1}.$$

This completes the proof of the lemma. □

An inspection of the proof of Lemma 3.6 indicates that for $0 < q < 1$, one can get rid of one factor ℓ_h. Now we can state an error estimate in L^∞-norm in time.

Theorem 3.3. *Let $v \equiv 0$, $f \in L^\infty(0,T;\dot H^q(\Omega))$, $-1 < q \le 0$, and u and u_h be the solutions of (1.1) and (1.4) with $f_h = P_h f$, respectively. Then with $\ell_h = |\ln h|$ and $t > 0$, there holds*

$$\|u_h(t) - u(t)\|_{L^2(\Omega)} + h\|\nabla(u_h(t) - u(t))\|_{L^2(\Omega)} \le Ch^{2+q}\ell_h^2\|f\|_{L^\infty(0,t;\dot H^q(\Omega))}.$$

Last, we derive an error estimate in L^2-norm in time. To this end, we need a discrete analogue of Theorem 2.3, whose proof follows identically and hence is omitted.

Lemma 3.7. *Let u_h be the solution of (1.4) with $v_h = 0$. Then for arbitrary $p > -1$*

$$\int_0^T |||P(\partial_t)u_h(t)|||^2_{\dot H^p(\Omega)} + |||u_h(t)|||^2_{\dot H^{p+2}(\Omega)}\,dt \le \int_0^T |||f_h(t)|||^2_{\dot H^p(\Omega)}dt.$$

Theorem 3.4. *Let $v \equiv 0$, $f \in L^\infty(0,T;\dot H^q(\Omega))$, $-1 < q \le 0$, and u and u_h be the solutions of (1.1) and (1.4) with $f_h = P_h f$, respectively. Then*

$$\|u_h - u\|_{L^2(0,T;L^2(\Omega))} + h\|\nabla(u_h - u)\|_{L^2(0,T;L^2(\Omega))} \le Ch^{2+q}\|f\|_{L^2(0,T;\dot H^q(\Omega))}.$$

Proof. We use the splitting (3.7). By Theorem 2.3 and Lemma 3.1

$$\|\widetilde{\varrho}\|_{L^2(0,T;L^2(\Omega))} + h\|\nabla\widetilde{\varrho}\|_{L^2(0,T;L^2(\Omega))} \le Ch^{2+q}\|u\|_{L^2(0,T;\dot H^{2+q}(\Omega))} \le Ch^{2+q}\|f\|_{L^2(0,T;\dot H^q(\Omega))}.$$

By (3.4), (3.8) and Lemmas 3.7 and 3.1, we have for $p = 0, 1$:

$$\int_0^T \|\widetilde{\vartheta}(t)\|^2_{\dot H^p(\Omega)}dt \le C\int_0^T |||\Delta_h(R_h u - P_h u)(t)|||^2_{\dot H^{p-2}(\Omega)}dt \le C\int_0^T |||(R_h u - P_h u)(t)|||^2_{\dot H^p(\Omega)}dt$$
$$\le Ch^{4+2q-2p}\|u(t)\|^2_{L^2(0,T;\dot H^{2+q}(\Omega))} \le Ch^{4+2q-2p}\|f(t)\|^2_{L^2(0,T;\dot H^q(\Omega))}.$$

Combing the preceding two estimates yields the desired assertion. □

4. A fully discrete scheme

Now we describe a fully discrete scheme for problem (1.1) based on the finite difference method introduced in [23]. To discretize the time-fractional derivatives, we divide the interval $[0,T]$ uniformly with a time step size $\tau = T/K$, $K \in \mathbb{N}$. We use the following discretization:

$$
\begin{aligned}
\partial_t^\alpha u(x,t_{n+1}) &= \frac{1}{\Gamma(1-\alpha)}\sum_{j=0}^{n}\int_{t_j}^{t_{j+1}}(t_{n+1}-s)^{-\alpha}\frac{\partial u(x,s)}{\partial s}\,ds \\
&= \frac{1}{\Gamma(1-\alpha)}\sum_{j=0}^{n}\frac{u(x,t_{j+1})-u(x,t_j)}{\tau}\int_{t_j}^{t_{j+1}}(t_{n+1}-s)^{-\alpha}\,ds + r_{\alpha,\tau}^{n+1} \\
&= \frac{1}{\Gamma(2-\alpha)}\sum_{j=0}^{n}d_{\alpha,j}\frac{u(x,t_{n+1-j})-u(x,t_{n-j})}{\tau^\alpha} + r_{\alpha,\tau}^{n+1},
\end{aligned}
\tag{4.1}
$$

where $d_{\alpha,j}=(j+1)^{1-\alpha}-j^{1-\alpha}$ with $j=0,1,2,\dots,n$ and $r_{\alpha,\tau}^{n+1}$ denotes the local truncation error. Lin and Xu [23, Lemma 3.1] (see also [34, Lemma 4.1]) showed that the truncation error $r_{\alpha,\tau}^{n+1}$ can be bounded by

$$
\left|r_{\alpha,\tau}^{n+1}\right| \le C\max_{0\le t\le T}\left|u_{tt}(x,t)\right|\tau^{2-\alpha}. \tag{4.2}
$$

Then the multi-term fractional derivative $P(\partial_t)u(t)$ at $t=t_{n+1}$ in (1.1) can be discretized by

$$
P(\partial_t)u(t_{n+1}) = P_\tau(\bar\partial_t)u(t_{n+1}) + R_\tau^{n+1}, \tag{4.3}
$$

where the discrete differential operator $P_\tau(\bar\partial_t)$ is defined by

$$
P_\tau(\bar\partial_t)u(t_{n+1}) := \frac{1}{\Gamma(2-\alpha)}\sum_{j=0}^{n}P_j\frac{u(x,t_{n+1-j})-u(x,t_{n-j})}{\tau^\alpha}, \tag{4.4}
$$

where the coefficients $\{P_j\}$ are defined by

$$
P_j = d_{\alpha,j} + \sum_{i=1}^{m}\frac{\Gamma(2-\alpha)b_i d_{\alpha_i,j}\tau^{\alpha-\alpha_i}}{\Gamma(2-\alpha_i)}, \quad j\in\mathbb{N}.
$$

Then by (4.2) the local truncation error R_τ^{n+1} of the approximation $P_\tau(\bar\partial_t)u(t_{n+1})$ is bounded by

$$
\left|R_\tau^{n+1}\right| \le C\max_{0\le t\le T}\left|u_{tt}(x,t)\right|\left(\tau^{2-\alpha}+\sum_{i=1}^{m}b_i\tau^{2-\alpha_i}\right) \le C\tau^{2-\alpha}\max_{0\le t\le T}\left|u_{tt}(x,t)\right|. \tag{4.5}
$$

By the monotonicity and convergence of $\{d_{\alpha,j}\}$ [23, Eq. (13)], we know that

$$
P_0 > P_1 > \ldots > 0 \quad \text{and} \quad P_j \to 0 \quad \text{for } j\to\infty. \tag{4.6}
$$

Now we arrive at the following fully discrete scheme: find $U^{n+1}\in X_h$ such that

$$
\left(P_\tau(\bar\partial_t)U^{n+1},\chi\right) + \left(\nabla U^{n+1},\nabla\chi\right) = \left(F^{n+1},\chi\right) \quad \forall\chi\in X_h, \tag{4.7}
$$

where $F^{n+1}=f(x,t_{n+1})$. Upon setting $\gamma=\Gamma(2-\alpha)\tau^\alpha$, the fully discrete scheme (4.7) is equivalent to finding $U^{n+1}\in X_h$ such that for all $\chi\in X_h$

$$
P_0\left(U^{n+1},\chi\right) + \gamma\left(\nabla U^{n+1},\nabla\chi\right) = \sum_{j=0}^{n-1}(P_j-P_{j+1})\left(U^{n-j},\chi\right) + P_n\left(U^0,\chi\right) + \gamma\left(F^{n+1},\chi\right). \tag{4.8}
$$

The next result gives the stability of the fully discrete scheme.

Lemma 4.1. *The fully discrete scheme (4.8) is unconditionally stable, i.e., for all $n\in\mathbb{N}$*

$$
\left\|U^n\right\|_{L^2(\Omega)} \le \left\|U^0\right\|_{L^2(\Omega)} + c\max_{1\le j\le n}\left\|F^j\right\|_{L^2(\Omega)}, \tag{4.9}
$$

where the constant c depends only on α and T.

Proof. The case $n=1$ is trivial. Then the proof proceeds by mathematical induction. By noting the monotone decreasing property of the sequence $\{P_n\}$ from (4.6) and choosing $\chi=U^{n+1}$ in (4.8), we deduce

$$
\begin{aligned}
P_0\|U^{n+1}\|_{L^2(\Omega)} &\le \sum_{j=0}^{n-1}(P_j-P_{j+1})\|U^{n-j}\|_{L^2(\Omega)}+P_n\|U^0\|_{L^2(\Omega)}+\gamma\|F^{n+1}\|_{L^2(\Omega)}\\
&\le \sum_{j=0}^{n-1}(P_j-P_{j+1})\|U^{n-j}\|_{L^2(\Omega)}+P_n\|U^0\|_{L^2(\Omega)}+\gamma\max_{1\le j\le n+1}\|F^j\|_{L^2(\Omega)}\\
&\le P_0\|U^0\|_{L^2(\Omega)}+\big(c(P_0-P_n)+\gamma\big)\max_{1\le j\le n+1}\|F^j\|_{L^2(\Omega)}.
\end{aligned}
$$

Using the monotonicity of $\{P_n\}$ again gives

$$c(P_0-P_n)+\gamma\le cP_0-(cP_N-\gamma).$$

It suffices to choose a constant c such that $cP_N-\gamma>0$. By taking $\tau=T/N$ and noting the concavity of the function $g(\tau)=(T+\tau)^{1-\alpha}$, we get $g(\tau)-g(0)\le g'(0)\tau$, i.e., $(T+\tau)^{1-\alpha}-T^{1-\alpha}\le(1-\alpha)T^{-\alpha}\tau$, and thus

$$d_{\alpha,N}=(N+1)^{1-\alpha}-N^{1-\alpha}=\big((T+\tau)^{1-\alpha}-T^{1-\alpha}\big)\tau^{\alpha-1}\le(1-\alpha)T^{-\alpha}\tau^{\alpha}.$$

Hence

$$P_N\le\left((1-\alpha)+\sum_{i=1}^{m}\frac{\Gamma(2-\alpha)b_i(1-\alpha_i)}{\Gamma(2-\alpha_i)}T^{\alpha-\alpha_i}\right)T^{-\alpha}\tau^{\alpha}=:C_0T^{-\alpha}\tau^{\alpha}.$$

Then by choosing $c=C_0^{-1}T^{\alpha}$ we obtain

$$P_0\|U^{n+1}\|_{L^2(\Omega)}\le P_0\|U^0\|_{L^2(\Omega)}+cP_0\max_{1\le j\le n+1}\|F^j\|_{L^2(\Omega)}.$$

The desired result follows by dividing both sides by P_0. □

Next we state an error estimate for the fully discrete scheme. In order to analyze the temporal discretization error, we assume that the solution is sufficiently smooth.

Theorem 4.1. *Let the solution u be sufficiently smooth, and $\{U^n\}\subset X_h$ be the solution of the fully discrete scheme (4.8) with U^0 such that*

$$\|U^0-v\|_{L^2(\Omega)}\le Ch^2\|v\|_{\dot H^2(\Omega)}.$$

Then there holds

$$\|u(t_n)-U^n\|_{L^2(\Omega)}\le C\Big(h^2\Big(\|v\|_{H^2(\Omega)}+\|f\|_{L^\infty(0,T;L^2(\Omega))}+\max_{0\le t\le t_n}\|u_t\|_{\dot H^2(\Omega)}\Big)+\tau^{2-\alpha}\max_{0\le t\le t_n}\|u_{tt}(t)\|_{L^2(\Omega)}\Big).$$

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

$$e^n=\big(u(t_n)-R_hu(t_n)\big)+\big(R_hu(t_n)-U^n\big)=:\varrho^n+\theta^n.$$

Here ϱ^n is a special case of $\varrho(t):=u(t)-R_hu(t)$ with $t=t_n$. Applying Lemma 3.1, we have

$$
\begin{aligned}
&\|\varrho(t)\|_{L^2(\Omega)}\le Ch^2\|u(t)\|_{\dot H^2(\Omega)}\le Ch^2\big(\|v\|_{H^2(\Omega)}+\|f\|_{L^\infty(0,T;H^1(\Omega))}\big),\\
&\|\partial_t\varrho(t)\|_{L^2(\Omega)}\le Ch^2\|\partial_tu(t)\|_{\dot H^2(\Omega)}.
\end{aligned}
$$

It suffices to bound the term θ^n. By comparing (1.1) and (4.7), we have the error equation

$$\big(P_\tau(\bar\partial_t)\theta^n,\chi\big)+\big(\nabla\theta^n,\nabla\chi\big)=\big(\omega^n,\chi\big),\tag{4.10}$$

where the right hand side ω^n is given by

$$\omega^n=R_hP_\tau(\bar\partial_t)u(t_n)-P(\partial_t)u(t_n)=-P_\tau(\bar\partial_t)\varrho_n-R_\tau^n:=\omega_1^n+\omega_2^n,$$

where the truncation error R_τ^n is defined in (4.3). Using the identity

$$\varrho^{j+1}-\varrho^j=\int_{t_j}^{t_{j+1}}\varrho_t(t)\,dt,$$

we can bound the term ω_1^n by

Table 1
Numerical results for the case with a smooth solution at $t = 1$ with $\beta = 0.2$ and $\alpha = 0.25, 0.5, 0.95$, discretized on a uniform mesh with $h = 2^{-10}$ and $\tau = 0.2 \times 2^{-k}$.

α	τ	1/10	1/20	1/40	1/80	1/160	Rate
$\alpha = 0.25$	L^2-norm	5.58e-4	1.73e-4	5.25e-5	1.51e-5	3.90e-6	≈1.81 (1.75)
$\alpha = 0.5$	L^2-norm	1.45e-3	5.11e-4	1.78e-4	6.17e-5	2.08e-5	≈1.55 (1.50)
$\alpha = 0.95$	L^2-norm	7.92e-3	3.79e-3	1.82e-3	8.73e-4	4.20e-4	≈1.06 (1.05)

$$\begin{aligned}
\|\omega_1^n\|_{L^2(\Omega)} &\le C\left\| \sum_{j=0}^{n-1} \frac{\varrho(t_{j+1})-\varrho(t_j)}{\tau} \int_{t_j}^{t_{j+1}} (t_n-s)^{-\alpha} + \sum_{i=1}^m b_i (t_n-s)^{-\alpha_i}\, ds \right\|_{L^2(\Omega)} \\
&\le C \sum_{j=0}^{n-1} \tau^{-1} \int_{t_j}^{t_{j+1}} \|\varrho_t(t)\|_{L^2(\Omega)}\, dt \int_{t_j}^{t_{j+1}} (t_n-s)^{-\alpha} + \sum_{i=1}^m b_i (t_n-s)^{-\alpha_i}\, ds \\
&\le Ch^2 \max_{0\le t\le t_n} \|u_t\|_{\dot H^2(\Omega)} \left(\int_0^{t_n} (t_n-s)^{-\alpha} + \sum_{i=1}^m b_i (t_n-s)^{-\alpha_i}\, ds \right) \\
&\le Ch^2 \max_{0\le t\le t_n} \|u_t\|_{\dot H^2(\Omega)}.
\end{aligned}$$

Meanwhile, the second term ω_2^n can be bounded using (4.5). Then by the stability from Lemma 4.1 for the error equation (4.10), we obtain

$$\begin{aligned}
\|\theta^n\|_{L^2(\Omega)} &\le C\left(\|\theta^0\|_{L^2(\Omega)} + \max_{1\le j\le n} \|\omega_1^j\|_{L^2(\Omega)} + \max_{1\le j\le n} \|\omega_2^j\|_{L^2(\Omega)} \right) \\
&\le C\left(h^2 \|v\|_{\dot H^2(\Omega)} + h^2 \max_{0\le t\le t_n} \|u_t\|_{\dot H^2(\Omega)} + \tau^{2-\alpha} \max_{0\le t\le t_n} \|u_{tt}(t)\|_{L^2(\Omega)} \right). \qquad \square
\end{aligned}$$

Remark 4.1. The error estimate in Theorem 4.1 holds only if the solution u is sufficiently smooth. There seems no known error estimate expressed in terms of the initial data (and right hand side) only for fully discrete schemes for nonsmooth initial data even for the single-term time-fractional diffusion equation with a Caputo fractional derivative.

5. Numerical experiments

In this part we present one- and two-dimensional numerical experiments to verify the error estimates in Sections 3 and 4. We shall discuss the cases of a homogeneous problem and an inhomogeneous problem separately.

5.1. The case of a smooth solution

Here we consider the following one-dimensional problem on the unit interval $\Omega = (0, 1)$ with $0 < \beta < \alpha < 1$

$$\begin{aligned}
\partial_t^\alpha u(x,t) + \partial_t^\beta u(x,t) - \partial_x^2 u(x,t) &= f, && 0 < x < 1,\ 0 \le t \le T, \\
u(0,t) = u(1,t) &= 0, && 0 \le t \le T, \\
u(x,0) &= v(x), && 0 \le x \le 1.
\end{aligned} \tag{5.1}$$

In order to verify the estimate in Theorem 4.1, we first check the case that the solution u is sufficiently smooth. To this end, we set initial data v to $v(x) = x(1-x)$ and the source term f to $f(x,t) = (2t^{2-\alpha}/\Gamma(3-\alpha) + 2t^{2-\beta}/\Gamma(3-\beta))(-x^2+x) + 2(1+t^2)$. Then the exact solution u is given by $u(x,t) = (1+t^2)(-x^2+x)$, which is very smooth.

In our computation, we divide the unit interval Ω into N equally spaced subintervals, with a mesh size $h = 1/N$. Similarly, we fix the time step size at $\tau = 1/K$. Here we choose N large enough so that the space discretization error is negligible, and the time discretization error dominates. We measure the accuracy of the numerical approximation U^n by the normalized L^2 error $\|U^n - u(t_n)\|_{L^2(\Omega)}/\|v\|_{L^2(\Omega)}$. In Table 1, we show the temporal convergence rates, indicated in the column `rate` (the number in bracket is the theoretical rate), for three different α values, which fully confirm the theoretical result, cf. also Fig. 1 for the plot of the convergence rates.

5.2. Homogeneous problems

In this part we present numerical results to illustrate the spatial convergence rates in Section 3. We performed numerical tests on the following three different initial data:

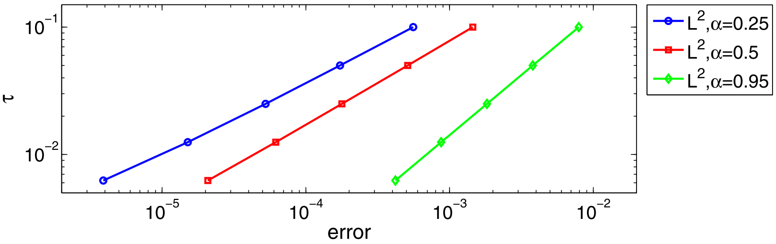


Fig. 1. Numerical results for the case with a smooth solution at $t = 1$ with $\beta = 0.2$ and $\alpha = 0.25, 0.5, 0.95$, discretized on a uniform mesh $h = 2^{-10}$ and $\tau = 0.2 \times 2^{-k}$.

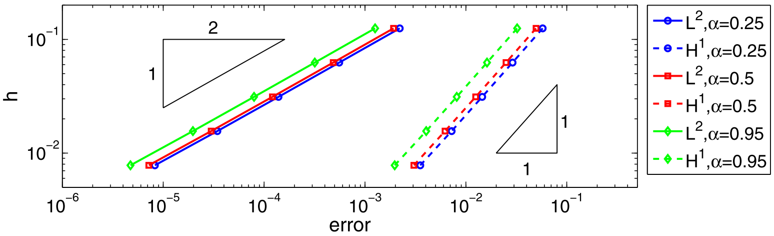


Fig. 2. Numerical results for example (2a) at $t = 1$ with $\beta = 0.2$ and $\alpha = 0.25$, 0.5, 0.95, discretized on a uniform mesh $h = 2^{-k}$ and $\tau = 2 \times 10^{-5}$.

Table 2
Numerical results for the nonsmooth case (2b) with $\alpha = 0.5$ and $\beta = 0.2$ at $t = 1, 0.01, 0.001$, discretized on a uniform mesh with $h = 2^{-k}$ and $\tau = t/(5 \times 10^4)$.

t	k	3	4	5	6	7	Rate
$t = 1$	L^2-norm	1.86e-3	4.64e-4	1.16e-4	2.87e-5	6.88e-6	≈2.02 (2.00)
	H^1-norm	4.89e-2	2.44e-2	1.22e-2	6.07e-3	2.96e-3	≈1.01 (1.00)
$t = 0.01$	L^2-norm	8.04e-3	2.00e-3	5.01e-4	1.24e-4	2.98e-5	≈2.03 (2.00)
	H^1-norm	2.31e-1	1.16e-1	5.79e-2	2.88e-2	1.40e-2	≈1.01 (1.00)
$t = 0.001$	L^2-norm	1.65e-2	4.14e-3	1.03e-3	2.56e-4	6.18e-4	≈2.01 (2.00)
	H^1-norm	5.15e-1	2.58e-1	1.29e-1	6.41e-2	3.13e-2	≈1.01 (1.00)

(2a) Smooth data: $v(x) = \sin(2\pi x)$ which belongs to $H^2(\Omega) \cap H_0^1(\Omega)$.
(2b) Nonsmooth data: $v(x) = \chi_{(0,1/2]}$ which lies in the space $\dot{H}^{\epsilon}(\Omega)$ for any $\epsilon \in [0, 1/2)$.
(2c) Very weak data: $v(x) = \delta_{1/2}(x)$, a Dirac $\delta_{1/2}(x)$-function concentrated at $x = 1/2$, which belongs to the space $\dot{H}^{-\epsilon}(\Omega)$ for any $\epsilon \in (1/2, 1]$.

In order to check the convergence rate of the semidiscrete scheme, we discretize the fractional derivatives with a small time step τ so that the temporal discretization error is negligible. In view of the possibly singular behavior as $t \to 0$, we set the time step τ to $\tau = t/(5 \times 10^4)$, with t being the terminal time. For each example, we measure the error $e(t) = u(t) - u_h(t)$ by the normalized errors $\|e(t)\|_{L^2(\Omega)}/\|v\|_{L^2(\Omega)}$ and $\|\nabla e(t)\|_{L^2(\Omega)}/\|v\|_{L^2(\Omega)}$. The normalization enables us to observe the behavior of the error with respect to time in case of nonsmooth initial data.

5.2.1. Numerical results for example (2a): smooth initial data

The numerical results show $O(h^2)$ and $O(h)$ convergence rates for the L^2- and H^1-norms of the error, respectively, for all three different α values, cf. Fig. 2. As the value of α increases from 0.25 to 0.95, the error at $t = 1$ decreases accordingly, which resembles that for the single-term time-fractional diffusion equation [12].

5.2.2. Numerical results for example (2b): nonsmooth initial data

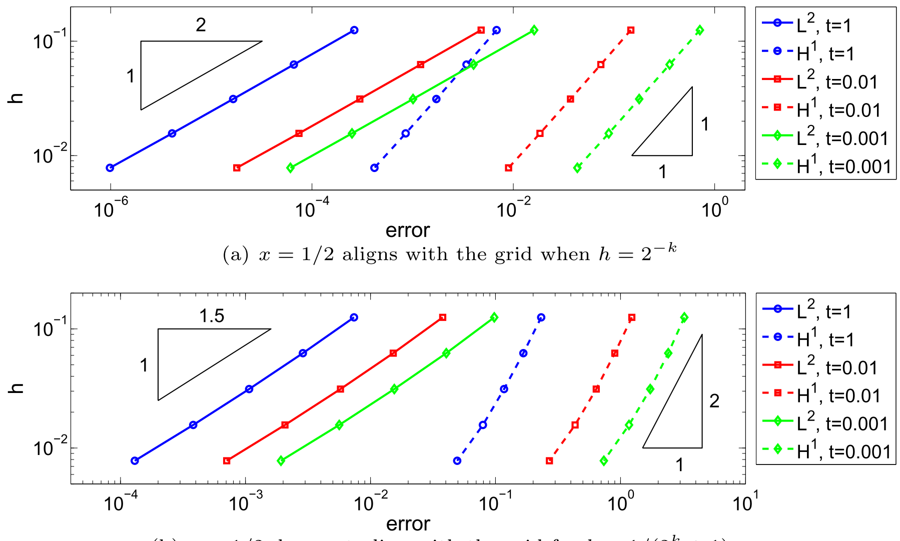
For nonsmooth initial data, we are particularly interested in errors for t close to zero, and thus we also present the errors at $t = 0.01$ and $t = 0.001$; see Table 2. The numerical results fully confirm the theoretically predicted rates for nonsmooth initial data. Further, in Table 3 we show the L^2-norm of the error for fixed $h = 2^{-6}$ and $t \to 0$. We observe that the error deteriorates as $t \to 0$. Upon noting $v \in \dot{H}^{1/2-\epsilon}(\Omega)$, it follows from Theorem 3.2 that the error grows like $O(t^{-3\alpha/4})$, which agrees well with the results in Table 3.

5.2.3. Numerical results for example (2c): very weak initial data

The numerical results show a superconvergence with a rate of $O(h^2)$ in the L^2-norm and $O(h)$ in the H^1-norm, cf. Fig. 3(a). This is attributed to the fact that in one-dimension the solution with the Dirac δ-function as the initial data is

Table 3
L^2-error with $\alpha = 0.5$ and $h = 2^{-6}$ for $t \to 0$ for nonsmooth initial data (2b).

t	1e-3	1e-4	1e-5	1e-6	1e-7	1e-8	Rate
Case (2b)	2.56e-4	5.39e-4	1.15e-3	2.91e-3	6.77e-3	1.55e-2	≈−0.37(−0.375)



(a) $x = 1/2$ aligns with the grid when $h = 2^{-k}$

(b) $x = 1/2$ does not align with the grid for $h = 1/(2^k + 1)$

Fig. 3. Numerical results for example (2c) with $\alpha = 0.5$, $\beta = 0.2$ at $t = 0.005$, 0.01, 1.0, uniform mesh in time with $\tau = t/(5 \times 10^4)$.

Table 4
Numerical results for example (3a) with $\alpha = 0.5$ and $\beta = 0.2$ at $t = 1, 0.01, 0.001$, discretized on a uniform mesh $h = 2^{-k}$ and $\tau = t/(5 \times 10^4)$.

t	k	3	4	5	6	7	Rate
$t = 1$	L^2-norm	1.76e-3	4.40e-4	1.10e-4	2.71e-5	6.53e-6	≈2.01 (2.00)
	H^1-norm	4.72e-2	2.36e-2	1.18e-2	5.86e-3	2.86e-3	≈1.01 (1.00)
$t = 0.01$	L^2-norm	6.34e-4	1.59e-4	3.96e-5	9.82e-6	2.38e-6	≈2.01 (2.00)
	H^1-norm	1.89e-2	9.46e-3	4.72e-3	2.35e-3	1.15e-3	≈1.01 (1.00)
$t = 0.001$	L^2-norm	4.55e-4	1.15e-4	2.88e-5	1.15e-6	1.73e-6	≈2.02 (2.00)
	H^1-norm	1.45e-2	7.31e-3	3.66e-3	1.82e-3	8.88e-4	≈1.01 (1.00)

smooth from both sides of the support point and the finite element spaces X_h have good approximation property. When the singularity point $x = 1/2$ is not aligned with the grid, Fig. 3(b) indicates an $O(h^{3/2})$ and $O(h^{1/2})$ convergence rate for the L^2- and H^1-norm of the error, respectively, which agrees with our theory.

5.3. Inhomogeneous problems

Now we consider the inhomogeneous problem with $v \equiv 0$ on the unit interval $\Omega = (0, 1)$ and test the following two examples:

(3a) Nonsmooth data: $f(x,t) = (\chi_{[1/2,1]}(t) + 1)\chi_{[0,1/2]}(x)$. The jump at $x = 1/2$ leads to $f(t,\cdot) \notin \dot{H}^1(\Omega)$; nonetheless, for any $\epsilon > 0$, $f \in L^\infty(0, T; \dot{H}^{1/2-\epsilon}(\Omega))$.

(3b) Very weak data: $f(x,t) = (\chi_{[1/2,1]}(t) + 1)\delta_{1/2}(x)$ where f involves a Dirac $\delta_{1/2}(x)$-function concentrated at $x = 0.5$.

5.3.1. Numerical results for example (3a)

Since the errors are bounded independently of the time, cf. Theorem 3.3, we only present the errors in L^∞ in time, i.e., $\|e(t)\|_{L^2(\Omega)}$ and $\|\nabla e(t)\|_{L^2(\Omega)}$. In Table 4, we present the L^2- and H^1-error at $t = 1$, 0.01, and 0.001. The numerical results agree well with our theoretical predictions, i.e., $O(h^2)$ and $O(h)$ convergence rates for the L^2- and H^1-norms of the error, respectively.

Table 5
Numerical results for example (3b) with $\alpha = 0.5$ and $\beta = 0.2$ at $t = 1, 0.01, 0.001$, discretized on a uniform mesh $h = 1/(2^k + 1)$ and $\tau = t/(5 \times 10^4)$.

t	k	3	4	5	6	7	Rate
$t = 1$	L^2-norm	1.02e-2	4.01e-3	1.49e-3	5.35e-4	1.82e-4	≈1.49 (1.50)
	H^1-norm	3.24e-1	2.35e-1	1.65e-1	1.11e-1	6.94e-2	≈0.50 (0.50)
$t = 0.01$	L^2-norm	4.66e-3	1.91e-3	7.29e-4	2.64e-4	9.02e-5	≈1.45 (1.50)
	H^1-norm	1.54e-1	1.14e-1	8.16e-2	5.54e-2	3.47e-2	≈0.55 (0.50)
$t = 0.001$	L^2-norm	4.30e-3	1.83e-3	7.12e-4	2.61e-4	8.97e-5	≈1.45 (1.50)
	H^1-norm	1.47e-1	1.11e-1	8.05e-2	5.50e-2	3.45e-2	≈0.55 (0.50)

Table 6
Numerical results for example (3b) with $\alpha = 0.5$ and $\beta = 0.2$ at $t = 1, 0.01, 0.001$, discretized on a uniform mesh with $h = 2^{-k}$ and $\tau = t/(5 \times 10^4)$.

t	k	3	4	5	6	7	Rate
$t = 1$	L^2-norm	5.35e-4	1.34e-4	3.35e-5	8.31e-6	2.01e-6	≈2.01 (1.50)
	H^1-norm	1.49e-2	7.48e-3	3.74e-3	1.86e-3	9.07e-4	≈1.01 (0.50)
$t = 0.01$	L^2-norm	6.67e-4	1.67e-4	4.17e-5	1.04e-5	2.52e-6	≈2.03 (1.50)
	H^1-norm	2.56e-2	1.29e-2	6.44e-3	3.20e-3	1.56e-3	≈1.02 (0.50)
$t = 0.001$	L^2-norm	8.19e-4	2.08e-4	5.22e-5	1.30e-5	3.19e-6	≈2.02 (1.50)
	H^1-norm	3.96e-2	2.00e-2	1.00e-3	4.98e-3	2.45e-3	≈1.01 (0.50)

Table 7
Numerical results for (4a) with $\alpha = 0.5$ and $\beta = 0.2$ at $t = 0.1, 0.01, 0.001$, discretized on a uniform mesh, $h = 2^{-k}$ and $\tau = t/10^4$.

t	k	3	4	5	6	7	Rate
$t = 0.1$	L^2-norm	5.25e-3	1.35e-3	3.38e-4	8.24e-5	1.98e-5	≈2.06 (2.00)
	H^1-norm	9.10e-2	4.53e-2	2.25e-2	1.09e-2	4.99e-3	≈1.04 (1.00)
$t = 0.01$	L^2-norm	1.25e-2	3.23e-3	8.09e-4	1.97e-4	4.65e-5	≈2.05 (2.00)
	H^1-norm	2.18e-1	1.08e-1	5.35e-2	2.62e-2	1.27e-2	≈1.05 (1.00)
$t = 0.001$	L^2-norm	3.02e-2	7.84e-3	1.97e-3	4.81e-4	1.16e-4	≈2.03 (2.00)
	H^1-norm	5.30e-1	2.64e-1	1.31e-1	6.38e-2	3.14e-2	≈1.04 (1.00)

5.3.2. Numerical results for example (3b)

In Table 5 we show convergence rates at three different times, i.e., $t = 0.1$, 0.01, and 0.001. Here the mesh size h is chosen to be $h = 1/(2^k + 1)$, and thus the support of the Dirac δ-function does not align with the grid. The results indicate an $O(h^{1/2})$ and $O(h^{3/2})$ convergence rate for the H^1- and L^2-norm of the error, respectively, which agrees well with the theoretical prediction. If the Dirac δ-function is supported at a grid point, both L^2- and H^1-norms of the error exhibit a superconvergence phenomenon, i.e., orders $O(h^2)$ and $O(h)$, respectively, which are one half order higher than theoretical ones, cf. Table 6. This superconvergence phenomenon still awaits theoretical justifications.

5.4. Examples in two-dimension

In this part, we present three two-dimensional examples on the unit square $\Omega = (0, 1)^2$.

(4a) Nonsmooth initial data: $v = \chi_{(0,1/2)\times(0,1)}$ and $f \equiv 0$.
(4b) Very weak initial data: $v = \delta_\Gamma$ with Γ being the union of $\{1/4\} \times [1/4, 3/4] \cup [1/4, 3/4] \times \{3/4\}$ clockwise and $[1/4, 3/4] \times \{1/4\} \cup \{3/4\} \times [1/4, 3/4]$ counterclockwise. The duality is defined by $\langle \delta_\Gamma, \phi\rangle = \int_\Gamma \phi(s)\,ds$. By Hölder's inequality and the continuity of the trace operator from $\dot H^{1/2+\epsilon}(\Omega)$ to $L^2(\Gamma)$ [2], we deduce $\delta_\Gamma \in H^{-1/2-\epsilon}(\Omega)$.
(4c) Nonsmooth right hand side: $f(x,t) = (\chi_{[1/20,1/10]}(t) + 1)\chi_{(0,1/2)\times(0,1)}(x)$ and $v \equiv 0$.

To discretize the problem, we divide each direction into $N = 2^k$ equally spaced subintervals, with a mesh size $h = 1/N$ so that the domain $[0, 1]^2$ is divided into N^2 small squares. We get a symmetric mesh by connecting the diagonal of each small square.

The numerical results for example (4a) are shown in Table 7, which agree well with Theorem 3.2, with a rate $O(h^2)$ and $O(h)$, respectively, for the L^2- and H^1-norm of the error. Interestingly, for example (4b), both the L^2-norm and H^1-norm of the error exhibit superconvergence, cf. Table 8. The numerical results for example (4c) confirm the theoretical results; see Table 9. The solution profiles for examples (4b) and (4c) at $t = 0.1$ are shown in Fig. 4, from which the nonsmooth region of the solution can be clearly observed.

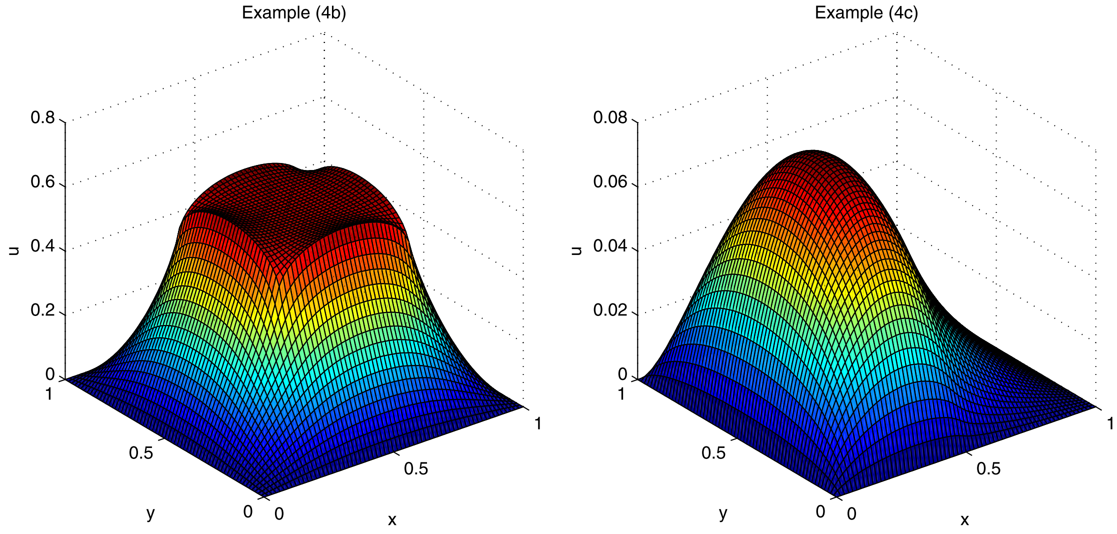


Fig. 4. Numerical solutions of examples (4b) and (4c) with $h = 2^{-6}$, $\alpha = 0.5$, $\beta = 0.2$ at $t = 0.1$.

Table 8
Numerical results for example (4b) with $\alpha = 0.5$ and $\beta = 0.2$ at $t = 0.1, 0.01, 0.001$ for a uniform mesh with $h = 2^{-k}$ and $\tau = t/10^4$.

t	k	3	4	5	6	7	Rate
$t = 0.1$	L^2-norm	1.18e-2	3.18e-3	8.41e-4	2.18e-4	5.41e-5	≈1.92 (1.50)
	H^1-norm	2.25e-1	1.13e-1	6.60e-2	3.40e-2	1.66e-2	≈0.92 (0.50)
$t = 0.01$	L^2-norm	2.82e-2	7.62e-3	2.28e-3	5.26e-4	1.25e-4	≈1.95 (2.00)
	H^1-norm	5.66e-1	3.09e-1	1.65e-1	8.52e-2	4.19e-2	≈0.94 (1.00)
$t = 0.001$	L^2-norm	6.65e-2	1.83e-3	4.98e-3	1.33e-3	3.30e-4	≈1.91 (2.00)
	H^1-norm	1.66e0	8.93e-1	4.75e-1	2.43e-1	1.21e-1	≈0.95 (1.00)

Table 9
Numerical results for example (4c) with $\alpha = 0.5$ and $\beta = 0.2$ at $t = 0.1, 0.01, 0.001$ for a uniform mesh with $h = 2^{-k}$ and $\tau = t/10^4$.

t	k	3	4	5	6	7	Rate
$t = 0.1$	L^2-norm	2.28e-3	5.86e-4	1.47e-4	3.58e-5	7.91e-6	≈2.07 (2.00)
	H^1-norm	3.97e-2	1.97e-2	9.77e-3	4.76e-3	2.13e-3	≈1.06 (1.00)
$t = 0.01$	L^2-norm	1.06e-3	2.73e-4	6.86e-5	1.67e-6	3.70e-7	≈2.06 (2.00)
	H^1-norm	1.85e-2	9.18e-3	4.56e-3	2.22e-3	9.94e-4	≈1.06 (1.00)
$t = 0.001$	L^2-norm	8.66e-4	2.28e-4	5.75e-5	1.40e-6	3.11e-7	≈2.04 (2.00)
	H^1-norm	1.56e-2	7.82e-3	3.88e-3	1.90e-3	8.47e-4	≈1.05 (1.00)

6. Concluding remarks

In this work, we have developed a simple numerical scheme based on the Galerkin finite element method for a multi-term time fractional diffusion equation which involves multiple Caputo fractional derivatives in time. A complete error analysis of the space semidiscrete Galerkin scheme is provided. The theory covers the practically very important case of nonsmooth initial data and right hand side. The analysis relies essentially on some new regularity results of the multi-term time fractional diffusion equation. Further, we have developed a fully discrete scheme based on a finite difference discretization of the Caputo fractional derivatives. The stability and error estimate of the fully discrete scheme were established, provided that the solution is smooth. The extensive numerical experiments in one- and two-dimension fully confirmed our convergence analysis: the empirical convergence rates agree well with the theoretical predictions for both smooth and nonsmooth data.

Acknowledgements

The authors are grateful to the anonymous referees for their constructive comments. The research of B. Jin has been supported by NSF Grant DMS-1319052, R. Lazarov was supported in parts by NSF Grant DMS-1016525 and also by Award

No. KUS-C1-016-04, made by King Abdullah University of Science and Technology (KAUST), and Y. Liu was supported by the Program for Leading Graduate Schools, MEXT, Japan.

References

[1] E.E. Adams, L.W. Gelhar, Field study of dispersion in a heterogeneous aquifer: 2. Spatial moments analysis, Water Res. Research 28 (12) (1992) 3293–3307.
[2] R. Adams, J. Fournier, Sobolev Spaces, Elsevier/Academic Press, Amsterdam, 2003.
[3] E. Bazhlekova, Properties of the fundamental and the impulse-response solutions of multi-term fractional differential equations, in: Complex Analysis and Applications '13, Sofia, 31 Oct.–2 Nov. 2013, 2013, pp. 55–64.
[4] H. Brunner, L. Ling, M. Yamamoto, Numerical simulations of 2D fractional subdiffusion problems, J. Comput. Phys. 229 (18) (2010) 6613–6622.
[5] E. Casas, C. Clason, K. Kunisch, Parabolic control problems in measure spaces with sparse solutions, SIAM J. Control Optim. 51 (1) (2013) 28–63.
[6] A.M.A. El-Sayed, I.L. El-Kalla, E.A.A. Ziada, Analytical and numerical solutions of multi-term nonlinear fractional orders differential equations, Appl. Numer. Math. 60 (8) (2010) 788–797.
[7] Z.-J. Fu, W. Chen, H.-T. Yang, Boundary particle method for Laplace transformed time fractional diffusion equations, J. Comput. Phys. 235 (2013) 52–66.
[8] S.B. Hadid, Y.F. Luchko, An operational method for solving fractional differential equations of an arbitrary real order, Panamer. Math. J. 6 (1) (1996) 57–73.
[9] H. Jiang, F. Liu, I. Turner, K. Burrage, Analytical solutions for the multi-term time–space Caputo–Riesz fractional advection–diffusion equations on a finite domain, J. Math. Anal. Appl. 389 (2) (2012) 1117–1127.
[10] B. Jin, R. Lazarov, J. Pasciak, Z. Zhou, Error analysis of semidiscrete finite element methods for inhomogeneous time-fractional diffusion, IMA J. Numer. Anal. (2013), in press.
[11] B. Jin, R. Lazarov, J. Pasciak, Z. Zhou, Galerkin FEM for fractional order parabolic equations with initial data in H^{-s}, $0 \le s \le 1$, in: Proc. 5th Conf. Numer. Anal. Appl., Springer, 2013, pp. 24–37.
[12] B. Jin, R. Lazarov, Z. Zhou, Error estimates for a semidiscrete finite element method for fractional order parabolic equations, SIAM J. Numer. Anal. 51 (1) (2013) 445–466.
[13] B. Jin, X. Lu, Numerical identification of a Robin coefficient in parabolic problems, Math. Comp. 81 (2012) 1369–1398.
[14] B. Jin, W. Rundell, An inverse problem for a one-dimensional time-fractional diffusion problem, Inverse Problems 28 (7) (2012) 075010, 19 pp.
[15] J. Katsikadelis, Numerical solution of multi-term fractional differential equations, ZAMM Z. Angew. Math. Mech. 89 (7) (2009) 593–608.
[16] J.F. Kelly, R.J. McGough, M.M. Meerschaert, Analytical time-domain Green's functions for power-law media, J. Acoust. Soc. Am. 124 (5) (2008) 2861–2872.
[17] A. Kilbas, H. Srivastava, J. Trujillo, Theory and Applications of Fractional Differential Equations, Elsevier, Amsterdam, 2006.
[18] L. Li, D. Xu, Alternating direction implicit Galerkin finite element method for the two-dimensional time fractional evolution equation, Numer. Math. Theory Methods Appl. 7 (1) (2014) 41–57.
[19] X. Li, C. Xu, A space–time spectral method for the time fractional diffusion equation, SIAM J. Numer. Anal. 47 (3) (2009) 2108–2131.
[20] X. Li, C. Xu, Existence and uniqueness of the weak solution of the space–time fractional diffusion equation and a spectral method approximation, Commun. Comput. Phys. 8 (5) (2010) 1016–1051.
[21] Z. Li, Y. Liu, M. Yamamoto, Initial-boundary value problems for multi-term time-fractional diffusion equations with positive constant coefficients, preprint, arXiv:1312.2112, 2013.
[22] Z. Li, M. Yamamoto, Initial-boundary value problems for linear diffusion equations with multiple time-fractional derivatives, preprint, arXiv:1306.2778, 2013.
[23] Y. Lin, C. Xu, Finite difference/spectral approximations for the time-fractional diffusion equation, J. Comput. Phys. 225 (2) (2007) 1533–1552.
[24] F. Liu, M.M. Meerschaert, R.J. McGough, P. Zhuang, X. Liu, Numerical methods for solving the multi-term time-fractional wave-diffusion equation, Frac. Cal. Appl. Anal. 16 (1) (2013) 9–25.
[25] Y. Luchko, Initial-boundary-value problems for the generalized multi-term time-fractional diffusion equations, J. Math. Anal. Appl. 374 (2) (2011) 538–548.
[26] Y. Luchko, R. Gorenflo, An operational method for solving fractional differential equations with the Caputo derivatives, Acta Math. Vietnam. 24 (2) (1999) 207–233.
[27] W. McLean, 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) (2010) 208–230.
[28] R. Metzler, J. Klafter, I.M. Sokolov, Anomalous transport in external fields: continuous time random walks and fractional diffusion equations extended, Phys. Rev. E 58 (2) (1998) 1621–1633.
[29] 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) (2011) 719–739.
[30] K. Mustapha, W. McLean, Superconvergence of a discontinuous Galerkin method for fractional diffusion and wave equations, SIAM J. Numer. Anal. 51 (1) (2013) 491–515.
[31] R. Nigmatulin, The realization of the generalized transfer equation in a medium with fractal geometry, Phys. Stat. Sol. B 133 (1986) 425–430.
[32] K. Sakamoto, M. Yamamoto, Initial value/boundary value problems for fractional diffusion-wave equations and applications to some inverse problems, J. Math. Anal. Appl. 382 (1) (2011) 426–447.
[33] R. Schumer, D.A. Benson, M.M. Meerschaert, B. Baeumer, Fractal mobile/immobile solute transport, Water Res. Research 39 (10) (2003) 1296, 13 pp.
[34] Z.-Z. Sun, X. Wu, A fully discrete difference scheme for a diffusion-wave system, Appl. Numer. Math. 56 (2) (2006) 193–209.
[35] V. Thomée, Galerkin Finite Element Methods for Parabolic Problems, Springer Series in Computational Mathematics, vol. 25, Springer-Verlag, Berlin, 2006.
[36] M. Zayernouri, G.E. Karniadakis, Exponentially accurate spectral and spectral element methods for fractional ODEs, J. Comput. Phys. 257 (Part A) (2014) 460–480.
[37] Y.-N. Zhang, Z.-Z. Sun, H.-W. Wu, Error estimates of Crank–Nicolson-type difference schemes for the subdiffusion equation, SIAM J. Numer. Anal. 49 (6) (2011) 2302–2322.
[38] J. Zhao, J. Xiao, Y. Xu, Stability and convergence of an effective finite element method for multiterm fractional partial differential equations, Abstr. Appl. Anal. (2013) 857205, 10 pp.