Numerical Methods for Time-Fractional Evolution Equations with Nonsmooth Data: a Concise Overview

Bangti Jin

Department of Computer Science, University College London, Gower Street, London, WC1E 2BT, UK.

Raytcho Lazarov

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

Zhi Zhou

Department of Applied Mathematics, The Hong Kong Polytechnic University, Kowloon, Hong Kong, P.R. China

Abstract

Over the past few decades, there has been substantial interest in evolution equations that involve a fractional-order derivative of order $\alpha \in (0, 1)$ in time, commonly known as subdiffusion, due to their many successful applications in engineering, physics, biology and finance. Thus, it is of paramount importance to develop and to analyze efficient and accurate numerical methods for reliably simulating such models, and the literature on the topic is vast and fast growing. The present paper gives a concise overview on numerical schemes for the subdiffusion model with nonsmooth problem data, which are important for the numerical analysis of many problems arising in optimal control, inverse problems and stochastic analysis. We focus on the following topics of the subdiffusion model: regularity theory, Galerkin finite element discretization in space, time-stepping schemes (including convolution quadrature and L1 type schemes), and space-time variational formulations, and compare the results with that for standard parabolic problems. Further, these aspects are showcased with illustrative numerical experiments and complemented with perspectives and pointers to relevant literature.

Keywords: time-fractional evolution, subdiffusion, nonsmooth solution, finite element method, time-stepping, initial correction, error estimates, space-time formulation

1. Introduction

Diffusion is one of the most prominent transport mechanisms found in nature. The classical diffusion model $\partial_t u - \Delta u = f$, which employs a first-order derivative $\partial_t u$ in time and the Laplace operator Δu in space, rests on the assumption that the particle motion is Brownian. One of the distinct features of Brownian motion is a linear growth of the mean squared particle displacement with the time t. Over the

*Corresponding author

Email addresses: b.jin@ucl.ac.uk, bangti.jin@gmail.com (Bangti Jin), lazarov@math.tamu.edu (Raytcho Lazarov), zhizhou@polyu.edu.hk (Zhi Zhou)

last few decades, a long list of experimental studies indicates that the Brownian motion assumption may not be adequate for accurately describing some physical processes, and the mean squared displacement can grow either sublinearly or superlinearly with the time t, which are known as subdiffusion and superdiffusion, respectively, in the literature. These experimental studies cover an extremely broad and diverse range of important practical applications in engineering, physics, biology and finance, including electron transport in Xerox photocopier [1], visco-elastic materials [2, 3], thermal diffusion in fractal domains [4], column experiments [5] and protein transport in cell membrane [6] etc. The underlying stochastic process for subdiffusion and superdiffusion is usually given by continuous time random walk and Lévy process, respectively, and the corresponding macroscopic model for the probability density function of the particle appearing at certain time instance t and location x is given by a diffusion model with a fractional-order derivative in time and in space, respectively. We refer interested readers to the comprehensive surveys [7, 8, 9] and the monograph [10] for an extensive list of practical applications and physical modeling in engineering, physic, and biology.

The present work surveys rigorous numerical methods for subdiffusion. The prototypical mathematical model for subdiffusion is as follows. Let $\Omega \subset \mathbb{R}^d$ $(d = 1, 2, 3)$ be a convex polygonal domain with a boundary $\partial\Omega$, and consider the following fractional-order parabolic problem for the function $u(x, t)$:

$$\begin{cases} \partial_t^\alpha u(x,t) - \Delta u(x,t) = f(x,t) & (x,t) \in \Omega \times (0,T], \\ u(x,t) = 0 & (x,t) \in \partial\Omega \times (0,T], \\ u(x,0) = v(x) & x \in \Omega, \end{cases} \tag{1.1}$$

where $T > 0$ is a fixed final time, $f \in L^\infty(0, T; L^2(\Omega))$ and $v \in L^2(\Omega)$ are given source term and initial data, respectively, and Δ is the Laplace operator in space. Here $\partial_t^\alpha u(t)$ denotes the Caputo fractional derivative in time t of order $\alpha \in (0, 1)$ [11, p. 70]

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

where $\Gamma(z)$ is the Gamma function defined by

$$\Gamma(z) = \int_0^\infty s^{z-1} e^{-s} \mathrm{d}s, \quad \Re(z) > 0.$$

It is named after geophysicist Michele Caputo [2], who first introduced it for describing the stress-strain relation in linear elasticity, although it was predated by the work of Armenian mathematician Mkhitar Djrbashian [12]. So more precisely, it should be called Djrbashian-Caputo fractional derivative. Note that the fractional derivative $\partial_t^\alpha u$ recovers the usual first-order derivative $u'(t)$ as $\alpha \to 1^-$, provided that the function u is sufficiently smooth [13, p. 100]. Thus the model (1.1) can be viewed as a fractional analogue of the classical parabolic equation. Therefore, it is natural and instructive to compare its analytical and numerical properties with that of standard parabolic problems.

Remark 1.1. *All the discussions below extend straightforwardly to a general second-order coercive and symmetric elliptic differential operator, given by* $\nabla \cdot (a(x)\nabla u(x)) - q(x)u(x)$ *with* $q \geq 0$ *a.e.*

Motivated by its tremendous success in the mathematical modeling of many important physical problems, over the last two decades there has been an explosive growth in the numerical algorithms, and mathematical analysis of the subdiffusion model. More recently this interest has been extended to related research areas, e.g., optimal control, inverse problems and stochastic fractional models. The literature on the topic is vast, and the list is still fast growing in the community of scientific and engineering computation, and more recently also in the community of numerical analysis; see, e.g., the recent special issues on the topic at *Journal of Computational Physics* [14] and *Computational Methods in Applied Mathematics* [15], for some important progress in the area of numerical methods for fractional evolution equations.

It is impossible to survey all important and relevant works in a short review. Instead, in this paper, we aim at only reviewing relevant works on the numerical methods for the subdiffusion model (1.1) with nonsmooth problem data in the sense that the initial data v belongs only to $L^2(\Omega)$ or the source term f is not compatible with the initial data or boundary condition. First, this choice allows us to highlight some distinct features common to many nonlocal models, especially how the smoothness of the problem data influences the solution and the corresponding numerical methods. It is precisely these features that pose substantial new mathematical and computational challenges when compared with standard parabolic problems, and extra care has to be taken when developing and analyzing relevant numerical methods. In particular, since the solution operators of the fractional model have limited smoothing property, a numerical method that requires high regularity of the exact solution will impose severe restrictions (compatibility conditions) on the data and generally does not work well and thus substantially limits its scope of potential applications. Finally, nonsmooth data analysis is fundamental to the rigorous error analysis of areas related to various applications, e.g., optimal control, inverse problems, and stochastic fractional diffusion (see, e.g., [16, 17, 18, 19]).

Amongst the numerous possible choices, we shall focus the review on the following four topics:

(i) Regularity theory in Sobolev spaces;

(ii) Spatial discretization via finite element methods (FEMs), e.g., standard Galerkin, lumped mass and finite volume element methods;

(iii) Temporal discretization via time-stepping schemes;

(iv) Space-time formulations (Galerkin or Petrov-Galerkin type).

In each aspect, we describe some representative results and leave most of technical proofs to the references. Further, in order to illustrate the distinct features of fractional diffusion, we compare the results with that for standard parabolic problems (see, e.g., [20]) and give some numerical illustrations of the mathematical theory. Finally, we complement each part with comments on future research problems and further references. The goal of the overview is to give readers a flavor of the numerical analysis of nonlocal problems and potential pitfalls in developing efficient numerical methods. The main contributions of the review

include summarizing known results that are otherwise dispersed in the vast literature (often in different notation), and offering some perspectives on unsolved issues.

Nonlocal models involving nonlocality in either time or space have evolved into a huge and extremely active and fruitful research area in recent years, due to the joint efforts of researchers in mathematical analysis, partial differential equation theory, numerical analysis, stochastic analysis, and many applied disciplines. The subdiffusion model (1.1) represents only one example in the zoology of nonlocal problems. We refer interested readers to the excellent surveys for other nonlocal problems and applications, namely, on problem involving fractional (spectral and integral) Laplacian [21, 22], on inverse problems with anomalous diffusion [18, 23], on regularity theory of nonlocal elliptic equations in bounded domains [24], and on nonlocal problems arising in peridynamics [25]. For an early overview on the numerical methods for fractional-order ordinary differential equations, we refer to the paper [26].

The rest of the paper is organized as follows. For the model (1.1), in Section 2, we describe the regularity theory and in Sections 3 and 4, we discuss the finite element methods and two popular classes of time stepping schemes, i.e., convolution quadrature and L1 type schemes, respectively. Then, in Section 5, we describe two space-time formulations for problem (1.1) with $v = 0$. We conclude the overview with some further discussions in Section 6. Throughout, the discussions focus on the case of nonsmooth problem data, and only references directly relevant are given. Obviously, the list of references is not meant to be complete in any sense, and strongly biased by the personal taste and limited by the knowledge of the authors. Throughout, the notation c denotes a generic constant which may change at each occurrence, but it is always independent of the discretization parameters h and τ etc. In the paper we use the standard notation on Sobolev spaces (see, e.g., [27]).

2. Regularity of the solution

First, we describe some regularity results for the model (1.1), which are crucial for rigorous numerical analysis. To this end, we need suitable function spaces. The most convenient one for our purpose is the space $\dot{H}^s(\Omega)$ defined as follows [20, Chapter 3]. Let $\{\lambda_j\}_{j=1}^\infty$ and $\{\varphi_j\}_{j=1}^\infty$ be respectively the eigenvalues (ordered nondecreasingly with multiplicity counted) and the $L^2(\Omega)$-orthonormal eigenfunctions of the negative Laplace operator $-\Delta$ on the domain Ω with a zero Dirichlet boundary condition. Then $\{\varphi_j\}_{j=1}^\infty$ forms an orthonormal basis in $L^2(\Omega)$. For any real number $s \geq -1$, we denote by $\dot{H}^s(\Omega)$ the Hilbert space consisting of the functions of the form

$$v = \sum_{j=1}^{\infty} \lambda_j^{\frac{s}{2}} \langle v, \varphi_j \rangle \varphi_j,$$

where $\langle \cdot, \cdot \rangle$ denotes the duality pairing between $H^{-1}(\Omega)$ and $H_0^1(\Omega)$, and it coincides with the usual $L^2(\Omega)$ inner product $(\cdot, \cdot)$ if the function $v \in L^2(\Omega)$. The induced norm $\|\cdot\|_{\dot{H}^s(\Omega)}$ is defined by

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

Then, $\|v\|_{\dot{H}^0(\Omega)} = \|v\|_{L^2(\Omega)} = (v,v)^{\frac{1}{2}}$ is the norm in $L^2(\Omega)$ and $\|v\|_{\dot{H}^{-1}(\Omega)} = \|v\|_{H^{-1}(\Omega)}$ is the norm in $H^{-1}(\Omega)$. Besides, it is easy to verify that $\|v\|_{\dot{H}^1(\Omega)} = \|\nabla v\|_{L^2(\Omega)}$ is also an equivalent norm in $H_0^1(\Omega)$ and $\|v\|_{\dot{H}^2(\Omega)} = \|\Delta v\|_{L^2(\Omega)}$ is equivalent to the norm in $H^2(\Omega)\cap H_0^1(\Omega)$, provided the domain Ω is convex [20, Section 3.1]. Note that the spaces $\dot{H}^s(\Omega)$, $s \geq -1$, form a Hilbert scale of interpolation spaces. Motivated by this, we denote $H_0^s(\Omega)$ to be the norm on the interpolation scale between $H_0^1(\Omega)$ and $L^2(\Omega)$ when s is in $[0,1]$ and $H_0^s(\Omega)$ to be the norm on the interpolation scale between $L^2(\Omega)$ and $H^{-1}(\Omega)$ when s is in $[-1,0]$. Then, the norms $\|\cdot\|_{H_0^s(\Omega)}$ and $\|\cdot\|_{\dot{H}^s(\Omega)}$ are equivalent for $s\in[-1,0]$ by interpolation.

There are several different ways to analyze problem (1.1). We outline one approach to derive regularity results by means of Laplace transform below. We denote the Laplace transform of a function $f:(0,\infty)\to\mathbb{R}$ by $\widehat{f}$ below. The starting point of the analysis is the following identity on the Laplace transform $\widehat{\partial_t^\alpha}(z)$ of the Caputo fractional derivative $\partial_t^\alpha u(t)$, $0<\alpha<1$ [11, Lemma 2.24, p. 98]

$$\widehat{\partial_t^\alpha u}(z) = z^\alpha \widehat{u}(z) - z^{\alpha-1}u(0).$$

By viewing $u(t)$ as a vector-valued function, applying Laplace transform to both sizes of (1.1) yields

$$z^\alpha\widehat{u}(z) - \Delta\widehat{u}(z) = \widehat{f}(z) + z^{\alpha-1}u(0),$$

i.e.,

$$\widehat{u}(z) = (z^\alpha - \Delta)^{-1}(\widehat{f}(z) + z^{\alpha-1}u(0)).$$

By inverse Laplace transform and the convolution rule, the solution $u(t)$ can be formally represented by a Duhamel type formula:

$$u(t) = F(t)v + \int_0^t E(t-s)f(s)\mathrm{d}s, \tag{2.1}$$

where the solution operators $F(t)$ and $E(t)$ are respectively defined by

$$F(t) := \frac{1}{2\pi\mathrm{i}}\int_{\Gamma_{\theta,\delta}} e^{zt}z^{\alpha-1}(z^\alpha-\Delta)^{-1}\,\mathrm{d}z \quad\text{and}\quad E(t) := \frac{1}{2\pi\mathrm{i}}\int_{\Gamma_{\theta,\delta}} e^{zt}(z^\alpha-\Delta)^{-1}\,\mathrm{d}z,$$

with integration over a contour $\Gamma_{\theta,\delta}$ in the complex plane $\mathbb{C}$ (oriented counterclockwise), defined by

$$\Gamma_{\theta,\delta} = \{z\in\mathbb{C}:|z|=\delta, |\arg z|\leq\theta\}\cup\{z\in\mathbb{C}: z=\rho e^{\pm\mathrm{i}\theta}, \rho\geq\delta\}.$$

Throughout, we fix $\theta\in(\frac{\pi}{2},\pi)$ so that $z^\alpha\in\Sigma_{\alpha\theta}\subset\Sigma_\theta := \{0\neq z\in\mathbb{C}:\arg(z)\leq\theta\}$, for all $z\in\Sigma_\theta$. Recall the following resolvent estimate for the Laplacian Δ with a homogenous Dirichlet boundary condition [28, Example 3.7.5 and Theorem 3.7.11]:

$$\|(z-\Delta)^{-1}\|\leq c_\phi|z|^{-1}, \quad \forall z\in\Sigma_\phi,\ \forall\,\phi\in(0,\pi), \tag{2.2}$$

where $\|\cdot\|$ denotes the operator norm from $L^2(\Omega)$ to $L^2(\Omega)$.

Equivalently, using the eigenfunction expansion $\{(\lambda_j,\varphi_j)\}_{j=1}^\infty$, these operators can be expressed as

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

Here $E_{\alpha,\beta}(z)$ is the two-parameter Mittag-Leffler function defined by [11, Section 1.8, pp. 40-45]

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

The Mittag-Leffler function $E_{\alpha,\beta}(z)$ is a generalization of the familiar exponential function e^z appearing in normal diffusion, and it can be evaluated efficiently (via contour integral and asymptotic expansion) [29, 30]. Since the solution operators involve only $E_{\alpha,\beta}(z)$ with z being a negative real argument, the following decay behavior $E_{\alpha,\beta}(z)$ is crucial to the smoothing properties of $F(t)$ and $E(t)$: for any $\alpha \in (0,1)$, the function $E_{\alpha,1}(-\lambda t^\alpha)$ decays only polynomially like $t^{-\alpha}$ as $t \to \infty$ [11, equation (1.8.28), p. 43], which contrasts sharply with the exponential decay for $e^{-\lambda t}$ appearing in normal diffusion. These important features directly translate into the limited smoothing property in both space and time for the solution operators $E(t)$ and $F(t)$.

Next, we state a few regularity results. The proof of these results can be found in, e.g., [31, 32, 33] (see, e.g., [33, Theorem 2.1 and Corollary 2.7] for part (i), and [31], [32, Theorem 3], and [34, Theorem 2.3] for part (ii)). All these results can be proven directly using the resolvent estimate (2.2).

Theorem 2.1. *Let $u(t)$ be the solution to problem (1.1). Then the following statements hold.*

(i) *If $v \in \dot{H}^s(\Omega)$ with $s \in (-1,2]$ and $f = 0$, then $u(t) \in \dot{H}^{s+2}(\Omega)$ and*

$$\|\partial_t^{(m)} u(t)\|_{\dot{H}^p(\Omega)} \le c t^{\frac{(s-p)\alpha}{2}-m} \|v\|_{\dot{H}^s(\Omega)}$$

with $0 \le p - s \le 2$ and any integer $m \ge 0$.

(ii) *If $v = 0$ and $f \in L^p(0,T;L^2(\Omega))$ with $1 < p < \infty$, then there holds*

$$\|u\|_{L^p(0,T;\dot{H}^2(\Omega))} + \|\partial_t^\alpha u\|_{L^p(0,T;L^2(\Omega))} \le c\|f\|_{L^p(0,T;L^2(\Omega))}.$$

Moreover, if $f \in L^\infty(0,T;L^2(\Omega))$, we have for any $\epsilon \in (0,1)$

$$\|u(t)\|_{\dot{H}^{2-\epsilon}(\Omega)} \le c\epsilon^{-1} t^{\epsilon\alpha} \|f\|_{L^\infty(0,t;L^2(\Omega))}.$$

(iii) *If $v = 0$ and $f \in C^{m-1}([0,T];L^2(\Omega))$, $\int_0^t (t-s)^{\alpha-1} \|\partial_s^{(m)} f(s)\|_{L^2(\Omega)} \mathrm{d}s < \infty$, then there holds*

$$\|\partial_t^{(m)} u(t)\|_{L^2(\Omega)} \le c \sum_{k=0}^{m-1} t^{\alpha-(m-k)} \|\partial_t^{(k)} f(0)\|_{L^2(\Omega)} + \int_0^t (t-s)^{\alpha-1} \|\partial_s^{(m)} f(s)\|_{L^2(\Omega)} \mathrm{d}s.$$

The estimate in Theorem 2.1(i) indicates that for homogeneous problems, the solution $u(t)$ is smooth in time $t > 0$ (actually analytic in a sector in the complex plane $\mathbb{C}$ [33, Theorem 2.1]), but has a weak singularity around $t = 0$. The strength of the singularity depends on the regularity of the initial data v: the smoother is v (measured in the space $\dot{H}^s(\Omega)$), the less singular is the solution u at the initial layer. Interestingly, even if the initial data v is very smooth, the solution u is generally not very smooth in time

in the fractional case, which also differs from the standard parabolic case. By now, it is well known that smooth solutions are produced by a small class of data [35]. The condition $0 \le p - s \le 2$ in Theorem 2.1(i) represents an essential restriction on the smoothing property of the solution operator $F(t)$ in space of order two. This restriction contrasts sharply with that for the standard diffusion equation: the following estimate

$$\|\partial_t^{(m)} u(t)\|_{\dot{H}^p(\Omega)} \le c t^{\frac{s-p}{2}-m} \|v\|_{\dot{H}^s(\Omega)}$$

holds for any $t > 0$ and any $p \ge s, m \ge 0$ (see, e.g. [20, Lemma 3.2, p. 39]). This means that the solution operator for standard parabolic problems is infinitely smoothing in space, as long as $t > 0$. The limited smoothing property in space of the model (1.1) represents one very distinct feature, which is generic for many other nonlocal (in time) models.

The first inequality in Theorem 2.1(ii) is often known as maximal L^p regularity, which is very useful in the mathematical and numerical analysis of nonlinear problems (see, e.g., [36, 37] for standard parabolic problems and [38] for subdiffusion). Theorem 2.1(iii) asserts that the temporal regularity of the solution $u(t)$ is essentially determined by that of the right hand side f. The solution $u(t)$ usually still has weak singularity near $t = 0$, even for a very smooth source term f, which differs again dramatically from standard parabolic problems. In order to have the mth order temporal differentiability in time t, for the inhomogeneous problem, it is necessary to impose the following (rather restrictive) compatibility conditions: $\partial_t^{(k)} f(0) = 0$, $k = 0, \ldots, m-1$. In the numerical analysis, it is important to take into account the initial singularity of the solution $u(t)$, which represents one of the main challenges in developing robust numerical methods.

Now we illustrate the results for the homogeneous problem.

Example 2.1. *Consider problem* (1.1) *on the unit interval* $\Omega = (0,1)$ *with*

(i) $v = \sin(\pi x)$ *and* $f = 0$*;*

(ii) $v = \delta_{0.5}(x)$*, with* $\delta_{0.5}(x)$ *the Dirac* δ *function concentrated at* $x = 0.5$*, and* $f = 0$*.*

The solution $u(t)$ *in case* (i) *is given by* $u(t) = E_{\alpha,1}(-\pi^2 t^\alpha)\sin(\pi x)$*. Since* $\sin(\pi t)$ *is a Dirichlet eigenfunction of the negative Laplacian* $-\Delta$ *on* Ω*, it is easy to see that for any* $s \ge 0$*,* $v \in \dot{H}^s(\Omega)$*, but the solution* $u(t)$ *has limited temporal regularity for any* $\alpha \in (0,1)$*: as* $t \to 0$*,* $E_\alpha(-\pi^2 t^\alpha) \sim 1 - \frac{\pi^2}{\Gamma(\alpha+1)} t^\alpha$*, which is continuous at* $t = 0$ *but has an unbounded first-order derivative. This observation clearly reflects the inherently limited smoothing property in time* t *of problem* (1.1)*. It contrasts sharply with the standard parabolic case,* $\alpha = 1$*, for which the solution* $u(t)$ *is given explicitly by* $u(t) = e^{-\pi^2 t}\sin(\pi x)$ *and is* $C^\infty[0,T]$ *in time. In Fig. 1, we show the solution profiles for* $\alpha = 0.5$ *and* $\alpha = 1$ *at two different time instances for case* (ii)*. Observe that the solution profile for* $\alpha = 1$ *decays much faster than that for* $\alpha = 0.5$*. For any* $t > 0$*, the solution* $u(t)$ *is very smooth in space for* $\alpha = 1$*, but it remains nonsmooth for* $\alpha = 0.5$*. In the latter case, the kink at* $x = 0.5$ *in the plot shows clearly the limited spatial smoothing property of the solution operator* $F(t)$*, and it remains no matter how long the problem evolves.*

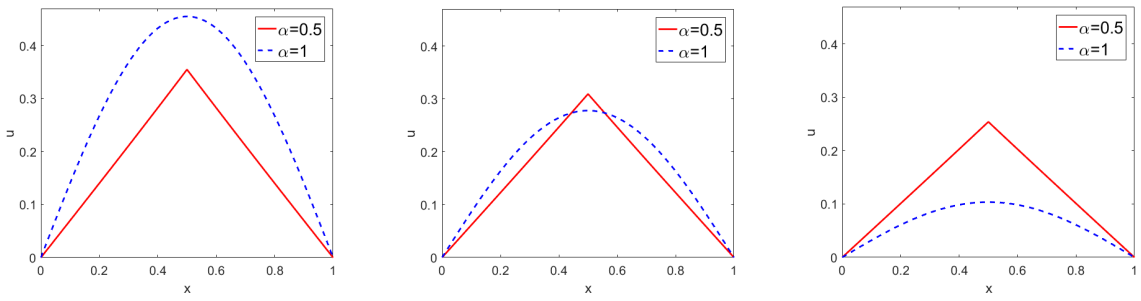

Figure 1: The solution profiles for Example 2.1(ii) at three time instances for $\alpha = 0.5$ and 1.

The analytical theory of problem (1.1) has been developed successfully in the last two decades, e.g., [31, 39, 40, 41, 42, 33, 43, 38, 44, 45, 46, 47]; see also the monograph [48] for closely related evolutionary integral equations. Eidelman and Kochubei [39] derived fundamental solutions to problem in the whole space using Fox H-functions, and derived various estimates, see also [49, 50];. Luchko [40] studied the existence and uniqueness of a strong solution. Sakamoto and Yamamoto [33] analyzed the problem by means of separation of variables, reducing it to an infinite system of fractional-order ODEs, studied the existence and uniqueness of weak solutions, and proved various regularity results including the asymptotic behavior of the solution for $t \to 0$ and $t \to \infty$. We note that the Laplace transform technique described above is essentially along the same line of reasoning. The important issue of properly interpreting the initial condition (for α close to zero) was discussed in [50, 51].

It is worth noting that techniques like separation of variables and Laplace transform are most convenient for analyzing time-independent elliptic operators. For time-dependent elliptic operators or nonlinear problems, e.g., time-dependent diffusion coefficients and Fokker-Planck equation, energy arguments [52] or perturbation arguments [53] can be used to show existence and uniqueness of the solution. However, the slightly more refined stability estimates, needed for numerical analysis of nonsmooth problem data, often do not directly follow and have to be derived separately. This represents one of the main obstacles in extending the numerical methods and their analysis below for the model problem (1.1) to these important classes of applied problems.

3. Spatially semidiscrete approximation

Now we describe several spatially semidiscrete finite element schemes for problem (1.1) using the standard notation from the classical monograph [20]. Semidiscrete methods are usually not directly implementable and rarely used in practical computations, but they are important for understanding the role of the regularity of problem data and also for the analysis of some space-time formulations and spectral, and rational approximations. 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 the mesh size h denoting the maximum diameter

of the elements. An approximate solution u_h is then sought in the finite element space $X_h \equiv X_h(\Omega)$ of continuous piecewise linear functions over the triangulation $\mathcal{T}_h$, defined by

$$X_h = \{\chi \in C_0(\Omega) : \ \chi|_\tau \in P_1(\tau), \quad \forall \tau \in \mathcal{T}_h\}. \tag{3.1}$$

where $P_1(\tau)$ denotes the space of linear polynomials on τ. To describe the schemes, we need the $L^2(\Omega)$ projection $P_h : L^2(\Omega) \to X_h$ and Ritz projection $R_h : \dot{H}^1(\Omega) \to X_h$, respectively, defined by (recall that $(\cdot,\cdot)$ denotes the $L^2(\Omega)$ inner product)

$$(P_h\psi, \chi) = (\psi, \chi) \quad \forall \chi \in X_h, \psi \in L^2(\Omega),$$
$$(\nabla R_h\psi, \nabla\chi) = (\nabla\psi, \nabla\chi) \quad \forall \chi \in X_h, \psi \in \dot{H}^1(\Omega).$$

Then by means of duality, the operator P_h can be boundedly extended to $\dot{H}^s(\Omega)$, $s \in [-1, 0]$. The following approximation properties of R_h and P_h are well known:

$$\|P_h\psi - \psi\|_{L^2(\Omega)} + h\|\nabla(P_h\psi - \psi)\|_{L^2(\Omega)} \le ch^q\|\psi\|_{H^q(\Omega)} \quad \forall \psi \in \dot{H}^q(\Omega), q = 1, 2, \tag{3.2}$$
$$\|R_h\psi - \psi\|_{L^2(\Omega)} + h\|\nabla(R_h\psi - \psi)\|_{L^2(\Omega)} \le ch^q\|\psi\|_{H^q(\Omega)} \quad \forall \psi \in \dot{H}^q(\Omega), q = 1, 2. \tag{3.3}$$

By multiplying both sides of equation (1.1) by a test function $\varphi \in H_0^1(\Omega)$, integrating over the domain Ω and then applying integration by parts formula yield the following weak formulation of problem (1.1): find $u(t) \in H_0^1(\Omega)$ for $t > 0$ such that

$$(\partial_t^\alpha u(t), \varphi) + a(u(t), \varphi) = (f, \varphi), \quad \forall \varphi \in H_0^1(\Omega), \quad t > 0, \text{with } u(0) = v, \tag{3.4}$$

where $a(u, \varphi) = (\nabla u, \nabla\varphi)$ for $u, \varphi \in H_0^1(\Omega)$ denotes the bilinear form for the elliptic operator $A = -\Delta$ (with a zero Dirichlet boundary condition). Then the spatially semidiscrete approximation of problem (1.1) is to find $u_h(t) \in X_h$ such that

$$[\partial_t^\alpha u_h(t), \chi] + a(u_h(t), \chi) = (f, \chi), \quad \forall \chi \in X_h, \ t > 0, \text{with } u_h(0) = v_h, \tag{3.5}$$

where $v_h \in X_h$ is an approximation of the initial data v, and the notation $[\cdot,\cdot]$ refers to a suitable inner product on the space X_h, approximating the usual $L^2(\Omega)$ inner product $(\cdot,\cdot)$. Following Thomée [20], we take $v_h = R_h v$ in case of smooth initial data $v \in \dot{H}^2(\Omega)$ and $v_h = P_h v$ in case of nonsmooth initial data, i.e., $v \in \dot{H}^s(\Omega)$, $-1 \le s \le 0$. Moreover, the spatially semidiscrete variational problem (3.5) can be rewritten into an operator form as

$$\partial_t^\alpha u_h(t) + A_h u_h(t) = f_h, \quad \forall \chi \in X_h, \ t > 0, \text{with } u_h(0) = v_h,$$

where A_h is a discrete approximation to the elliptic operator A on the space X_h, and will be given below.

With the help of the abstract form (3.5), we shall present three main finite element type discretization methods in space, i.e., standard Galerkin finite element (SG) method, lumped mass (LM) method and finite volume element (FVE) method. In passing, we note that in principle any other spatial discretization

methods, e.g., finite difference methods [54, 55], collocation, and spectral methods [54, 56] can also be used. Our choice of the FEMs is motivated by nonsmooth problem data, for which FEMs are most convenient for analysis.

3.1. *Standard Galerkin finite element.*

The SG method is obtained from (3.5) when the approximate inner product $[\cdot,\cdot]$ is chosen to be the usual $L^2(\Omega)$ inner product $(\cdot,\cdot)$. The SG method was first developed and rigorously analyzed for nonsmooth data in [57, 58, 34] for problem (1.1) on convex polygonal domains, and in [59] for nonconvex domains.

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 semidiscrete problem (3.5) as

$$\partial_t^\alpha u_h(t) - \Delta_h u_h(t) = f_h(t) \text{ for } t\ge 0 \quad \text{with} \quad u_h(0) = v_h. \tag{3.6}$$

Now we introduce the semidiscrete analogues of $F(t)$ and $E(t)$ for $t>0$:

$$F_h(t) := \frac{1}{2\pi\mathrm{i}}\int_{\Gamma_{\theta,\delta}} e^{zt} z^{\alpha-1}(z^\alpha-\Delta_h)^{-1}\,\mathrm{d}z \qquad \text{and} \qquad E_h(t) := \frac{1}{2\pi\mathrm{i}}\int_{\Gamma_{\theta,\delta}} e^{zt}(z^\alpha-\Delta_h)^{-1}\,\mathrm{d}z.$$

Then the solution $u_h(t)$ of the semidiscrete problem (3.6) can be expressed by:

$$u_h(t) = F_h(t)v_h + \int_0^t E_h(t-s)f_h(s)\,\mathrm{d}s. \tag{3.7}$$

Now we give pointwise-in-time $L^2(\Omega)$ error estimates for the semidiscrete Galerkin approximation u_h.

Theorem 3.1. *Let u be the solution of problem* (1.1) *and u_h be the solution of problem* (3.6)*, respectively. Then with $\ell_h = |\log h|$, for any $t>0$, the following error estimates hold:*

(i) *If $f\equiv 0$, $v\in L^2(\Omega)$, and $v_h = P_h v$, then*

$$\|(u-u_h)(t)\|_{L^2(\Omega)} \le ch^2\ell_h t^{-\alpha}\|v\|_{L^2(\Omega)}.$$

(ii) *If $f\equiv 0$, $v\in \dot H^2(\Omega)$, $v_h = R_h v$, then*

$$\|(u-u_h)(t)\|_{L^2(\Omega)} \le ch^2\|\Delta v\|_{L^2(\Omega)}.$$

(iii) *If $f\in L^\infty(0,T;L^2(\Omega))$ and $v\equiv 0$, then*

$$\|(u-u_h)(t)\|_{L^2(\Omega)} \le ch^2\ell_h^2\|f\|_{L^\infty(0,t;L^2(\Omega))}.$$

Proof. We only briefly sketch the proof for part (i) to give a flavor, and refer interested readers to [57, 34] for further details. In a customary way, we split the error $u_h(t)-u(t)$ into two terms as

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

By the approximation property of the $L^2(\Omega)$ projection P_h and Theorem 2.1, we have for any $t > 0$

$$\|\varrho(t)\|_{L^2(\Omega)} \le ch^2\|u(t)\|_{H^2(\Omega)} \le ch^2 t^{-\alpha}\|v\|_{L^2(\Omega)}.$$

So it remains to obtain proper estimates on $\vartheta(t)$. Obviously, $P_h\partial_t^\alpha \varrho = \partial_t^\alpha P_h(P_h u - u) = 0$ and using the identity $\Delta_h R_h = P_h\Delta$ [20, equation (1.34), p. 11], we deduce that ϑ satisfies:

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

Then with the help of Duhamel's formula (3.7), $\vartheta(t)$ can be represented by

$$\begin{aligned}\vartheta(t) &= -\int_0^t E_h(t-s)\Delta_h(R_h u - P_h u)(s)\,\mathrm{d}s\\ &= \int_0^t (-\Delta_h)^{1-\epsilon}E_h(t-s)(-\Delta_h)^{\epsilon}(R_h u - P_h u)(s)\,\mathrm{d}s,\end{aligned}$$

where the constant $\epsilon \in (0,1)$ is to be chosen below, and $(-\Delta_h)^\epsilon$ is the fractional power of $-\Delta_h$ defined in the spectral sense. That is, if (λ_j^h, ϕ_j^h) are the eigenvalues and eigenfunctions of $-\Delta_h$, then for $v \in X_h$, $(-\Delta_h)^\epsilon v = \sum_j (\lambda_j^h)^\epsilon (v, \phi_j^h)\phi_j^h$. Consequently,

$$\|\vartheta(t)\|_{L^2(\Omega)} \le \int_0^t \|(-\Delta_h)^{1-\epsilon}E_h(t-s)\|\|(-\Delta_h)^{\epsilon}(R_h u - P_h u)(s)\|_{L^2(\Omega)}\mathrm{d}s.$$

Now recall the following smoothing property of the semidiscrete solution operator $E_h(t)$

$$\|(-\Delta_h)^s E_h(t)\| \le ct^{-1+(1-s)\alpha} \quad \forall s \in [0,1],$$

which follows directly from the resolvent estimate (2.2) (for Δ_h), and the inverse estimate for FEM functions

$$\|(-\Delta_h)^s(R_h u - P_h u)(t)\|_{L^2(\Omega)} \le ch^{-2s}\|(R_h u - P_h u)(t)\|_{L^2(\Omega)} \quad \forall s \in [0,1].$$

Thus, by the triangle inequality and the approximation properties of R_h and P_h, cf. (3.2) and (3.3), we deduce

$$\|(R_h u - P_h u)(t)\|_{L^2(\Omega)} \le \|(R_h u - u)(t)\|_{L^2(\Omega)} + \|(P_h u - u)(t)\|_{L^2(\Omega)} \le ch^2\|u(t)\|_{H^2(\Omega)}.$$

The preceding estimates together with Theorem 2.1 imply

$$\begin{aligned}\|\vartheta(t)\|_{L^2(\Omega)} &\le ch^{2-2\epsilon}\int_0^t (t-s)^{\epsilon\alpha-1}\|u(s)\|_{H^2(\Omega)}\,\mathrm{d}s\\ &\le ch^{2-2\epsilon}\|v\|_{L^2(\Omega)}\int_0^t (t-s)^{\epsilon\alpha-1}s^{-\alpha}\,\mathrm{d}s\\ &\le c\epsilon^{-1}h^{2-2\epsilon}t^{-\alpha}\|v\|_{L^2(\Omega)}.\end{aligned}$$

The desired assertion follows by choosing $\epsilon = 1/\ell_h$. □

Remark 3.1. *The error estimates in Theorem 3.1 are stated directly with respect to the regularity of problem data, i.e., the initial condition v and the source term f, which can both be nonsmooth (e.g., in*

$L^2(\Omega)$ *only) and do not need to satisfy any compatibility condition (e.g., with zero boundary condition). The results do not directly assume any regularity on the solution* u*, although the proof relies crucially on sharp regularity estimates in Theorem 2.1. This clearly shows the central role of regularity analysis in the nonsmooth data analysis.*

It is instructive to compare the error estimate in Theorem 3.1 with that for standard parabolic problems. For example, in the latter case, for the homogeneous problem with $v \in L^2(\Omega)$*, the following error estimate holds [20, Theorem 3.5, p. 47]:*

$$\|(u - u_h)(t)\|_{L^2(\Omega)} \le ch^2 t^{-1} \|v\|_{L^2(\Omega)}.$$

This estimate is comparable with that in Theorem 3.1(i)*, apart from the log factor* ℓ_h*, which can be overcome using an operator trick due to Fujita and Suzuki [60]. Hence, in the limit* $\alpha \to 1^-$*, the result in the fractional case essentially recovers that for the standard parabolic case. The log factor* ℓ_h *in the error estimate for the inhomogeneous problem in Theorem 3.1*(*iii*) *is due to the limited smoothing property, cf. Theorem 2.1*(ii)*. It is unclear whether the factor* ℓ_h *is intrinsic or due to the limitation of the proof technique.*

Upon extension, the following error estimate analogous to Theorem 3.1(i) *holds for very weak initial data, i.e.,* $v \in H^{-s}(\Omega)$*,* $0 \le s \le 1$ *[58, Theorem 2]:*

$$\|(u - u_h)(t)\|_{L^2(\Omega)} \le ch^{2-s} \ell_h t^{-\alpha} \|v\|_{H^{-s}(\Omega)}.$$

3.2. Two variants (lumped mass and finite volume) of Galerkin method.

Now we discuss two variants of the standard Galerkin FEM, i.e., lumped mass FEM and finite volume element method. These methods have also been analyzed for nonsmooth data, but less extensively [57, 34, 61, 62]. These variants are essential for some applications: the lumped mass FEM is important for preserving qualitative properties of the approximations, e.g., positivity [63, 64], while the finite volume method inherits the local conservation property of the physical problem.

First, we describe the lumped mass FEM (see, e.g. [20, Chapter 15, pp. 239–244]), where the mass matrix is replaced by a diagonal matrix with the row sums of the original mass matrix as its diagonal elements. Specifically, let z_j^τ, $j = 1, \ldots, d+1$ be the vertices of a d-simplex $\tau \in \mathcal{T}_h$. Consider the following quadrature formula

$$Q_{\tau,h}(f) = \frac{|\tau|}{d+1} \sum_{j=1}^{d+1} f(z_j^\tau) \approx \int_\tau f \mathrm{d}x \quad \forall f \in C(K).$$

where $|K|$ denotes the area/volume of the simplex K. Then we define an approximate $L^2(\Omega)$-inner product $(\cdot, \cdot)_h$ in X_h by

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

The lumped mass FEM is to find $\bar{u}_h(t) \in X_h$ such that

$$(\partial_t^\alpha \bar{u}_h, \chi)_h + a(\bar{u}_h, \chi) = (f, \chi) \quad \forall \chi \in X_h,\ t > 0, \quad \text{with } \bar{u}_h(0) = P_h v.$$

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

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

Remark 3.2. *For a rectangular domain Ω, first partitioned into a uniform square mesh and then triangulated into a mesh by connecting the lower left corner with the upper right corner of each square, the discrete Laplacian $\bar{\Delta}_h$ is identical with the canonical five-point finite difference approximation of the Laplace operator. Such a relation may allow extending the error analysis below to various finite difference approximations of problem* (1.1).

Also, we introduce an approximate $L^2(D)$ projection operator $\bar{P}_h : L^2(\Omega) \to X_h$ by

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

Then with $f_h = \bar{P}_h f$, the lumped mass FEM can be written in an operator form as

$$\partial_t^\alpha \bar{u}_h(t) - \bar{\Delta}_h \bar{u}_h(t) = f_h(t) \quad \text{for } t \geq 0 \quad \text{with } \bar{u}_h(0) = P_h v. \tag{3.8}$$

Next, we describe the finite volume element (FVE) method (see, e.g., [65]). It is based on a discrete version of the local conservation law

$$\int_V \partial_t^\alpha u(t)\mathrm{d}x - \int_{\partial V} \frac{\partial u(t)}{\partial n}\mathrm{d}s = \int_V f(t)\,\mathrm{d}x, \quad \text{for } t > 0, \tag{3.9}$$

valid for any $V \subset \Omega$ with a piecewise smooth boundary ∂V, with n being the unit outward normal vector to the boundary ∂V. The FVE requires the identity (3.9) to be satisfied for $V = V_j,\ j = 1,\ldots,N$, which are disjoint and known as control volumes associated with the nodes P_j of the triangulation $\mathcal{T}_h$. Then the discrete problem reads: find $\widetilde{u}_h(t) \in X_h$ such that

$$\int_{V_j} \partial_t^\alpha \widetilde{u}_h(t)\,\mathrm{d}x - \int_{\partial V} \frac{\partial \widetilde{u}_h(t)}{\partial n}\mathrm{d}s = \int_{V_j} f(t)\,\mathrm{d}x, \quad \text{for } t \geq 0, \quad \text{with } \widetilde{u}_h(0) = P_h v, \tag{3.10}$$

Following [66, 67], it can be recast as a Galerkin method, by letting

$$Y_h = \left\{\varphi \in L^2(\Omega) : \varphi|_{V_j} = \text{constant}, \ \ j = 1,2,...,N;\ \varphi = 0 \ \text{ outside } \ \cup_{j=1}^N V_j\right\},$$

introducing the interpolation operator $J_h : C(\Omega) \to Y_h$ by $(J_h v)(P_j) = v(P_j)$, $j = 1,\ldots,N$, and then defining an approximate $L^2(\Omega)$ inner product $\langle\chi,\psi\rangle_h = (\chi, J_h\psi)$ for all $\chi,\psi \in X_h$, the FVE method (3.10) can be reformulated by

$$\langle\partial_t^\alpha \widetilde{u}_h(t),\chi\rangle_h + a(\widetilde{u}(t),\chi) = (f(t), J_h\chi) \quad \text{for } t \geq 0 \quad \text{with } \widetilde{u}_h(0) = P_h v \in X_h.$$

In order to be consistent with the abstract form (3.5), we slightly perturb the right hand side to $(f(t),\chi)$. Then the FVE is to find $\widetilde{u}_h \in X_h$ such that

$$\langle\partial_t^\alpha \widetilde{u}_h(t),\chi\rangle_h + a(\widetilde{u}(t),\chi) = (f(t),\chi) \quad \text{for } t \geq 0 \quad \text{with } \widetilde{u}_h(0) = P_h v. \tag{3.11}$$

Thus, it corresponds to (3.5) with $[\cdot,\cdot] = \langle\cdot,\cdot\rangle_h$. By introducing the discrete Laplacian $-\widetilde{\Delta}_h : X_h \to X_h$, corresponding to the inner product $\langle\cdot,\cdot\rangle$, defined by

$$-\langle\widetilde{\Delta}_h\psi,\chi\rangle_h = (\nabla\psi,\nabla\chi) \quad \forall\psi,\chi\in X_h,$$

and a projection operator $\widetilde{P}_h : L^2(\Omega) \to X_h$ defined by

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

In this way, the FVE method (3.11) can be written with $f_h = \widetilde{P}_h f$ in an operator form as

$$\partial_t^\alpha\widetilde{u}_h(t) - \widetilde{\Delta}_h\widetilde{u}_h(t) = f_h(t) \quad \text{for } t\ge 0 \quad \text{with } \widetilde{u}_h(0) = P_h v. \tag{3.12}$$

For the convergence analysis of the LM and FVE methods, we recall a useful quadrature error operator $Q_h : X_h \to X_h$ defined by

$$(\nabla Q_h\chi,\nabla\psi) = \epsilon_h(\chi,\psi) := [\chi,\psi] - (\chi,\psi) \quad \forall\chi,\psi\in X_h. \tag{3.13}$$

190 The operator Q_h represents the quadrature error in a special way. It satisfies the following error estimate; see [68, Lemma 2.4] for LM method and [67, Lemma 2.2] for FVE method.

Lemma 3.1. *Let A_h be $-\bar{\Delta}_h$ or $-\widetilde{\Delta}_h$, and Q_h be the operator defined by* (3.13)*. Then there holds*

$$\|\nabla Q_h\chi\|_{L^2(\Omega)} + h\|A_h Q_h\chi\|_{L^2(\Omega)} \le ch^{p+1}\|\nabla^p\chi\|_{L^2(\Omega)} \quad \forall\chi\in X_h, \quad p=0,1.$$

Furthermore, if the meshes are symmetric (for details and illustration, see [68, Section 5, Fig. 2 and 3]), then there holds

$$\|Q_h\chi\|_{L^2(\Omega)} \le ch^2\|\chi\|_{L^2(\Omega)} \quad \forall\chi\in X_h. \tag{3.14}$$

Theorem 3.2. *Let u be the solution of problem* (1.1) *and $\bar{u}_h$ be the solution of problem* (3.8) *or* (3.12)*, respectively. Then under condition* (3.14)*, the following estimates are valid for $t>0$ and $\ell_h = |\ln h|$.*

(i) *If $f\equiv 0$, $v\in L^2(\Omega)$ and $v_h = P_h v$, then*

$$\|(\bar{u}_h - u)(t)\|_{L^2(\Omega)} \le ch^2\ell_h t^{-\alpha}\|v\|_{L^2(\Omega)}. \tag{3.15}$$

(ii) *If $v\equiv 0$, $f\in L^\infty(0,T;\dot{H}^q(\Omega))$, $-1<q\le 0$, and $f_h = P_h f$, then*

$$\|(\bar{u}_h - u)(t)\|_{L^2(\Omega)} \le ch^{2+q}\ell_h^2\|f\|_{L^\infty(0,t;\dot{H}^q(\Omega))}.$$

Proof. We only sketch the proof for part (i). For the analysis, we split the error $\bar{u}_h(t) - u(t)$ 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 standard Galerkin FEM solution, cf. (3.6). Upon noting Theorem 3.1 for $\|(u_h - u)(t)\|_{L^2(\Omega)}$, it suffices to show

$$\|\delta(t)\|_{L^2(\Omega)} \le ch^2 \ell_h t^{-\alpha} \|v\|_{L^2(\Omega)}.$$

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

$$\partial_t^\alpha \delta(t) + A_h \delta(t) = -A_h Q_h \partial_t^\alpha u_h(t) \quad \text{for } t > 0, \quad \text{with } \delta(0) = 0,$$

where the operator A_h denotes either $-\bar{\Delta}_h$ or $-\widetilde{\Delta}_h$. By Duhamel's formula (3.7), $\delta(t)$ is given by

$$\begin{aligned}\delta(t) &= -\int_0^t E_h(t-s) A_h Q_h \partial_t^\alpha u_h(s) \mathrm{d}s \\ &= -\int_0^t A_h^{1-\epsilon} E_h(t-s) A_h^\epsilon Q_h \partial_t^\alpha u_h(s) \mathrm{d}s.\end{aligned}$$

Then the smoothing property of E_h, the inverse estimate and the quadrature error assumption (3.14) imply

$$\begin{aligned}\|\delta(t)\|_{L^2(\Omega)} &\le \int_0^t \|E_h(t-s) A_h^{1-\epsilon}\| \|A_h^\epsilon Q_h \partial_t^\alpha u_h(s)\|_{L^2(\Omega)} \mathrm{d}s \\ &\le ch^{-2\epsilon} \int_0^t (t-s)^{\epsilon\alpha - 1} \|Q_h \partial_t^\alpha u_h(s)\|_{L^2(\Omega)} \mathrm{d}s \\ &\le ch^{2-2\epsilon} \int_0^t (t-s)^{\epsilon\alpha - 1} \|\partial_t^\alpha u_h(s)\|_{L^2(\Omega)} \mathrm{d}s.\end{aligned}$$

Last, the (discrete) stability result $\|\partial_t^\alpha u_h(t)\|_{L^2(\Omega)} \le ct^{-\alpha} \|v_h\|_{L^2(\Omega)}$ (which follows analogously as Theorem 2.1(i)) and the $L^2(\Omega)$-stability of P_h imply

$$\begin{aligned}\|\delta(t)\|_{L^2(\Omega)} &\le ch^{2-2\epsilon} \int_0^t (t-s)^{\epsilon\alpha - 1} s^{-\alpha} \|u_h(0)\|_{L^2(\Omega)} \mathrm{d}s \\ &\le c\epsilon^{-1} h^{2-2\epsilon} t^{-\alpha} \|v_h\|_{L^2(\Omega)} \le c\epsilon^{-1} h^{2-2\epsilon} t^{-\alpha} \|v\|_{L^2(\Omega)}.\end{aligned}$$

Then the desired assertion follows immediately by choosing $\epsilon = 1/\ell_h$. □

Remark 3.3. *The quadrature error condition* (3.14) *is satisfied for symmetric meshes [68, Section 5]. If condition* (3.14) *does not hold, we are able to show only a suboptimal $O(h)$-convergence rate for $L^2(\Omega)$-norm of the error [57, Theorem 4.5], which is reminiscent of that in the classical parabolic case, e.g. [68, Theorem 4.4].*

Generally, the FEM analysis in the fractional case is much more delicate than the standard parabolic case due to the less standard solution operators and the limited smoothing property. Nonetheless, the results in the two cases are largely comparable, and the overall proof strategy is often similar. The Laplace approach described above represents only one way to analyze the spatially semidiscrete schemes. Recently, Karaa [69] gave a unified analysis of all three methods for the homogeneous problem based on an energy argument, which generalizes the corresponding technique for standard parabolic problems in [20, Chapter 3]. However, the analysis of the inhomogeneous case by the energy argument is still missing.

The energy type argument is generally more tricky in the fractional case. This is due to the nonlocality of the fractional derivative $\partial_t^\alpha u$ and consequently that many powerful PDE tools, like integration by parts formula and product rule, are either invalid or require substantial modification. See also [70, 71] for some results on a related subdiffusion model.

3.3. Illustrations and outstanding issues on semidiscrete methods

Now we illustrate the three spatially semidiscrete methods with very weak initial data.

Example 3.1. *Consider problem* (1.1) *on the unit square* $\Omega = (0,1)^2$ *with* $f = 0$ *and very weak initial data* $v = \delta_\Gamma$*, with* Γ *being the line segment* $\{\frac{1}{2}\} \times [\frac{1}{4}, \frac{3}{4}]$ *and* $\langle \delta_\Gamma, \phi\rangle = \int_\Gamma \phi(s)\mathrm{d}s$*. One may view* (v, χ) *for* $\chi \in X_h \subset H^{-\frac{1}{2}-\epsilon}(\Omega)$ *as duality pairing between the spaces* $H^{-\frac{1}{2}-\epsilon}(\Omega)$ *and* $\dot{H}^{\frac{1}{2}+\epsilon}(\Omega)$ *for any* $\epsilon > 0$ *so that* $\delta_\Gamma \in H^{-\frac{1}{2}-\epsilon}(\Omega)$*. Indeed, it follows from Hölder's inequality and trace theorem [27] that*

$$\|\delta_\Gamma\|_{H^{-\frac{1}{2}-\epsilon}(\Omega)} = \sup_{0\neq\phi\in\dot{H}^{\frac{1}{2}+\epsilon}(\Omega)} \frac{|\int_\Gamma \phi(s)\mathrm{d}s|}{\|\phi\|_{H^{\frac{1}{2}+\epsilon}(\Omega)}} \le |\Gamma|^{\frac{1}{2}} \sup_{0\neq\phi\in\dot{H}^{\frac{1}{2}+\epsilon}(\Omega)} \frac{\|\phi\|_{L^2(\Gamma)}}{\|\phi\|_{H^{\frac{1}{2}+\epsilon}(\Omega)}}.$$

Table 1: The errors $\|(u-u_h)(t)\|_{L^2(\Omega)}$ of the FEM approximation $u_h(t)$ (by three semidiscrete schemes) for Example 3.1 with $\alpha = 0.5$, at $t = 0.001, 0.01, 0.1$, discretized on a uniform mesh, $h = 1/M$.

(a) Galerkin FEM

M	10	20	40	80	160	rate
$t = 0.001$	9.01e-3	3.14e-3	8.85e-4	2.44e-4	6.63e-5	≈ 1.87 (1.50)
$t = 0.01$	2.45e-3	1.14e-3	3.13e-4	8.47e-5	2.27e-5	≈ 1.90 (1.50)
$t = 0.1$	7.43e-4	3.98e-4	1.08e-4	2.88e-5	7.66e-6	≈ 1.91 (1.50)

(b) lumped mass method

M	10	20	40	80	160	rate
$t = 0.001$	6.77e-2	2.29e-2	8.38e-3	3.01e-3	1.07e-3	≈ 1.48 (1.50)
$t = 0.01$	2.21e-2	7.29e-3	2.66e-3	9.54e-4	3.40e-4	≈ 1.48 (1.50)
$t = 0.1$	7.07e-3	2.31e-3	8.41e-4	3.02e-4	1.08e-4	≈ 1.48 (1.50)

(c) finite volume element method

M	10	20	40	80	160	rate
$t = 0.001$	1.89e-2	5.56e-3	1.95e-3	6.87e-4	2.42e-4	≈ 1.50 (1.50)
$t = 0.01$	5.81e-3	1.76e-3	6.17e-4	2.17e-4	7.66e-5	≈ 1.50 (1.50)
$t = 0.1$	1.80e-3	5.63e-4	1.97e-4	6.90e-5	2.43e-5	≈ 1.51 (1.50)

The reference solution used for computing the $L^2(D)$ error is computed by a fully discrete scheme based on the standard Galerkin method in space with $h = 1/1280$ and BDF3 convolution quadrature in time

with $\tau = T/1000$, cf. (4.4) below. The numerical results obtained by standard Galerkin method, lumped mass method and finite volume element method are given in Table 1 (a), (b) and (c), respectively. In the tables, the numbers in the bracket in the last column refer to the theoretical rate from Remark 3.1. The empirical convergence rate for the very weak data δ_Γ agrees well with the theoretical rate; Interestingly, for the standard Galerkin scheme, the $L^2(\Omega)$-norm of the error exhibits super-convergence. This is attributed to the fact that the singularity of the solution is supported on the interface Γ and it is fully aligned with the mesh. It is observed that for all three semidiscrete schemes, the error increases as the time $t \to 0^+$, which concurs with the weak solution singularity at the initial time $t = 0$.

We end this section with some future research problems. Despite the maturity of the FEM analysis, there are still a few interesting questions on the FEMs for the model (1.1) which are not well understood and await further research:

(i) So far the analysis is mostly concerned with a time-independent coefficient, which can be treated conveniently using the semigroup type techniques. The time dependent case requires different techniques, and the nonlocality of the operator $\partial_t^\alpha u$ prevents a straightforward adaptation of known techniques for standard parabolic problems [72]. Encouraging results in this direction were established in a recent work of Mustapha [73], where optimal error estimates for the homogeneous problem were obtained using an energy argument.

(ii) All existing works focus on linear finite elements, and there seems to be no study on high-order finite elements for nonsmooth problem data. It is unclear whether there are similar nonsmooth data estimates, as in the parabolic case [20, Chapter 3] (see, e.g., [74, p. 397] for smooth data). This problem is interesting in view of the limited smoothing property of the solution operators in space in Theorem 2.1, which has played a major role in the error analysis. Thus it is of substantial interest to develop and to analyze high-order schemes in space.

(iii) The study on nonlinear subdiffusion models is rather limited, and there seems to be no error estimate with respect to the data regularity, especially for nonsmooth problem data. The recent progress [36, 32] in discrete maximal ℓ^p regularity results may provide useful tools for this purpose, which have proven extremely powerful for the study of nonlinear parabolic problems [37]. One outstanding issue seems to be sharp regularity estimates for general problem data.

4. Fully discrete schemes by time-stepping

One outstanding challenge for solving the subdiffusion model lies in the accurate and efficient discretization of the fractional derivative $\partial_t^\alpha u$. Roughly speaking, there are two predominant groups of numerical methods for time stepping, i.e., convolution quadrature and finite difference type methods, e.g., L1 scheme and L1-2 scheme. The former relies on approximating the (Riemann-Liouville) fractional derivative in the

Laplace domain (i.e., symbol), whereas the latter approximates the Caputo derivative directly by piecewise polynomials. These two approaches have their pros and cons: convolution quadrature (CQ) is quite flexible and often much easier to analyze, since by construction, it inherits excellent numerical stability property of the underlying schemes for ODEs, but it is often restricted to uniform grids. The finite difference type methods are very flexible in construction and implementation and can easily generalize to nonuniform grids, but often challenging to analyze, especially for nonsmooth problem data. Generally, these schemes are only first-order accurate when implemented straightforwardly, unless restrictive compatibility conditions are fulfilled. Hence, suitable corrections to the straightforward implementation are needed in order to restore the desired high-order convergence.

In this section, we review these two popular classes of time-stepping schemes on uniform grids. Specifically, let $\{t_n = n\tau\}_{n=0}^N$ be a uniform partition of the time interval $[0, T]$, with a time step size $\tau = T/N$. The case of general nonuniform time grids is also of interest, e.g., in resolving initial or interior layer, but the analysis seems not well understood at present; we refer interested readers to the references [62, 75, 76, 77] for some recent progress on nonuniform grids.

4.1. Convolution quadrature

Convolution quadrature (CQ) was first proposed by Lubich in a series of works [78, 79, 80] for discretizing Volterra integral equations. It has been widely applied in discretizing the Riemann-Liouville fractional derivative (see, e.g., [81, 82, 83]). One distinct feature is that the construction requires only that Laplace transform of the kernel be known. Specifically, CQ approximates the Riemann-Liouville derivative $\partial_t^\alpha \varphi(t_n)$, which is defined by

$$ {}^R\partial_t^\alpha \varphi := \frac{\mathrm{d}}{\mathrm{d}t} \frac{1}{\Gamma(1-\alpha)} \int_0^t (t-s)^{-\alpha} \varphi(s) \mathrm{d}s,$$

(with $\varphi(0) = 0$) by a discrete convolution (with the shorthand notation $\varphi^n = \varphi(t_n)$)

$$\bar\partial_\tau^\alpha \varphi^n := \frac{1}{\tau^\alpha} \sum_{j=0}^n b_j \varphi^{n-j}. \tag{4.1}$$

The weights $\{b_j\}_{j=0}^\infty$ are the coefficients in the power series expansion

$$\delta_\tau(\zeta)^\alpha = \frac{1}{\tau^\alpha} \sum_{j=0}^\infty b_j \zeta^j, \tag{4.2}$$

where $\delta_\tau(\zeta) = \delta(\zeta)/\tau$ is the characteristic polynomial of a linear multistep method for ODEs, with $\delta(\zeta) = \delta_1(\zeta)$. There are several possible choices of the characteristic polynomial, e.g., backward differentiation formula, trapezoidal rule, Newton-Gregory method and Runge-Kutta methods. The most popular one is the backward differentiation formula of order k (BDFk), $k = 1, \ldots, 6$, for which $\delta(\zeta)$ is given by

$$\delta_\tau(\zeta) := \frac{1}{\tau} \sum_{j=1}^k \frac{1}{j} (1-\zeta)^j, \quad j = 1, 2, \ldots.$$

The special case $k = 1$, i.e., the backward Euler convolution quadrature, is commonly known as Grünwald-Letnikov approximation in the literature [84] and the coefficients b_j are given explicitly by the following recurrence relation

$$b_0 = 1, \quad b_j = -\frac{\alpha - j + 1}{j} b_{j-1}.$$

Generally, the weights b_j can be evaluated efficiently via recursion or discrete Fourier transform [85, 86].

The CQ discretization first reformulates problem (1.1) by the Riemann-Liouville derivative ${}^R\partial_t^\alpha \varphi$, using the defining relation for the Caputo derivatives [11, p. 91] $\partial_t^\alpha \varphi(t) = {}^R\partial_t^\alpha(\varphi - \varphi(0))$, into the form

$${}^R\partial_t^\alpha(u - v) - \Delta u = f.$$

Then the time stepping scheme based on the CQ for problem (1.1) is to seek approximations U^n to the exact solution $u(t_n)$ by

$$\bar\partial_\tau^\alpha (U - v)^n - \Delta U^n = f(t_n), \quad n = 1, \dots, N, \tag{4.3}$$

with $U^0 = v$. It can be combined with spatially semidiscrete schemes described in Section 3 to arrive at fully discrete schemes, which are implementable on computers. Our discussions below focus on the temporal error for time-stepping schemes, and omit the spatial error.

If the exact solution u is smooth and has sufficiently many vanishing derivatives at $t = 0$, then the approximation U^n converges at a rate of $O(\tau^k)$ uniformly in time t [79, Theorem 3.1]. However, it generally only exhibits a first-order accuracy when solving fractional evolution equations even for smooth v and f [87, 83], since the requisite compatibility condition is usually not satisfied. This loss of accuracy is one distinct feature for most time stepping schemes, since they are usually derived under the assumption that the solution u is sufficiently smooth, which holds only if the problem data satisfy certain rather restrictive compatibility conditions. In summary, they tend to lack robustness with respect to the regularity of problem data.

This observation on accuracy loss has motivated some research works. For fractional ODEs, one successful idea is to use starting weights [78] to correct the CQ in discretizing $\partial_t^\alpha \varphi(t_n)$ by

$$\bar\partial_\tau^\alpha \varphi^n = \tau^{-\alpha} \sum_{j=0}^{n} b_{n-j} \varphi^j + \sum_{j=0}^{M} w_{n,j} \varphi^j,$$

where $M \in \mathbb{N}$ and the weights $w_{n,j}$ depend on the fractional order α and the convergence order k. The purpose of the starting term $\sum_{j=0}^{M} w_{n,j}\varphi^j$ is to capture all leading singularities so as to recover a uniform $O(\tau^k)$ rate. The weights $w_{n,j}$ have to be computed at every time step, which involves solving a linear system with Vandermonde type matrices and may lead to instability issue (if a large M is needed, which is likely the case when α is close to zero). This idea works well for fractional ODEs (see, e.g., [88]); however, its extension to fractional PDEs essentially seems to boil down to expanding the solution into (fractional-order) power series in t, which would impose certain strong compatibility conditions on the source f.

The more promising idea for the model (1.1) is initial correction. It corrects only the first few steps of the schemes. This idea can at least be traced back to the work [89], where a correction of the BDF2 CQ was developed for an integro-differential equation with a positive memory term. Then it was applied as an abstract framework in [87] for BDF2 CQ in order to achieve a uniform second-order convergence for semilinear fractional diffusion-wave equations (which is slightly different from the model (1.1)) with smooth data. Further, BDF2 CQ was extended to subdiffusion and diffusion wave equations in [83] and very recently also general BDFk [90]. In particular, in the work [90], by a careful analysis of the error representation in the Laplace domain, a set of simple algebraic criteria was derived for constructing correction schemes. In the following, we describe the correction scheme for the BDF CQ derived in [90].

To restore the k^{th}-order accuracy for BDFk CQ, the scheme in [90] corrects the starting $k-1$ steps by (as usual, the summation disappears if the upper index is smaller than the lower one)

$$\begin{cases} \bar\partial_\tau^\alpha (U-v)^n - \Delta U^n = a_n^{(k)}(\Delta v + f(0)) + f(t_n) + \sum_{\ell=1}^{k-2} b_{\ell,n}^{(k)} \tau^\ell \partial_t^{(\ell)} f(0), & 1 \le n \le k-1, \\ \bar\partial_\tau^\alpha (U-v)^n - \Delta U^n = f(t_n), & k \le n \le N. \end{cases} \tag{4.4}$$

where the coefficients $a_n^{(k)}$ and $b_{\ell,n}^{(k)}$ are given in Table 2. When compared with the vanilla CQ scheme (4.3), the additional terms are constructed so as to improve the overall accuracy of the scheme to $O(\tau^k)$ for an initial data $v \in D(\Delta)$ and a possibly incompatible right-hand side f [90]. The only difference between the corrected scheme (4.4) and the standard scheme (4.3) lies in the correction terms at the starting $k-1$ steps for BDFk. Hence, the scheme (4.4) is easy to implement. The correction is also minimal in the sense that there is no other correction scheme which uses fewer correction steps while attaining the same accuracy. The corrected scheme (4.4) satisfies the following error estimates [90, Theorem 2.4].

Theorem 4.1. *Let $f \in C^{k-1}([0,T];L^2(\Omega))$ and $\int_0^t (t-s)^{\alpha-1}\|\partial_s^{(k)} f(s)\|_{L^2(\Omega)}\mathrm{d}s < \infty$. Then for the solution U^n to (4.4), the following error estimates hold for any $t_n > 0$.*

(i) *If $\Delta v \in L^2(\Omega)$, then*

$$\|U^n - u(t_n)\|_{L^2(\Omega)} \le c\tau^k \bigg(t_n^{\alpha-k}\|f(0)+\Delta v\|_{L^2(\Omega)} + \sum_{\ell=1}^{k-1} t_n^{\alpha+\ell-k}\|\partial_t^{(\ell)} f(0)\|_{L^2(\Omega)} + \int_0^{t_n} (t_n-s)^{\alpha-1}\|\partial_s^{(k)} f(s)\|_{L^2(\Omega)}\mathrm{d}s \bigg).$$

(ii) *If $v \in L^2(\Omega)$, then*

$$\|U^n - u(t_n)\|_{L^2(\Omega)} \le c\tau^k \bigg(t_n^{-k}\|v\|_{L^2(\Omega)} + \sum_{\ell=0}^{k-1} t_n^{\alpha+\ell-k}\|\partial_t^{(\ell)} f(0)\|_{L^2(\Omega)} + \int_0^{t_n} (t_n-s)^{\alpha-1}\|\partial_s^{(k)} f(s)\|_{L^2(\Omega)}\mathrm{d}s \bigg).$$

Remark 4.1. *Note that similar to Theorem 3.1, the error estimates in Theorem 4.1 depend only on the regularity of the source term f and the initial condition v, which can be incompatible or nonsmooth,*

rather than the regularity of the solution u. For nonsmooth initial data v, e.g., $v \in L^2(\Omega)$, the term Δv is no longer in the space $L^2(\Omega)$, and so the scheme (4.4) (*and the scheme* (4.8) *below*) *should be interpreted in a distributional sense or the scheme is stated for the spatially discrete operator. Further, although no restrictive condition on the spatial regularity of the source term f, in order to achieve high-order temporal convergence of any scheme, it necessarily possess suitable temporal regularity, as indicated by $f \in C^{k-1}([0,T]; L^2(\Omega))$ and $\int_0^t (t-s)^{\alpha-1} \|\partial_s^{(k)} f(s)\|_{L^2(\Omega)} \mathrm{d}s < \infty$. Should these conditions fail, the temporal convergence rate deteriorates accordingly, due to a lack of necessary solution regularity, cf. Theorem 2.1. Theorem 4.1 implies that for any fixed $t_n > 0$, the rate is $O(\tau^k)$ for BDFk CQ. In order to have a uniform $O(\tau^k)$ rate, the following compatibility conditions are needed:*

$$f(0) + \Delta v = 0 \quad \textit{and} \quad \partial_t^{(\ell)} f(0) = 0, \quad \ell = 1, \ldots, k-1.$$

Otherwise, the estimate deteriorates as $t \to 0$, in accordance with the regularity theory in Theorem 2.1: the solution (*and its derivatives*) *exhibits weak singularity at $t = 0$.*

Remark 4.2. *The case $k = 1$ corresponds to the backward Euler CQ, and it does not require any correction in order to achieve a first-order convergence.*

In passing, we note that not all CQ schemes require initial correction in order to recover high-order convergence. One notable example is Runge-Kutta CQ; see [91] for semilinear parabolic problems and [92] for the subdiffusion model as well as its fast implementation. Further, a proper weighted average of shifted standard Grunwald-Letnikov approximations can also lead to high-order approximations [93]. CQ schemes can exhibit superconvergence at points that may be different from the grid points, which can also be effectively exploited to develop high-order schemes. For example, the Grunwald-Letnikov formula is only first-order accurate at the grid point $t = t_n$, but it is actually second-order accurate at the intermediate point $t = t_n - \frac{\alpha}{2}\tau$ [94], which results in a scheme analogous to the classical Crank-Nicolson scheme for standard parabolic equations. However, the corrected versions of these approximations have not yet been developed for the general case, except for a fractional variant of Crank-Nicolson scheme [95].

4.2. Piecewise polynomial interpolation

Now we describe time stepping schemes based on piecewise polynomial approximation, especially interpolation. These schemes are essentially of finite difference nature, and the most prominent one is the L1

Table 2: The coefficients $a_j^{(k)}$ and $b_{\ell,j}^{(k)}$ [90, Tables 1 and 2].

BDFk	$a_1^{(k)}$	$a_2^{(k)}$	$a_3^{(k)}$	$a_4^{(k)}$	$a_5^{(k)}$
$k=2$	$\frac{1}{2}$				
$k=3$	$\frac{11}{12}$	$-\frac{5}{12}$			
$k=4$	$\frac{31}{24}$	$-\frac{7}{6}$	$\frac{3}{8}$		
$k=5$	$\frac{1181}{720}$	$-\frac{177}{80}$	$\frac{341}{240}$	$-\frac{251}{720}$	
$k=6$	$\frac{2837}{1440}$	$-\frac{2543}{720}$	$\frac{17}{5}$	$-\frac{1201}{720}$	$\frac{95}{288}$

BDFk		$b_{\ell,1}^{(k)}$	$b_{\ell,2}^{(k)}$	$b_{\ell,3}^{(k)}$	$b_{\ell,4}^{(k)}$	$b_{\ell,5}^{(k)}$
$k=3$	$\ell=1$	$\frac{1}{12}$	0			
$k=4$	$\ell=1$	$\frac{1}{6}$	$-\frac{1}{12}$	0		
	$\ell=2$	0	0	0		
$k=5$	$\ell=1$	$\frac{59}{240}$	$-\frac{29}{120}$	$\frac{19}{240}$	0	
	$\ell=2$	$\frac{1}{240}$	$-\frac{1}{240}$	0	0	
	$\ell=3$	$\frac{1}{720}$	0	0	0	
$k=6$	$\ell=1$	$\frac{77}{240}$	$-\frac{7}{15}$	$\frac{73}{240}$	$-\frac{3}{40}$	0
	$\ell=2$	$\frac{1}{96}$	$-\frac{1}{60}$	$\frac{1}{160}$	0	0
	$\ell=3$	$-\frac{1}{360}$	$\frac{1}{720}$	0	0	0
	$\ell=4$	0	0	0	0	0

scheme. The L1 approximation of the Caputo derivative $\partial_t^\alpha u(t_n)$ is given by [54, Section 3]

$$\begin{aligned}
\partial_t^\alpha u(t_n) &= \frac{1}{\Gamma(1-\alpha)} \sum_{j=0}^{n-1} \int_{t_j}^{t_{j+1}} \frac{\partial u(s)}{\partial s} (t_n - s)^{-\alpha} \, \mathrm{d}s \\
&\approx \frac{1}{\Gamma(1-\alpha)} \sum_{j=0}^{n-1} \frac{u(t_{j+1}) - u(t_j)}{\tau} \int_{t_j}^{t_{j+1}} (t_n - s)^{-\alpha} \mathrm{d}s \\
&= \sum_{j=0}^{n-1} b_j \frac{u(t_{n-j}) - u(t_{n-j-1})}{\tau^\alpha} \\
&= \tau^{-\alpha} [b_0 u(t_n) - b_{n-1} u(t_0) + \sum_{j=1}^{n-1} (b_j - b_{j-1}) u(t_{n-j})] =: L_1^n(u),
\end{aligned} \tag{4.5}$$

where the weights b_j are given by

$$b_j = ((j+1)^{1-\alpha} - j^{1-\alpha})/\Gamma(2-\alpha), \ j = 0, 1, \ldots, N-1.$$

In essence, it approximates the function u by a continuous piecewise linear interpolation, in a manner similar to the backward Euler method. Thus it can be viewed as a fractional analogue of the latter. It was

shown in [54, equation (3.3)] and [96, Lemma 4.1] that the local truncation error of the L1 approximation is bounded by

$$|\partial_t^\alpha u(t_n) - L_1^n(u)| \le c(u)\tau^{2-\alpha}, \tag{4.6}$$

where the constant $c(u)$ depends on $\|u\|_{C^2([0,T])}$. Thus, it requires that the solution u be twice continuously differentiable in time. Since its first appearance, the L1 scheme has been widely used in practice, and currently it is one of the most popular and successful numerical methods for solving the model (1.1). With the L1 scheme in time, we arrive at the following time stepping scheme: Given $U^0 = v$, find $U^n \in \dot{H}^1(\Omega)$ for $n = 1, 2, \ldots, N$

$$L_1^n(U) - \Delta U^n = f(t_n). \tag{4.7}$$

We have the following temporal error estimate for the scheme (4.7) [97, 38]. This is achieved by means of discrete Laplace transform, and it is rather technical, since the discrete Laplace transform of the weights b_j involves the fairly wieldy polylogarithmic function. See also [98] for a different analysis via an energy argument. Formally, the error estimate is nearly identical with that for the backward Euler CQ. Thus, in stark contrast to the $O(\tau^{2-\alpha})$ rate expected from the local truncation error (4.6) for smooth solutions, the L1 scheme is generally only first-order accurate, even for smooth initial data or source term.

Theorem 4.2. *Let u and U^n be the solutions of problems* (1.1) *and* (4.7)*, respectively. Then for any $\beta \in [0, 2]$, there holds*

$$\|u(t_n) - U^n\|_{L^2(\Omega)} \le c\tau t_n^{\beta\alpha/2-1}\|u_0\|_{\dot{H}^\beta(\Omega)} + c\tau\bigg(t_n^{\alpha-1}\|f(0)\|_{L^2(\Omega)} + \int_0^{t_n}(t_n - s)^{\alpha-1}\|f'(s)\|_{L^2(\Omega)}\mathrm{d}s\bigg).$$

Very recently, a corrected L1 scheme was developed by Yan et al [99] (see also [100, 101] for related works from the group). The corrected scheme is given by

$$\begin{cases} L_1^1(U) - \Delta U^1 - \frac{1}{2}\Delta U^0 = f(t_1) + \frac{1}{2}f(0), \quad n = 1 \\ \qquad\qquad L_1^n(U) - \Delta U^n = f(t_n), \quad n \ge 2. \end{cases} \tag{4.8}$$

It is noteworthy that it requires only correcting the first step, and incidentally, the correction term is identical with that for BDF2 CQ, one special case of the scheme (4.4) [83]. The following error estimate holds for the corrected scheme. Note that the stated regularity requirement on the source term f may be not optimal for $\alpha > 1/2$.

Theorem 4.3. *Let u and U^n be the solutions of problems* (1.1) *and* (4.8)*, respectively. Then for any $\beta \in [0, 2]$, there holds*

$$\begin{aligned}\|u(t_n) - U^n\|_{L^2(\Omega)} \le c\tau^{2-\alpha}\bigg(&t_n^{(\beta/2+1)\alpha-2}\|v\|_{\dot{H}^\beta(\Omega)} + t_n^{2\alpha-2}\|f(0)\|_{L^2(\Omega)} + t_n^{2\alpha-1}\|f'(0)\|_{L^2(\Omega)} \\ &+ \int_0^{t_n}(t_n - s)^{2\alpha-1}\|f''(s)\|_{L^2(\Omega)}\mathrm{d}s\bigg).\end{aligned}$$

There have been several important efforts in extending the L1 scheme to high-order schemes by using high-order local polynomials [102, 103, 104] and superconvergent points [105]. For example, the L1-2 scheme due to Gao et al [103] applies a piecewise linear approximation on the first subinterval, and a quadratic approximation on the other subintervals to improve the numerical accuracy. However, the performance of these time stepping schemes for nonsmooth data is not even fully examined numerically, not to mention rigorous theoretical analysis.

Besides, Mustapha and McLean developed several discontinuous Galerkin methods [106, 107, 108] for a variant of the model (1.1):

$$\partial_t u - {}^R\partial_t^{1-\alpha}\Delta u = f,$$

together with suitable boundary and initial conditions. Formally, this model can be derived from (1.1) by applying the Riemann-Liouville operator ${}^R\partial_t^{1-\alpha}$ to both sides of the equation and then redefining the source term $\tilde{f} = {}^R\partial_t^{1-\alpha} f$. The resulting schemes are similar to piecewise polynomial interpolation described above. However, the nonsmooth error estimates are mostly unavailable, except for the piecewise constant discontinuous Galerkin method (for the homogeneous problem) [109, 110]; see also [111] for a Crank-Nicolson type scheme for a related model.

4.3. Illustrations and outstanding issues

Now we illustrate the performance of the corrected time stepping schemes, i.e., (4.4) and (4.8), with both smooth and nonsmooth initial data. In order to observe the convergence rate for high-order BDF CQs, we employ Multiprecision Computing Toolbox for MATLAB (`https://www.advanpix.com`, last accessed on September 10, 2018) in the computation.

Example 4.1. *Consider problem* (1.1) *on* $\Omega = (0,1)$ *with* $v = x\sin(2\pi x) \in \dot{H}^2(\Omega)$ *and* $f = 0$.

In this example, the initial data v is smooth. The numerical results, i.e., the $L^2(\Omega)$ error $\|u_h(t_N) - u_h^N\|_{L^2(\Omega)}$ at $t_N = 1$, by the vanilla BDFk CQ schemes are presented in Table 3 (a), where the reference solution $u_h(t_N)$ is computed using the corrected BDF6 CQ scheme with a much finer temporal mesh with $\tau = T/3200$. The numerical results show only a first-order empirical convergence rate, for all vanilla BDFk CQ, $k \geq 2$, which shows clearly the lack of robustness of the naive CQ scheme (4.3) with respect to problem data regularity, despite the good regularity of the initial data v (actually it belongs to the space $\dot{H}^2(\Omega)$). In sharp contrast, the corrected scheme (4.4) can achieve the desired convergence rate; see Table 3(b). These observations remain valid for the L1 scheme and its corrected version; see Tables 4. These results agree well with the theoretical predictions from Theorems 4.1 and 4.3. It is worth noting that the desired rate for the corrected L1 scheme only kicks in at a relatively small time step size, and its precise mechanism remains unclear. These results show clearly the effectiveness of the idea of initial correction for restoring the desired high-order convergence.

Table 3: The L^2-norm error for Example 4.1 at $t_N = 1$, by the BDFk CQ with $h = 1/100$.

(a) uncorrected BDFk

α	N	50	100	200	400	800	rate
0.5	BDF2	4.94e-3	2.48e-3	1.24e-3	6.20e-4	3.10e-4	$\approx$ 1.00 (1.00)
	BDF3	4.99e-3	2.49e-3	1.24e-3	6.21e-4	3.11e-4	$\approx$ 1.00 (1.00)
	BDF4	4.99e-3	2.49e-3	1.24e-3	6.21e-4	3.11e-4	$\approx$ 1.00 (1.00)
	BDF5	4.99e-3	2.49e-3	1.24e-3	6.21e-4	3.11e-4	$\approx$ 1.00 (1.00)
	BDF6	4.96e-3	2.49e-3	1.24e-3	6.21e-4	3.11e-4	$\approx$ 1.00 (1.00)

(b) Corrected BDFk

α	$k\backslash N$	50	100	200	400	800	rate
0.25	2	5.87e-5	1.45e-5	3.59e-6	8.95e-7	2.23e-7	$\approx$ 2.00 (2.00)
	3	2.39e-6	2.88e-7	3.53e-8	4.38e-9	5.45e-10	$\approx$ 3.00 (3.00)
	4	1.49e-7	8.72e-9	5.27e-10	3.24e-11	2.01e-12	$\approx$ 4.02 (4.00)
	5	1.33e-8	3.57e-10	1.06e-11	3.22e-13	9.91e-15	$\approx$ 5.02 (5.00)
	6	1.12e-5	1.54e-9	2.68e-13	4.02e-15	6.16e-17	$\approx$ 6.04 (6.00)
0.5	2	1.77e-4	4.34e-5	1.08e-5	2.68e-6	6.69e-7	$\approx$ 2.00 (2.00)
	3	7.85e-6	9.44e-7	1.16e-7	1.43e-8	1.78e-9	$\approx$ 3.01 (3.00)
	4	5.23e-7	3.04e-8	1.83e-9	1.12e-10	6.97e-12	$\approx$ 4.02 (4.00)
	5	4.86e-8	1.30e-9	3.85e-11	1.17e-12	3.60e-14	$\approx$ 5.03 (5.00)
	6	2.82e-5	2.99e-9	1.01e-12	1.51e-14	2.32e-16	$\approx$ 6.05 (6.00)
0.75	2	4.58e-4	1.12e-4	2.78e-5	6.92e-6	1.73e-6	$\approx$ 2.00 (2.00)
	3	2.39e-5	2.85e-6	3.49e-7	4.31e-8	5.36e-9	$\approx$ 3.01 (3.00)
	4	1.80e-6	1.04e-7	6.22e-9	3.81e-10	2.36e-11	$\approx$ 4.02 (4.00)
	5	2.51e-7	4.90e-9	1.44e-10	4.35e-12	1.34e-13	$\approx$ 5.03 (5.00)
	6	1.65e-3	4.20e-7	4.17e-12	6.10e-14	9.31e-16	$\approx$ 6.06 (6.00)

Example 4.2. *Consider problem* (1.1) *on* $\Omega = (0, 1)$ *with* $v = \chi_{(0,1/2)}(x) \in \dot{H}^{\frac{1}{2}-\epsilon}(\Omega)$ (*with* χ_S *being the characteristic function of the set* S) *and* $f = 0$.

In this example, the initial data v is nonsmooth. It is observed that the uncorrected schemes, e.g., BDFk CQ and L1 scheme, can only achieve a first-order convergence (at any fixed t_N); see Table 6 for an illustration. In contrast, both corrected BDFk CQ in (4.4) and corrected L1 scheme (4.8), can achieve the desired convergence rate, cf Tables 5 and 6, agreeing excellently with the theoretical predictions in Theorem 4.1 and 4.3. These results clearly illustrate the potential of initial correction for restoring high-order convergence and robustness of the schemes.

Table 4: The L^2-norm error for Example 4.1 at $t_N = 0.01$, by the L1 scheme (4.4) with $h = 1/100$ and $N = 1000 \times 2^k$.

scheme	$\alpha\backslash k$	1	2	3	4	5	6	rate
L1	0.3	1.28e-4	6.38e-5	3.19e-5	1.59e-5	7.97e-6	3.99e-6	≈ 1.00 (1.00)
	0.5	1.99e-4	9.94e-5	4.96e-5	2.47e-5	1.24e-5	6.17e-6	≈ 1.00 (1.00)
	0.7	2.86e-3	1.41e-4	6.98e-5	3.46e-5	1.72e-5	8.34e-6	≈ 1.01 (1.00)
corrected L1	0.3	7.94e-8	2.79e-8	9.67e-9	3.28e-9	1.09e-9	3.56e-10	≈ 1.63 (1.70)
	0.5	1.90e-6	6.93e-7	2.50e-7	8.95e-8	3.19e-8	1.14e-8	≈ 1.49 (1.50)
	0.7	1.97e-5	8.06e-6	3.29e-6	1.34e-6	5.44e-7	2.21e-7	≈ 1.30 (1.30)

Table 5: The L^2-norm error for Example 4.2 at $t_N = 1$, by the corrected scheme (4.4) with $h = 1/100$.

α	$k\backslash N$	50	100	200	400	800	rate
0.25	2	5.70e-5	1.41e-5	3.49e-6	8.70e-7	2.17e-7	≈ 2.00 (2.00)
	3	2.31e-6	2.78e-7	3.42e-8	4.24e-9	5.27e-10	≈ 3.01 (3.00)
	4	1.44e-7	8.40e-9	5.08e-10	3.12e-11	1.93e-12	≈ 4.02 (4.00)
	5	1.28e-8	3.44e-10	1.02e-11	3.09e-13	9.57e-15	≈ 5.02 (5.00)
	6	1.10e-5	1.59e-9	2.57e-13	3.84e-15	7.51e-17	≈ 5.87 (6.00)
0.5	2	1.75e-4	4.30e-5	1.07e-5	2.66e-6	6.63e-7	≈ 2.00 (2.00)
	3	7.75e-6	9.32e-7	1.14e-7	1.41e-8	1.76e-9	≈ 3.01 (3.00)
	4	5.14e-7	2.99e-8	1.80e-9	1.11e-10	6.86e-12	≈ 4.02 (4.00)
	5	4.77e-8	1.28e-9	3.78e-11	1.15e-12	3.53e-14	≈ 5.03 (5.00)
	6	2.98e-5	2.83e-9	9.90e-13	1.48e-14	2.04e-16	≈ 6.12 (6.00)
0.75	2	4.79e-4	1.17e-4	2.91e-5	7.23e-6	1.80e-6	≈ 2.00 (2.00)
	3	2.52e-5	3.01e-6	3.68e-7	4.55e-8	5.65e-9	≈ 3.01 (3.00)
	4	1.91e-6	1.10e-7	6.60e-9	4.04e-10	2.50e-11	≈ 4.02 (4.00)
	5	2.88e-7	5.23e-9	1.53e-10	4.64e-12	1.43e-13	≈ 5.03 (5.00)
	6	1.65e-3	3.25e-7	4.47e-12	6.51e-14	9.44e-16	≈ 6.10 (6.00)

We conclude this section with two research directions on time stepping schemes that deserve further investigation.

(i) Nonsmooth error analysis for time-stepping schemes is still in its infancy. So far all known results are only for uniform grids, and although not presented, all the proofs rely essentially on Laplace transform and its discrete analogue. It is of immense interest to develop energy type arguments that yield nonsmooth data error estimates, which, among the potential advantages, might allow deriving results for nonuniform grids. Likewise, correction schemes are also only developed for uniform grids. This is partially due to the fact that the current construction of corrections essentially relies on

Table 6: The L^2-norm error for Example 4.2 at $t_N = 0.01$, by the L1 scheme (4.4) with $h = 1/100$ and $N = 1000 \times 2^k$.

scheme	$\alpha\backslash k$	1	2	3	4	5	6	rate
L1	0.3	1.20e-4	5.97e-5	2.99e-5	1.49e-5	7.46e-6	3.73e-6	≈ 1.00 (1.00)
	0.5	1.70e-4	8.50e-5	4.22e-5	2.10e-5	1.05e-5	5.25e-6	≈ 1.00 (1.00)
	0.7	1.83e-4	9.06e-5	4.50e-5	2.24e-5	1.11e-5	5.55e-6	≈ 1.01 (1.00)
corrected L1	0.3	9.32e-8	3.43e-8	1.20e-8	4.05e-9	1.34e-9	4.34e-10	≈ 1.61 (1.70)
	0.5	1.90e-6	6.90e-7	2.48e-7	8.87e-8	3.16e-8	1.12e-8	≈ 1.49 (1.50)
	0.7	9.83e-6	4.01e-6	1.64e-6	6.66e-7	2.71e-7	1.10e-7	≈ 1.30 (1.30)

Laplace transform of the kernel and its discrete analogue.

(ii) The error estimates are only derived for problems with a time-independent elliptic operator, and there are no analogous results for time-dependent elliptic operators, including time-dependent coefficient and certain nonlinear problems.

5. Space-time formulations

Due to the nonlocality of the fractional derivative $\partial_t^\alpha u$, at each time step one has to use the numerical solutions at all preceding time levels, which incurs huge storage issues. Thus, in terms of storage, the advantages of time stepping schemes, when compared to space-time schemes, are not as pronounced as in the case of standard parabolic problems, and it is very natural to consider space-time discretization. Naturally, any such construction would rely on a proper weak formulation of the fractional derivative, which is only well understood for the Riemann-Liouville derivative ${}^R\partial_t^\alpha u$ at present. Thus, the idea so far is mostly restricted to problem (1.1) with $v = 0$, for which the Riemann-Liouville and Caputo derivatives coincide, and we shall not distinguish the two fractional derivatives in this section. It is noteworthy that there is a very large body of literature on space-time formulations for time fractional evolution equations (or more general space-time fractional models), and we shall only mention two such formulations below. However, the theoretical understanding of these formulations remains rather limited in the context of nonsmooth problem data, and there is a huge demand on developing rigorous theoretical analysis.

Throughout, let $I = (0, T)$, and the space $\widetilde{H}_L^s(I)$ consists of functions whose extension by zero belong to $H^s(-\infty, T)$. On the cylindrical domain $Q_T = \Omega \times I$, we denote the $L^2(Q_T)$-inner product by $(\cdot, \cdot)_{L^2(Q_T)}$.

5.1. Standard Galerkin formulation

In an influential work, Li and Xu [112] proposed one first rigorous space-time formulation for problem (1.1), which was extended and refined by many other researchers (see, e.g., [56, 113, 114, 115, 116] and the references therein). For any $s \in [0, 1]$, we denote by

$$B^s(Q_T) = H^s(I; L^2(\Omega)) \cap L^2(I; H_0^1(\Omega)),$$

with a norm defined by

$$\|v\|_{B^s(Q_T)} = (\|v\|^2_{H^s(I;L^2(\Omega))} + \|v\|^2_{L^2(I;H^1(\Omega))})^{1/2}.$$

The foundation of the method is the following important identity [112, Lemma 2.6]

$$({}^{R}_{0}\partial_t^{\alpha} w, v)_{L^2(I)} = ({}^{R}_{0}\partial_t^{\frac{\alpha}{2}} w, {}^{R}_{t}\partial_T^{\frac{\alpha}{2}} v)_{L^2(I)} \quad \forall w \in \widetilde{H}^1_L(I), v \in \widetilde{H}^{\frac{\alpha}{2}}_L(I), \tag{5.1}$$

where ${}^{R}_{0}\partial_t^{\gamma} w$ and ${}^{R}_{t}\partial_T^{\gamma} w$ denote the left-sided and right-sided Riemann-Liouville fractional derivatives, respectively, and for $\gamma \in (0,1)$, and are defined by

$$\begin{aligned}
{}^{R}_{0}\partial_t^{\gamma} w(t) &= \frac{\mathrm{d}}{\mathrm{d}t}\frac{1}{\Gamma(1-\gamma)}\int_0^t (t-s)^{-\gamma} w(s)\mathrm{d}s,\\
{}^{R}_{t}\partial_T^{\gamma} w(t) &= -\frac{\mathrm{d}}{\mathrm{d}t}\frac{1}{\Gamma(1-\gamma)}\int_t^T (s-t)^{-\gamma} w(s)\mathrm{d}s.
\end{aligned}$$

By multiplying both sides of problem (1.1) with $v \in B^{\frac{\alpha}{2}}(Q_T)$, integrating over the cylindrical domain Q_T, applying the formula (5.1) in time and integration by parts in space, we obtain the following bilinear form on the space $B^{\frac{\alpha}{2}}(Q_T)$:

$$a(u,v) = ({}^{R}_{0}\partial_t^{\frac{\alpha}{2}} u, {}^{R}_{t}\partial_T^{\frac{\alpha}{2}} v)_{L^2(Q_T)} + (\nabla u, \nabla v)_{L^2(Q_T)}.$$

Hence, the weak formulation of problem (1.1) is given by: for $f \in L^2(Q_T)$, find $u \in B^{\frac{\alpha}{2}}(Q_T)$ such that

$$a(u,v) = (f,v)_{L^2(Q_T)} \quad \forall v \in B^{\frac{\alpha}{2}}(Q_T). \tag{5.2}$$

Clearly, the bilinear form $a(\cdot,\cdot)$ is not symmetric, since the Riemann-Liouville derivatives ${}^{R}_{0}\partial_t^{\gamma} u(t)$ and ${}^{R}_{t}\partial_T^{\gamma} u(t)$ differ. Nonetheless, it is continuous on the space $B^{\frac{\alpha}{2}}(Q_T)$. Further, since the inner product $({}^{R}_{0}\partial_t^{\frac{\alpha}{2}} v,\ {}^{R}_{t}\partial_T^{\frac{\alpha}{2}} v)_{L^2(I)}$ involving Riemann-Liouville derivatives actually induces an equivalent norm on the space $H^{\frac{\alpha}{2}}(I)$ (e.g., by means of Fourier transform) (see, e.g., [112, Lemma 2.5] and [117, Lemma 4.2]):

$$({}^{R}_{0}\partial_t^{\frac{\alpha}{2}} v,\ {}^{R}_{t}\partial_T^{\frac{\alpha}{2}} v)_{L^2(I)} \geq \|v\|^2_{H^{\frac{\alpha}{2}}(I)},$$

we have the following coercivity of the bilinear form $a(\cdot,\cdot)$

$$a(u,u) \geq c\|u\|^2_{B^{\frac{\alpha}{2}}(Q_T)}.$$

Then the well-posedness of the weak formulation (5.2) follows directly from Lax-Milgram theorem.

To discretize the weak formulation, Li and Xu [112] employed a spectral approximation for the case of one-dimensional spatial domain Ω. Specifically, let $P_N(I)$ (respectively $P_M(\Omega)$) be the polynomial space of degree less than or equal to N (respectively M) with respect to t (respectively x) (not to be confused with the $L^2(\Omega)$ projection operator P_h on the FEM space X_h in Section 3). For the spectral approximation in space, the authors employ the space $P^0_M(\Omega) := P_M(\Omega) \cap H^1_0(\Omega)$, and since $v = 0$, it is natural to construct the approximation space (in time):

$$P^E_N(I) := \{v \in P_N(I) : v(0) = 0\}.$$

Then for a given pair of integers M, N, let $L := (M, N)$ and the tensor product space $S_L := P_M^0(\Omega) \otimes P_N^E(I) \subset B^{\frac{\alpha}{2}}(Q_T)$. The space-time spectral Galerkin approximation to problem (1.1) reads: find $u_L \in S_L$ such that

$$a(u_L, v_L) = (f, v_L)_{L^2(Q_T)} \quad \forall v_L \in S_L.$$

The well-posedness of the discrete problem follows directly from Lax-Milgram theorem as the continuous case. The authors also provided optimal error estimates in the energy norm. However, the $L^2(Q_T)$ error estimate for the approximation remains unclear, since the regularity of the adjoint problem is not well understood (interestingly, a similar issue arises when applying the Galerkin FEM to two-point boundary value problems involving either the Riemann-Liouville or Caputo derivative [118, 117]). Clearly the construction extends directly to rectangular domains.

Note that in order to achieve high-order convergence, the standard polynomial approximation space requires high regularity of the solution u in time (in usual Sobolev spaces), which is nontrivial to ensure *a priori*, in view of the limited smoothing property of the solution operators. Hence, recently, there has been immense interest in developing schemes that can take care of the solution singularity directly or indirectly. In the context of space-time formulations, singularity enriched trial and/or test spaces, e.g., generalized Jacobi polynomials [119] (including Jacobi poly-fractonomials [120]) and Müntz polynomials [115], are extremely promising and have demonstrated very encouraging numerical results. However, the rigorous convergence analysis of such schemes can be very challenging, and is strikingly missing for nonsmooth problem data.

5.2. Petrov-Galerkin formulation

Now we introduce a Petrov-Galerkin formulation recently developed in [121]. Let $V(Q_T) = L^2(I; H_0^1(\Omega))$ and by $V^*(Q_T)$ its dual, and for any $0 < s < 1$, define the space $B^s(Q_T)$ by

$$B^s(Q_T) = \widetilde{H}_L^s(I; H^{-1}(\Omega)) \cap L^2(I; H_0^1(\Omega)).$$

The space is endowed with the norm

$$\|v\|_{B^s(Q_T)} = (\|\partial_t^s v\|_{V^*(Q_T)}^2 + (\nabla v, \nabla v)_{L^2(Q_T)})^{1/2}.$$

Here we have slightly abused the notation $B^s(Q_T)$ since it differs from that in Section 5.1. Then we define the bilinear form $a(\cdot, \cdot) : B^\alpha(Q_T) \times V(Q_T) \to \mathbb{R}$ by

$$a(v, \phi) := (\partial_t^\alpha v, \phi)_{L^2(Q_T)} + (\nabla u, \nabla v)_{L^2(Q_T)}.$$

The Petrov-Galerkin weak formulation of problem (1.1) reads: find $u \in B^\alpha(Q_T)$ such that

$$a(u, \phi) = (f, \phi)_{L^2(Q_T)} \quad \forall \phi \in V(Q_T). \tag{5.3}$$

The bilinear form $a(\cdot,\cdot)$ is continuous on $B^\alpha(Q_T)\times V(Q_T)$, and it satisfies the following inf-sup condition

$$\sup_{\phi\in V(Q_T)}\frac{a(v,\phi)}{\|\phi\|_{V(Q_T)}}\geq \|v\|_{B^\alpha(Q_T)}\quad \forall v\in B^\alpha(Q_T)$$

and a compatibility condition, i.e., $\sup_{v\in B^\alpha(Q_T)} a(v,\phi)>0$ for any $0\neq\phi\in V(Q_T)$ [121, Lemma 2.4]. Thus the well-posedness of the space-time formulation (5.3) follows directly from the Babuska-Brezzi theory.

Now a Petrov-Galerkin method can be derived as follows. Let X_h be the FEM space defined in (3.1). Also, take a uniform partition of the time interval I with grid points $t_n=n\tau$, $n=0,\ldots,N$, and time step-size $\tau=T/N$. Following [122], define a set of "fractionalized" piecewise constant basis functions $\phi_n(t)$, $n=1,\ldots,N$, by

$$\phi_n(t)=(t-t_{n-1})^\alpha\chi_{[t_{n-1},T]}(t).$$

(recall χ_S denotes the characteristic function of the set S.) It is easy to verify that

$$\phi_n(t)=\Gamma(\alpha+1){}_0I_t^\alpha\chi_{[t_{n-1},T]}(t)\quad\text{ and }\quad \partial_t^\alpha\phi_k(t)=\Gamma(\alpha+1)\chi_{[t_{n-1},T]}(t).$$

Clearly, $\phi_k\in\widetilde{H}_L^{\alpha+s}(0,T)$ for any $s\in[0,1/2)$. One benefit of this construction is to build the potential singularity of form t^α into the basis functions directly, which may allow resolving common solution singularities effectively.

Further, we introduce the following two spaces

$$V_\tau=\operatorname{span}(\{\phi_n(t)\}_{n=1}^N)\quad\text{and}\quad W_\tau:=\operatorname{span}(\{\chi_{[t_{n-1},T]}(t)\}_{n=1}^N).$$

Then the solution space $B_{h,\tau}^\alpha\subset B^\alpha(Q_T)$ and the test space $V_{h,\tau}(Q_T)\subset V(Q_T)$ are respectively defined by the tensor product spaces $B_{h,\tau}^\alpha(Q_T):=X_h\otimes V_\tau$ and $V_{h,\tau}(Q_T):=X_h\otimes W_\tau$. The space-time Petrov-Galerkin FEM problem of (1.1) reads: given $f\in V^*(Q_T)$, find $u_{h\tau}\in B_{h,\tau}^\alpha(Q_T)$ such that

$$a(u_{h\tau},\phi)=(f,\phi)_{L^2(Q_T)}\quad\forall\phi\in V_{h,\tau}(Q_T).\tag{5.4}$$

Algorithmically, it leads to a time-stepping like scheme, and thus admits an efficient practical implementation. The existence and the stability of the solution $u_{h\tau}$ follows from the discrete inf-sup condition [121, Lemma 3.3]

$$\sup_{\phi\in V_{h,\tau}(Q_T)}\frac{a(v,\phi)}{\|\phi\|_{V(Q_T)}}\geq c_\alpha\|v\|_{B^\alpha(Q_T)}\quad\forall v\in B_{h,\tau}^\alpha(Q_T).$$

This condition was shown using the $L^2(I)$ stability of the projection operator from V_τ to W_τ. It is interesting to note that the constant in the $L^2(I)$-stability of the operator depends on the fractional order α and deteriorates as $\alpha\to 1$. Note that for standard parabolic problems ($\alpha=1$), it depends on the time step size τ, leading to an undesirable CFL-condition, a fact shown in [123]. This indicates one significant difference between the fractional model and the standard parabolic model in the context of space-time formulations. In passing, we also note a different Petrov-Galerkin formulation proposed very recently

in [124], whose numerical realization, however, has not been carried out yet and needs computational verification.

Next, we give two error estimates for the space-time Petrov-Galerkin approximation $u_{h\tau}$ obtained by (5.4), [121, Theorems 5.2 and 5.3], in $H^s_L(0,T;L^2(\Omega))$- and $L^2(Q_T)$-norms, respectively.

Theorem 5.1. *Let $f \in \widetilde{H}^s_L(0,T;L^2(\Omega))$ with $0 \le s \le 1$, and u and $u_{h\tau}$ be the solutions of (5.3) and (5.4), respectively. Then there holds*

$$\|u - u_{h\tau}\|_{B^\alpha(Q_T)} \le c(\tau^s + h)\|f\|_{\widetilde{H}^s_L(0,T;L^2(\Omega))},$$
$$\|u - u_{h\tau}\|_{L^2(Q_T)} \le c(\tau^{\alpha+s} + h^2)\|f\|_{\widetilde{H}^s_L(0,T;L^2(\Omega))}.$$

Remark 5.1. *The time stepping schemes in Section 4 rely on sampling the governing equation at time grids and approximation(s) of the derivatives at these grid points, which implicitly assumes $f \in C([0,T];L^2(\Omega))$. In contrast, the space-time formulations can still be well defined for $f \in L^2(0,T;H^s(\Omega))$ for any $s \ge 0$, which hence allows treating problems with very weak source term provided that the relevant integrals can be evaluated with requisite accuracy.*

5.3. Numerical illustrations, comments and research questions

Now we present some numerical results to show the performance of the space-time Petrov-Galerkin FEM, cf. (5.4) with nonsmooth data.

Example 5.1. *Consider problem (1.1) on the unit square domain $\Omega = (0,1)^2$ with $v \equiv 0$ and*

(a) *$f = \sin(t)\chi_{\{x \le \frac{1}{2}\}}(x,y) \in \widetilde{H}^1_L(0,T;\dot{H}^\beta(\Omega))$ with $\beta < 0.5$;*

(b) *$f = t^{-0.2}\chi_{\{x \le \frac{1}{2}\}}(x,y) \in \widetilde{H}^s_L(0,T;\dot{H}^\beta(\Omega))$ with $\beta < 0.5$ and $s < 0.3$.*

To compute the numerical solution $u_{h\tau}$, we fix the spatial mesh size h at $h = 1/160$ and set the temporal step size τ to $\tau = T/N$, where the reference solution u_h is computed using a much finer temporal mesh, i.e., $\tau = T/2560$. The error $\|u_h - u_{h\tau}\|_{L^2(Q_T)}$ with $T = 1$ is presented in Table 7. Irrespective of the compatibility of the source term with the initial data, we observe a convergence rate in the $L^2(Q_T)$-norm of order $O(\tau^{\alpha+1})$ and $O(\tau^{\alpha+0.3})$ for cases (a) and (b), respectively, which fully supports the theoretical results in Theorem 5.1.

We conclude this sections with two important research problems on space-time formulations.

(i) The development of space-time formulations relies crucially on proper variational formulations for the fractional derivative, and this is relatively well understood only for the Riemann-Liouville fractional derivative but not yet for the Caputo one. This is largely the main reason for the restriction to the case of a zero initial data in our discussions. Thus, it is of significant interest to develop proper techniques for handling nonsmooth initial data in the Caputo case. Nonetheless, sufficiently smooth initial data, e.g., $v \in \dot{H}^2(\Omega)$, can be easily accommodated as follows: let $w := u - v$, then w satisfies $w(0) = 0$ and $\partial_t^\alpha w - \Delta w = f + \Delta v$, for which both space-time formulations apply directly.

Table 7: The $L^2(Q_T)$-norm error for Example 5.1 with $T = 1$, by the space-time Petrov-Galerkin FEM scheme with $\tau = T/N$ and $h = 1/160$.

	$\alpha\backslash N$	20	40	80	160	320	640	rate
	0.3	1.46e-2	5.97e-3	2.47e-3	1.04e-3	4.51e-4	2.00e-4	≈ 1.23 (1.30)
(a)	0.5	7.14e-1	2.62e-1	9.59e-2	3.54e-2	1.35e-2	5.48e-3	≈ 1.41 (1.50)
	0.7	5.45e-3	1.84e-3	5.97e-4	1.89e-4	6.10e-5	2.10e-5	≈ 1.61 (1.70)
	0.3	1.59e-1	1.25e-1	9.71e-2	7.53e-2	5.78e-2	4.31e-2	≈ 0.40 (0.60)
(b)	0.5	1.52e-1	1.06e-1	7.23e-2	3.18e-2	2.05e-2	1.33e-2	≈ 0.62 (0.80)
	0.7	1.21e-1	6.93e-2	3.75e-2	1.96e-2	1.01e-2	5.23e-3	≈ 0.95 (1.00)

(ii) Nonpolynomial type approximation spaces for trial and test functions lead to interesting new schemes, supported by extremely promising numerical results. However, the performance may depend strongly on the exponent of the fractional powers, and it would be of substantial interest to analyze the precise influence and to develop strategies to adapt the algorithmic parameter(s) automatically. Many theoretical questions, e.g., unique solvability, stability and optimal convergence rates, surrounding such schemes, of either Galerkin or Petrov-Galerkin type, are largely open.

6. Concluding remarks

In this paper, we have concisely surveyed relevant results in the area of numerical methods for the subdiffusion problem with nonsmooth problem data, with a focus on the state of the art of the following topics: regularity theory, finite element discretization, time-stepping schemes and space-time formulations. We compared the theoretical results with that for standard parabolic problems, and also provided illustrative numerical results. We also outlined a few interesting research problems that would lead to further developments and deeper theoretical understanding, and pointed out the most relevant references. Thus, it may serve as a brief introduction to this fast growing area of numerical analysis.

The subdiffusion model represents one of the simplest models in the zoology of fractional diffusion or anomalous diffusion. The authors believe that many of the analysis may be extended to more complex nonlocal-in-time models, e.g., diffusion wave model, multi-term, distributed-order model, tempered subdiffusion, nonsingular Caputo-Fabrizio fractional derivative, and space-time fractional models. Up to now, these more complex models have received little attention in numerical analysis, have scarcely been studied in the context of nonsmooth problem data, and their distinct features remain largely to be explored both analytically and numerically, in spite of the explosion of their use in mathematical modeling.

Acknowledgements

The research of Z. Zhou is partially supported by a start-up grant from the Hong Kong Polytechnic University and a grant from Research Grants Council of the Hong Kong Special Administrative Region (Project No. 25300818).

References

[1] H. Scher, E. Montroll, Anomalous transit-time dispersion in amorphous solids, Phys. Rev. 12 (6) (1975) 2455–2477. doi:https://doi.org/10.1103/PhysRevB.12.2455.

[2] M. Caputo, Linear models of dissipation whose Q is almost frequency independentII, Geophys. J. Int. 13 (5) (1967) 529–539. doi:10.1111/j.1365-246X.1967.tb02303.x.

[3] M. Ginoa, S. Cerbelli, H. E. Roman, Fractional diffusion equation and relaxation in complex viscoelastic materials, Phys. A 191 (1–4) (1992) 449–453. doi:10.1016/0378-4371(92)90566-9.

[4] R. R. Nigmatullin, The realization of the generalized transfer equation in a medium with fractal geometry, Phys. Stat. Sol. B 133 (1986) 425–430. doi:10.1002/pssb.2221330150.

[5] Y. Hatano, N. Hatano, Dispersive transport of ions in column experiments: an explanation of long-tailed profiles, Water Resources Res. 34 (1998) 1027–1033. doi:10.1029/98WR00214.

[6] S. C. Kou, Stochastic modeling in nanoscale biophysics: subdiffusion within proteins, Ann. Appl. Stat. 2 (2) (2008) 501–535. doi:10.1214/07-AOAS149.

[7] R. Metzler, J. Klafter, The random walk's guide to anomalous diffusion: a fractional dynamics approach, Phys. Rep. 339 (1) (2000) 1–77. doi:10.1016/S0370-1573(00)00070-3.

[8] R. Metzler, J. Klafter, The restaurant at the end of the random walk: recent developments in the description of anomalous transport by fractional dynamics, J. Phys. A 37 (31) (2004) R161–R208. doi:10.1088/0305-4470/37/31/R01.

[9] R. Metzler, J.-H. Jeon, A. G. Cherstvy, E. Barkai, Anomalous diffusion models and their properties: non-stationarity, non-ergodicity, and ageing at the centenary of single particle tracking, Phys. Chem. Chem. Phys. 16 (44) (2014) 24128–24164. doi:10.1039/C4CP03465A.

[10] J. Klafter, I. M. Sokolov, First Steps in Random Walks, Oxford University Press, Oxford, 2011.

[11] A. A. Kilbas, H. M. Srivastava, J. J. Trujillo, Theory and Applications of Fractional Differential Equations, Elsevier Science B.V., Amsterdam, 2006.

[12] M. M. Djrbashian, Harmonic Analysis and Boundary Value Problems in the Complex Domain, Birkhäuser Verlag, Basel, 1993. doi:10.1007/978-3-0348-8549-2.

[13] J. Nakagawa, K. Sakamoto, M. Yamamoto, Overview to mathematical analysis for fractional diffusion equations—new mathematical aspects motivated by industrial collaboration, J. Math-for-Ind. 2A (2010) 99–108.

[14] G. E. Karniadakis, J. S. Hesthaven, I. Podlubny, Special issue on "Fractional PDEs: theory, numerics, and applications" [Editorial], J. Comput. Phys. 293 (2015) 1–3. doi:10.1016/j.jcp.2015.04.007.

[15] B. Jin, R. Lazarov, P. Vabishchevich, Preface: Numerical analysis of fractional differential equations, Comput. Methods Appl. Math. 17 (4) (2017) 643–646. doi:10.1515/cmam-2017-0036.

[16] Z. Zhou, W. Gong, Finite element approximation of optimal control problems governed by time fractional diffusion equation, Comput. Math. Appl. 71 (1) (2016) 301–318. doi:10.1016/j.camwa.2015.11.014.

[17] B. Jin, B. Li, Z. Zhou, Pointwise-in-time error estimates for an optimal control problem with subdiffusion constraint, IMA J. Numer. Anal. (2018) in press, arXiv:1707.08808.

[18] B. Jin, W. Rundell, A tutorial on inverse problems for anomalous diffusion processes, Inverse Problems 31 (3) (2015) 035003, 40. doi:10.1088/0266-5611/31/3/035003.

[19] Y. Yan, Galerkin finite element methods for stochastic parabolic partial differential equations, SIAM J. Numer. Anal. 43 (4) (2005) 1363–1384. doi:10.1137/040605278.

[20] V. Thomée, Galerkin Finite Element Methods for Parabolic Problems, 2nd Edition, Springer-Verlag, Berlin, 2006.

[21] M. Kwaśnicki, Ten equivalent definitions of the fractional laplace operator, Frac. Cal. Appl. Anal. 20 (1) (2017) 7–51. doi:10.1515/fca-2017-0002.

[22] A. Bonito, J. P. Borthagaray, R. H. Nochetto, E. Otárola, A. J. Salgado, Three numerical methods for fractional diffusion, Comput. Vis. Sci. (2018) in press, doi:10.1007/s00791–018–0289–y.

[23] M. Salo, The fractional Calderón problem, Preprint, arXiv:1711.06103 (2017).

[24] X. Ros-Oton, Nonlocal elliptic equations in bounded domains: a survey, Publ. Mat. 60 (1) (2016) 3–26.
URL http://projecteuclid.org/euclid.pm/1450818481

[25] Q. Du, M. Gunzburger, R. B. Lehoucq, K. Zhou, Analysis and approximation of nonlocal diffusion problems with volume constraints, SIAM Rev. 54 (4) (2012) 667–696. doi:10.1137/110833294.

[26] K. Diethelm, N. J. Ford, A. D. Freed, Y. Luchko, Algorithms for the fractional calculus: a selection of numerical methods, Comput. Methods Appl. Mech. Engrg. 194 (6-8) (2005) 743–773. doi:10.1016/j.cma.2004.06.006.

[27] R. A. Adams, J. J. F. Fournier, Sobolev Spaces, 2nd Edition, Elsevier/Academic Press, Amsterdam, 2003.

[28] W. Arendt, C. J. Batty, M. Hieber, F. Neubrander, Vector-valued Laplace Transforms and Cauchy Problems, 2nd Edition, Birkhäuser, Basel, 2011.

[29] R. Gorenflo, J. Loutchko, Y. Luchko, Computation of the Mittag-Leffler function $E_{\alpha,\beta}(z)$ and its derivative, Fract. Calc. Appl. Anal. 5 (4) (2002) 491–518.

[30] H. Seybold, R. Hilfer, Numerical algorithm for calculating the generalized Mittag-Leffler function, SIAM J. Numer. Anal. 47 (1) (2008/09) 69–88. doi:10.1137/070700280.

[31] E. G. Bajlekova, Fractional Evolution Equations in Banach Spaces, Ph.D. thesis, Eindhoven University of Technology (2001).

[32] B. Jin, B. Li, Z. Zhou, Discrete maximal regularity of time-stepping schemes for fractional evolution equations, Numer. Math. 138 (1) (2018) 101–131. doi:10.1007/s00211-017-0904-8.

[33] 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. doi:10.1016/j.jmaa.2011.04.058.

[34] B. Jin, R. Lazarov, J. Pasciak, Z. Zhou, Error analysis of semidiscrete finite element methods for inhomogeneous time-fractional diffusion, IMA J. Numer. Anal. 35 (2) (2015) 561–582. doi:10.1093/imanum/dru018.

[35] M. Stynes, Too much regularity may force too much uniqueness, Fract. Calc. Appl. Anal. 19 (6) (2016) 1554–1562. doi:10.1515/fca-2016-0080.

[36] B. Kovács, B. Li, C. Lubich, A-stable time discretizations preserve maximal parabolic regularity, SIAM J. Numer. Anal. 54 (6) (2016) 3600–3624. doi:10.1137/15M1040918.

[37] G. Akrivis, B. Li, C. Lubich, Combining maximal regularity and energy estimates for time discretizations of quasilinear parabolic equations, Math. Comp. 86 (306) (2017) 1527–1552. doi:10.1090/mcom/3228.

[38] B. Jin, B. Li, Z. Zhou, Numerical analysis of nonlinear subdiffusion equations, SIAM J. Numer. Anal. 56 (1) (2018) 1–23. doi:10.1137/16M1089320.

[39] S. D. Eidelman, A. N. Kochubei, Cauchy problem for fractional diffusion equations, J. Diff. Eq. 199 (2) (2004) 211–255. doi:10.1016/j.jde.2003.12.002.

[40] Y. Luchko, Maximum principle for the generalized time-fractional diffusion equation, J. Math. Anal. Appl. 351 (1) (2009) 218–223. doi:10.1016/j.jmaa.2008.10.018.

[41] A. V. Pskhu, The fundamental solution of a diffusion-wave equation of fractional order, Izv. Ross. Akad. Nauk Ser. Mat. 73 (2) (2009) 141–182. doi:10.1070/IM2009v073n02ABEH002450.

[42] W. McLean, Regularity of solutions to a time-fractional diffusion equation, ANZIAM J. 52 (2) (2010) 123–138. doi:10.1017/S1446181111000617.

[43] A. N. Kochubei, Asymptotic properties of solutions of the fractional diffusion-wave equation, Fract. Calc. Appl. Anal. 17 (3) (2014) 881–896. doi:10.2478/s13540-014-0203-3.

[44] M. Allen, L. Caffarelli, A. Vasseur, A parabolic problem with a fractional time derivative, Arch. Ration. Mech. Anal. 221 (2) (2016) 603–630. doi:10.1007/s00205-016-0969-z.

[45] Y. Liu, W. Rundell, M. Yamamoto, Strong maximum principle for fractional diffusion equations and an application to an inverse source problem, Fract. Calc. Appl. Anal. 19 (4) (2016) 888–906. doi:10.1515/fca-2016-0048.

[46] M. Yamamoto, Weak solutions to non-homogeneous boundary value problems for time-fractional diffusion equations, J. Math. Anal. Appl. 460 (1) (2018) 365–381. doi:10.1016/j.jmaa.2017.11.048.

[47] C. Gal, M. Warma, Fractional-in-time semilinear parabolic equations and applications, Preprint, available at https://hal.archives-ouvertes.fr/hal-01578788 (2017).

[48] J. Prüss, Evolutionary Integral Equations and Applications, Birkhäuser Verlag, Basel, 1993. doi:10.1007/978-3-0348-8570-6.

[49] W. R. Schneider, W. Wyss, Fractional diffusion and wave equations, J. Math. Phys. 30 (1) (1989) 134–144. doi:10.1063/1.528578.

[50] R. Gorenflo, Y. Luchko, M. Yamamoto, Time-fractional diffusion equation in the fractional Sobolev spaces, Fract. Calc. Appl. Anal. 18 (3) (2015) 799–820. doi:10.1515/fca-2015-0048.

[51] L. Li, J.-G. Liu, A generalized definition of Caputo derivatives and its application to fractional ODEs, SIAM J. Math. Anal. 50 (3) (2018) 2867–2900. doi:10.1137/17M1160318.

[52] V. Vergara, R. Zacher, Optimal decay estimates for time-fractional and other nonlocal subdiffusion equations via energy methods, SIAM J. Math. Anal. 47 (1) (2015) 210–239. doi:10.1137/130941900.

[53] I. Kim, K.-H. Kim, S. Lim, An $L_q(L_p)$-theory for the time fractional evolution equations with variable coefficients, Adv. Math. 306 (2017) 123–176. doi:10.1016/j.aim.2016.08.046.

[54] Y. Lin, C. Xu, Finite difference/spectral approximations for the time-fractional diffusion equation, J. Comput. Phys. 225 (2) (2007) 1533–1552. doi:10.1016/j.jcp.2007.02.001.

[55] Y.-N. Zhang, Z.-Z. Sun, Alternating direction implicit schemes for the two-dimensional fractional subdiffusion equation, J. Comput. Phys. 230 (24) (2011) 8713–8728. doi:10.1016/j.jcp.2011.08.020.

[56] F. Chen, Q. Xu, J. S. Hesthaven, A multi-domain spectral method for time-fractional differential equations, J. Comput. Phys. 293 (2015) 157–172. doi:10.1016/j.jcp.2014.10.016.

[57] 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. doi:10.1137/120873984.

[58] B. Jin, R. Lazarov, J. Pasciak, Z. Zhou, Galerkin FEM for fractional order parabolic equations with initial data in H^{-s}, $0 \leq s \leq 1$, in: Numerical Analysis and its Applications, Vol. 8236 of Lecture Notes in Comput. Sci., Springer, Heidelberg, 2013, pp. 24–37. doi:10.1007/978-3-642-41515-9_3.

[59] K. N. Le, W. McLean, B. Lamichhane, Finite element approximation of a time-fractional diffusion problem for a domain with a re-entrant corner, ANZIAM J. 59 (1) (2017) 61–82. doi:10.1017/S1446181116000365.

[60] H. Fujita, T. Suzuki, Evolution problems, in: Handbook of Numerical Analysis, Vol. II, Handb. Numer. Anal., II, North-Holland, Amsterdam, 1991, pp. 789–928.

[61] S. Karaa, K. Mustapha, A. K. Pani, Finite volume element method for two-dimensional fractional subdiffusion problems, IMA J. Numer. Anal. 37 (2) (2017) 945–964. doi:10.1093/imanum/drw010.

[62] N. Kopteva, Error analysis of the L1 method on graded and uniform meshes for a fractional-derivative problem in two and three dimensions, Preprint, arXiv:1709.09136 (2017).

[63] P. Chatzipantelidis, Z. Horváth, V. Thomée, On preservation of positivity in some finite element methods for the heat equation, Comput. Methods Appl. Math. 15 (4) (2015) 417–437. doi:10.1515/cmam-2015-0018.

[64] B. Jin, R. Lazarov, V. Thomée, Z. Zhou, On nonnegativity preservation in finite element methods for subdiffusion equations, Math. Comp. 86 (307) (2017) 2239–2260. doi:10.1090/mcom/3167.

[65] B. R. Baliga, S. V. Patankar, A new finite-element formulation for convection-diffusion problems, Numerical Heat Transfer 3 (4) (1980) 393–409. doi:10.1080/01495728008961767.

[66] S.-H. Chou, Q. Li, Error estimates in L^2, H^1 and L^∞ in covolume methods for elliptic and parabolic problems: a unified approach, Math. Comp. 69 (229) (2000) 103–120. doi:10.1090/S0025-5718-99-01192-8.

[67] P. Chatzipantelidis, R. Lazarov, V. Thomée, Some error estimates for the finite volume element method for a parabolic problem, Comput. Methods Appl. Math. 13 (3) (2013) 251–279. doi:10.1515/cmam-2012-0006.

[68] P. Chatzipantelidis, R. Lazarov, V. Thomée, Some error estimates for the lumped mass finite element method for a parabolic problem, Math. Comp. 81 (277) (2012) 1–20. doi:10.1090/S0025-5718-2011-02503-2.

[69] S. Karaa, Semidiscrete finite element analysis of time fractional parabolic problems: a unified approach, SIAM J. Numer. Anal. 56 (3) (2018) 1673–1692. doi:10.1137/17M1134160.

[70] S. Karaa, K. Mustapha, A. K. Pani, Optimal error analysis of a FEM for fractional diffusion problems by energy arguments, J. Sci. Comput. 74 (1) (2018) 519–535. doi:10.1007/s10915-017-0450-7.

[71] S. Karaa, A. K. Pani, Error analysis of a FVEM for fractional order evolution equations with nonsmooth initial data, ESAIM: Math. Model. Numer. Anal. 52 (2) (2018) 773–801. doi:10.1051/m2an/2018029.

[72] M. Luskin, R. Rannacher, On the smoothing property of the Galerkin method for parabolic equations, SIAM J. Numer. Anal. 19 (1) (1982) 93–113. doi:10.1137/0719003.

[73] K. Mustapha, FEM for time-fractional diffusion equations, novel optimal error analyses, Math. Comp. 87 (313) (2018) 2259–2272. doi:10.1090/mcom/3304.

[74] A. Quarteroni, A. Valli, Numerical Approximation of Partial Differential Equations, Springer-Verlag, Berlin, 1994.

[75] Y.-N. Zhang, Z.-Z. Sun, H.-L. Liao, Finite difference methods for the time fractional diffusion equation on non-uniform meshes, J. Comput. Phys. 265 (2014) 195–210. doi:10.1016/j.jcp.2014.02.008.

[76] M. Stynes, E. O'Riordan, J. L. Gracia, Error analysis of a finite difference method on graded meshes for a time-fractional diffusion equation, SIAM J. Numer. Anal. 55 (2) (2017) 1057–1079. doi:10.1137/16M1082329.

[77] H.-L. Liao, D. Li, J. Zhang, Sharp error estimate of the nonuniform L1 formula for linear reaction-subdiffusion equations, SIAM J. Numer. Anal. 56 (2) (2018) 1112–1133. doi:10.1137/17M1131829.

[78] C. Lubich, Discretized fractional calculus, SIAM J. Math. Anal. 17 (3) (1986) 704–719. doi:10.1137/0517050.

[79] C. Lubich, Convolution quadrature and discretized operational calculus. I, Numer. Math. 52 (2) (1988) 129–145. doi:10.1007/BF01398686.

[80] C. Lubich, Convolution quadrature revisited, BIT 44 (3) (2004) 503–514. doi:10.1023/B:BITN.0000046813.23911.2d.

[81] S. B. Yuste, L. Acedo, An explicit finite difference method and a new von Neumann-type stability analysis for fractional diffusion equations, SIAM J. Numer. Anal. 42 (5) (2005) 1862–1874. doi:10.1137/030602666.

[82] F. Zeng, C. Li, F. Liu, I. Turner, The use of finite difference/element approaches for solving the time-fractional subdiffusion equation, SIAM J. Sci. Comput. 35 (6) (2013) A2976–A3000. doi:10.1137/130910865.

[83] B. Jin, R. Lazarov, Z. Zhou, Two fully discrete schemes for fractional diffusion and diffusion-wave equations with nonsmooth data, SIAM J. Sci. Comput. 38 (1) (2016) A146–A170. doi:10.1137/140979563.

[84] K. B. Oldham, J. Spanier, The Fractional Calculus, Academic Press, New York-London, 1974, theory and applications of differentiation and integration to arbitrary order.

[85] I. Podlubny, Fractional Differential Equations, Academic Press, Inc., San Diego, CA, 1999.

[86] E. Sousa, How to approximate the fractional derivative of order $1 < \alpha \leq 2$, Internat. J. Bifur. Chaos Appl. Sci. Engrg. 22 (4) (2012) 1250075, 13. doi:10.1142/S0218127412500757.

[87] E. Cuesta, C. Lubich, C. Palencia, Convolution quadrature time discretization of fractional diffusion-wave equations, Math. Comp. 75 (254) (2006) 673–696. doi:10.1090/S0025-5718-06-01788-1.

[88] W. Cao, F. Zeng, Z. Zhang, G. E. Karniadakis, Implicit-explicit difference schemes for nonlinear fractional differential equations with nonsmooth solutions, SIAM J. Sci. Comput. 38 (5) (2016) A3070–A3093. doi:10.1137/16M1070323.

[89] C. Lubich, I. H. Sloan, V. Thomée, Nonsmooth data error estimates for approximations of an evolution equation with a positive-type memory term, Math. Comp. 65 (213) (1996) 1–17. doi:10.1090/S0025-5718-96-00677-1.

[90] B. Jin, B. Li, Z. Zhou, Correction of high-order BDF convolution quadrature for fractional evolution equations, SIAM J. Sci. Comput. 39 (6) (2017) A3129–A3152. doi:10.1137/17M1118816.

[91] C. Lubich, A. Ostermann, Runge-Kutta approximation of quasi-linear parabolic equations, Math. Comp. 64 (210) (1995) 601–627. doi:10.2307/2153442.

[92] M. Fischer, Fast and parallel Runge-Kutta approximation of fractional evolution equations, Preprint, arXiv:1803.05335 (2018).

[93] M. Chen, W. Deng, Fourth order accurate scheme for the space fractional diffusion equations, SIAM J. Numer. Anal. 52 (3) (2014) 1418–1438. doi:10.1137/130933447.

[94] Y. Dimitrov, Numerical approximations for fractional differential equations, J. Fract. Calc. Appl. 5 (suppl. 3S) (2014) Paper no. 22, 45 pp.

[95] B. Jin, B. Li, Z. Zhou, An analysis of the Crank–Nicolson method for subdiffusion, IMA J. Numer. Anal. 38 (1) (2018) 518–541. doi:10.1093/imanum/drx019.

[96] Z.-Z. Sun, X. Wu, A fully discrete difference scheme for a diffusion-wave system, Appl. Numer. Math. 56 (2) (2006) 193–209. doi:10.1016/j.apnum.2005.03.003.

[97] B. Jin, R. Lazarov, Z. Zhou, An analysis of the L1 scheme for the subdiffusion equation with nonsmooth data, IMA J. Numer. Anal. 36 (1) (2016) 197–221. doi:10.1093/imanum/dru063.

[98] B. Jin, Z. Zhou, An analysis of Galerkin proper orthogonal decomposition for subdiffusion, ESAIM Math. Model. Numer. Anal. 51 (1) (2017) 89–113. doi:10.1051/m2an/2016017.

[99] Y. Yan, M. Khan, N. J. Ford, An analysis of the modified L1 scheme for time-fractional partial pifferential equations with nonsmooth data, SIAM J. Numer. Anal. 56 (1) (2018) 210–227. doi:10.1137/16M1094257.

[100] Y. Xing, Y. Yan, A higher order numerical method for time fractional partial differential equations with nonsmooth data, J. Comput. Phys. 357 (2018) 305–323. doi:10.1016/j.jcp.2017.12.035.

[101] N. J. Ford, Y. Yan, An approach to construct higher order time discretisation schemes for time fractional partial differential equations with nonsmooth data, Frac. Calc. Appl. Anal. 20 (5) (2017) 1076–1105. doi:10.1515/fca-2017-0058.

[102] C. Lv, C. Xu, Error analysis of a high order method for time-fractional diffusion equations, SIAM J. Sci. Comput. 38 (5) (2016) A2699–A2724. doi:10.1137/15M102664X.

[103] G.-H. Gao, Z.-Z. Sun, H.-W. Zhang, A new fractional numerical differentiation formula to approximate the Caputo fractional derivative and its applications, J. Comput. Phys. 259 (2014) 33–50. doi:10.1016/j.jcp.2013.11.017.

[104] K. Mustapha, B. Abdallah, K. M. Furati, A discontinuous Petrov-Galerkin method for time-fractional diffusion equations, SIAM J. Numer. Anal. 52 (5) (2014) 2512–2529. doi:10.1137/140952107.

[105] A. A. Alikhanov, A new difference scheme for the time fractional diffusion equation, J. Comput. Phys. 280 (2015) 424–438. doi:10.1016/j.jcp.2014.09.031.

[106] W. McLean, K. Mustapha, Convergence analysis of a discontinuous Galerkin method for a subdiffusion equation, Numer. Algorithms 52 (1) (2009) 69–88. doi:10.1007/s11075-008-9258-8.

[107] K. Mustapha, W. McLean, Piecewise-linear, discontinuous Galerkin method for a fractional diffusion equation, Numer. Algorithms 56 (2) (2011) 159–184. doi:10.1007/s11075-010-9379-8.

[108] K. Mustapha, Time-stepping discontinuous Galerkin methods for fractional diffusion problems, Numer. Math. 130 (3) (2015) 497–516. doi:10.1007/s00211-014-0669-2.

[109] W. McLean, K. Mustapha, Time-stepping error bounds for fractional diffusion problems with non-
715 smooth initial data, J. Comput. Phys. 293 (2015) 201–217. doi:10.1016/j.jcp.2014.08.050.

[110] Y. Yang, Y. Yan, N. J. Ford, Some time stepping methods for fractional diffusion problems with non-smooth data, Comput. Methods Appl. Math. 18 (1) (2018) 129–146. doi:10.1515/cmam-2017-0037.

[111] M. Gunzburger, J. Wang, A second-order Crank-Nicolson scheme for time-fractional PDEs, Int. J. Numer. Anal. & Model. (2017) in press.

720 [112] X. Li, C. Xu, A space-time spectral method for the time fractional diffusion equation, SIAM J. Numer. Anal. 47 (3) (2009) 2108–2131. doi:10.1137/080718942.

[113] M. Zayernouri, M. Ainsworth, G. E. Karniadakis, A unified Petrov-Galerkin spectral method for fractional PDEs, Comput. Methods Appl. Mech. Engrg. 283 (2015) 1545–1569. doi:10.1016/j.cma.2014.10.051.

725 [114] C. Huang, Z. Zhang, Convergence of a p-version/hp-version method for fractional differential equations, J. Comput. Phys. 286 (2015) 118–127. doi:10.1016/j.jcp.2015.01.025.

[115] D. Hou, M. T. Hasan, C. Xu, Müntz spectral methods for the time-fractional diffusion equation, Comput. Methods Appl. Math. 18 (1) (2018) 43–62. doi:10.1515/cmam-2017-0027.

[116] B. Li, H. Luo, X. Xie, Analysis of a time-stepping scheme for time fractional diffusion problems with
730 nonsmooth data, Preprint, arXiv:1804.10552 (2018).

[117] B. Jin, R. Lazarov, J. Pasciak, W. Rundell, Variational formulation of problems involving fractional order differential operators, Math. Comp. 84 (296) (2015) 2665–2700. doi:10.1090/mcom/2960.

[118] V. J. Ervin, J. P. Roop, Variational formulation for the stationary fractional advection dispersion equation, Numer. Methods Part. Diff. Eq. 22 (3) (2006) 558–576. doi:10.1002/num.20112.

735 [119] S. Chen, J. Shen, L.-L. Wang, Generalized Jacobi functions and their applications to fractional differential equations, Math. Comp. 85 (300) (2016) 1603–1638. doi:10.1090/mcom3035.

[120] M. Zayernouri, G. E. Karniadakis, Fractional Sturm-Liouville eigen-problems: theory and numerical approximation, J. Comput. Phys. 252 (2013) 495–517. doi:10.1016/j.jcp.2013.06.031.

[121] B. Duan, B. Jin, R. Lazarov, J. Pasciak, Z. Zhou, Space-time Petrov-Galerkin FEM for fractional dif-
740 fusion problems, Comput. Methods Appl. Math. 18 (1) (2018) 1–20. doi:10.1515/cmam-2017-0026.

[122] B. Jin, R. Lazarov, Z. Zhou, A Petrov-Galerkin finite element method for fractional convection-diffusion equations, SIAM J. Numer. Anal. 54 (1) (2016) 481–503. doi:10.1137/140992278.

[123] S. Larsson, M. Molteni, Numerical solution of parabolic problems based on a weak space-time formulation, Comput. Methods Appl. Math. 17 (1) (2017) 65–84. doi:10.1515/cmam-2016-0027.

[124] M. Karkulik, Variational formulation of time-fractional parabolic equations, Comput. Math. Appl. 75 (11) (2018) 3929–3938. doi:10.1016/j.camwa.2018.03.003.