HIGH-ORDER SPLITTING FINITE ELEMENT METHODS FOR THE SUBDIFFUSION EQUATION WITH LIMITED SMOOTHING PROPERTY

BUYANG LI, ZONGZE YANG, AND ZHI ZHOU

ABSTRACT. In contrast with the diffusion equation which smoothens the initial data to C^∞ for $t > 0$ (away from the corners/edges of the domain), the subdiffusion equation only exhibits limited spatial regularity. As a result, one generally cannot expect high-order accuracy in space in solving the subdiffusion equation with nonsmooth initial data. In this paper, a new splitting of the solution is constructed for high-order finite element approximations to the subdiffusion equation with nonsmooth initial data. The method is constructed by splitting the solution into two parts, i.e., a time-dependent smooth part and a time-independent nonsmooth part, and then approximating the two parts via different strategies. The time-dependent smooth part is approximated by using high-order finite element method in space and convolution quadrature in time, while the steady nonsmooth part could be approximated by using smaller mesh size or other methods that could yield high-order accuracy. Several examples are presented to show how to accurately approximate the steady nonsmooth part, including piecewise smooth initial data, Dirac–Delta point initial data, and Dirac measure concentrated on an interface. The argument could be directly extended to subdiffusion equations with nonsmooth source data. Extensive numerical experiments are presented to support the theoretical analysis and to illustrate the performance of the proposed high-order splitting finite element methods.

1. Introduction

This article is concerned with the construction and analysis of high-order finite element methods for solving the subdiffusion equation in a polygonal/polyhedral domain $\Omega \subset \mathbb{R}^d$, with $d \geq 1$, i.e.,

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

where f and u^0 are given source function and initial value, respectively, $\Delta : H^2(\Omega) \cap H_0^1(\Omega) \to L^2(\Omega)$ is the Dirichlet Laplacian operator, $T \in (0, \infty]$ represents a specified terminal time. Here $\partial_t^\alpha u$ denotes the Djrbashian–Caputo fractional time derivative of order $\alpha \in (0,1)$ [22, p. 92] and [13, Section 2.3]:

$$\partial_t^\alpha u(t) = \frac{1}{\Gamma(1-\alpha)} \int_0^t (t-s)^{-\alpha} u'(s)\, \mathrm{d}s \quad \text{with} \quad \Gamma(1-\alpha) = \int_0^\infty s^{-\alpha} e^{-s} \mathrm{d}s;$$

The subdiffusion equation in (1.1) has received much attention in recent years in physics, engineering, biology and finance, due to their capability for describing anomalously slow diffusion processes, also known as subdiffusion. At a microscopic level, subdiffusive processes can be described by continuous time random walk with a heavy-tailed waiting time distribution, which displays local motion occasionally interrupted by long sojourns and trapping effects. These transport processes are characterized by a sublinear growth of the mean squared displacement of the particle with the time, as opposed to linear growth for Brownian motion. The model (1.1) has found many successful

2010 *Mathematics Subject Classification.* Primary: 65M30, 65M15, 65M12.

Key words and phrases. subdiffusion, limited smoothing property, nonsmooth data, finite element method, high-order, convolution quadrature, error estimate.

practical applications, e.g., subsurface flows [1], thermal diffusion in media with fractal geometry [39], transport column experiments [12] and heat conduction with memory [46], to name but a few. See [35] for physical modeling and a long list of applications.

One of the main difficulties in the numerical approximation to the subdiffusion equation, compared to the standard parabolic equations, is the weak singularity of the solution at $t = 0$ and the limited regularity pick up with respect to the initial data. In general, for the subdiffusion equation with a nonsmooth initial value $u^0 \in L^2(\Omega)$ and a temporally smooth source function $f(x, t)$, the solution generally exhibits the following type of weak singularity at $t = 0$:

$$\|\partial_t^m u(\cdot, t)\|_{L^2} \le C_m t^{-m} \quad \text{for } m \ge 0, \tag{1.2}$$

and the spatial regularity pick up is limited to

$$\|u(\cdot, t)\|_{H^2} \le Ct^{-1}. \tag{1.3}$$

Higher-order spatial regularity generally cannot be expected for $u^0 \in L^2(\Omega)$. This limited smoothing property was shown in the paper [42] of Sakamoto and Yamamoto with the following two-sided stability (with $f \equiv 0$):

$$\frac{c_1}{1+t^\alpha}\|u^0\|_{\dot{H}^s(\Omega)} \le \|u(t)\|_{\dot{H}^{s+2}(\Omega)} \le c_2 t^{-\alpha}\|u^0\|_{\dot{H}^s(\Omega)}. \tag{1.4}$$

The limited regularity of the solution, as shown in (1.2)–(1.4), causes many difficulties in developing high-order temporal and spatial discretizations for the subdiffusion equation when the initial data is nonsmooth. Many efforts have been made in overcoming these difficulties. In particular, high-order temporal discretizations for the subdiffusion equation have been developed based on graded mesh in [23, 25, 38, 44] for $u^0 \in H_0^1(\Omega) \cap H^2(\Omega)$, discontinuos Galerkin method in [37] for $u^0 \in H^{5/2}(\Omega) \cap H_0^1(\Omega)$ plus a compatibility condition $\Delta u^0 = 0$ on $\partial\Omega$, BDF convolution quadrature in [17] and Runge–Kutta convolution quadrature in [3] for $u^0 \in L^2(\Omega)$, exponential convolution quadrature [31] and exponential spectral method [30] for semilinear problems with $u^0 \in L^\infty(\Omega)$. See also [4, 8, 24, 26] for *a posteriori* error analysis of several popular time stepping schemes. We refer interested reader to the recent monograph [18] for more detailed information.

The spatial discretization using the standard Galerkin finite element method (FEM) or the lumped mass Galerkin FEM for solving the subdiffusion equation with nonsmooth initial data was studied in [15]. The second order convergence in $L^2(\Omega)$ norm was established and it is optimal with respect to the $L^2(\Omega)$ initial data. In these works, the error analysis was carried using the Mittag–Leffler functions. Error estimates of the standard Galerkin FEM with initial data in $\dot{H}^q(\Omega)$ with $q \in (-1, 0)$ can be found in [14]. The Laplace transform approach, initially introduced for parabolic equations by Fujita and Suzuki [9, Chapter 7], was adapted to the subdiffusion equation in [18, Chapter 2.3] to remove a logarithmic factor in the error estimates. See [16] and [18, Chapter 2] for concise overviews. The energy argument, which represents one of the most commonly used strategies for standard parabolic equations, is much more involved for the subdiffusion equation. This is due to the nonlocality of the fractional derivative $\partial_t^\alpha u$, which causes that many useful tools (such as integration by parts formula and product rule) are either invalid or requiring substantial modifications. Some first encouraging theoretical results in this important direction were obtained by Mustapha [36], where optimal error estimates for the homogeneous problem were obtained using an energy argument; see also [28] for the time-fractional Fokker–Planck equation. A unified analysis of different kinds of FEMs for the homogeneous subdiffusion problem based on an energy argument, which generalizes the corresponding technique for standard parabolic problems in [45, Chapter 3], were given by Karaa [19]. The numerical analysis of subdiffusion equations with irregular domains and finite volume element methods were discussed in [27] and [20, 21], respectively.

In all the aforementioned work for the subdiffusion equation, people only considered the error analysis of the piecewise linear finite element method for nonsmooth initial data. This is attributed to the limited smoothing property of the subdiffusion equation, which only smoothen the initial

data $u^0 \in L^2(\Omega)$ to $u(t) \in H^2(\Omega)$ for $t > 0$; cf. (1.4) with $s = 0$. This is in sharp contrast to the standard parabolic equation, which smoothen the initial data $u^0 \in L^2(\Omega)$ to $u(t) \in C^\infty(\Omega)$ for $t > 0$. Consequently, the standard finite element approximation to the subdiffusion equation only has second-order convergence in $L^2(\Omega)$ for initial data in $L^2(\Omega)$. The development of high-order finite element methods in space in case of nonsmooth initial data, a common scenario in associated inverse problems [42, 48, 49], remains challenging and missing in the literature. If the problem data is smooth and compatible to the boundary condition, then the solution could be smooth enough. In this case, it is possible to construct high-order spatial approximation. For example, the high-order hybridizable discontinuous Galerkin method was proposed and analyzed in [5] under the assumption that the solution is smooth enough.

In this paper, we construct a splitting method which allows to develop high-order finite element methods for solving the subdiffusion equation with nonsmooth initial data. For $u^0 \in L^2(\Omega)$, we split the solution into a time-dependent regular part $u^r(t) \in \dot{H}^{2m+2}(\Omega)$ and several time-independent singular parts u_j^s with $j = 1, \ldots, m$. Then the time-dependent smooth part $u^r(t)$ is approximated by using high-order finite element methods in space and convolution quadrature in time generated by k-step BDF method, with $k = 1, 2, \ldots, 6$. If we denote by $U_h^{r,n}$ the fully discrete solution approximating $u^r(t_n)$, then the following result is proved (see Theorems 2.1 and 3.1):

$$\|U_h^{r,n} - u^r(t_n)\|_{L^2(\Omega)} \le c\Big(h^{2m+2} t_n^{-(1+m)\alpha} + \tau^k t_n^{-k-m\alpha}\Big)\|u^0\|_{L^2(\Omega)} \quad \text{for all } t \in (0, T].$$

Note that the integer m could be arbitrarily large, and hence we have developed an arbitrarily high-order FEM approximation for the smooth part. This argument also works for weaker initial data $u^0 \in \dot{H}^s(\Omega)$ with $s < 0$. Meanwhile, the singular parts u_j^s are independent of time and they can be approximated by solving several elliptic equations with nonsmooth sources. This is illustrated in Section 4 for several exemplary nonsmooth data, e.g., piecewise smooth functions, Dirac–Delta point source, and Dirac measure concentrated on an interface. More generally, the time-independent nonsmooth part can be approximated by the standard FEM using smaller mesh size without increasing the overall computational cost significantly. This is possible as the nonsmooth part is time-independent and therefore avoids the time stepping procedure. As a result, the high-order finite element approximation to the subdiffusion equation can be realized by the novel splitting strategy. This strategy works for all second-order elliptic operators with smooth coefficients even though we only consider the negative Laplacian $-\Delta$ for simplicity of presentation. As far as we know, this is the first attempt to develop spatially high-order finite element methods for the subdiffusion equation with nonsmooth initial data. In addition, the argument in this paper can be easily extended to the case of nonsmooth source term.

The rest of the paper is organized as follows. In Section 2, we present the splitting method and the high-order finite element approximation to the regular part of the solution. High-order time-stepping schemes for the regular part and the corresponding error estimates are presented in Section 3. High-order finite element approximations to the singular parts are discussed in Section 4. The extension to nonsmooth source term is discussed in Section 4. Finally, in Section 6, we present several numerical examples to illustrate the high-order convergence of the proposed splitting FEMs in comparison with the standard Galerkin FEMs for the subdiffusion equation.

2. Construction of high-order spatial discretizations

Let $\{\lambda_j\}_{j=1}^\infty$ and $\{\varphi_j\}_{j=1}^\infty$ be the eigenvalues (ordered nondecreasingly with multiplicity counted) and the $L^2(\Omega)$-orthonormal eigenfunctions, respectively, of the elliptic operator $A = -\Delta : H^2(\Omega) \cap H_0^1(\Omega) \to L^2(\Omega)$ under the zero boundary condition. Then $\{\varphi_j\}_{j=1}^\infty$ forms an orthonormal basis in $L^2(\Omega)$. For any real number $s \ge 0$, we denote by $\dot{H}^s(\Omega)$ the Hilbert space with the induced norm

$\|\cdot\|_{\dot{H}^s(\Omega)}$ defined by

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

In particular, $\|v\|_{\dot{H}^0(\Omega)} = \|v\|_{L^2(\Omega)} = (v,v)^{\frac{1}{2}}$ is the norm in $L^2(\Omega)$. For $s<0$, we define the space $\dot{H}^s(\Omega) = \dot{H}^{-s}(\Omega)'$. It is straightforward to verify that $\|v\|_{\dot{H}^1(\Omega)} = \|\nabla v\|_{L^2(\Omega)}$ is an equivalent norm of $H_0^1(\Omega)$ and $\|v\|_{\dot{H}^2(\Omega)} = \|Av\|_{L^2(\Omega)}$ is a norm of $H^2(\Omega)\cap H_0^1(\Omega)$; see [45, Section 3.1]. Moreover, for any integer $m\geq 0$, $v\in \dot{H}^{2m+2}(\Omega)$ if and only if $A^{j-1}v \in H^2(\Omega)\cap H_0^1(\Omega)$ and $A^j v\in \dot{H}^{2m+2-2j}(\Omega)$ for $j=1,\ldots,m+1$, and

$$\|v\|_{\dot{H}^{2m+2}(\Omega)} \sim \|A^{m+1}v\|_{L^2(\Omega)} \quad \forall\, v\in \dot{H}^{2m+2}(\Omega).$$

It is known that the solution to problem (1.1) can be written as (cf. [13, eq. (6.24)])

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

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

$$F(t) := \frac{1}{2\pi \mathrm{i}}\int_{\Gamma_{\theta,\kappa}} e^{zt} z^{\alpha-1}(z^\alpha + A)^{-1}\,\mathrm{d}z \quad \text{and} \quad E(t) := \frac{1}{2\pi\mathrm{i}}\int_{\Gamma_{\theta,\kappa}} e^{zt}(z^\alpha + A)^{-1}\,\mathrm{d}z, \tag{2.2}$$

respectively, with integration over a contour $\Gamma_{\theta,\kappa}$ in the complex plane $\mathbb{C}$ (oriented counterclockwise), defined by $\Gamma_{\theta,\kappa} = \{z\in\mathbb{C}: |z|=\kappa, |\arg z|\leq\theta\}\cup\{z\in\mathbb{C}: z=\rho e^{\pm \mathrm{i}\theta}, \rho\geq\kappa\}$. 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 $z\in\Sigma_\theta$. The following resolvent estimat will be frequently used (see [2, Example 3.7.5 and Theorem 3.7.11]):

$$\|(z+A)^{-1}\| \leq c_\theta |z|^{-1}, \quad \forall z\in\Sigma_\theta,\ \forall\,\theta\in(0,\pi), \tag{2.3}$$

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

Equivalently, the solution operators in (2.2) can also 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 [22, Section 1.8, pp. 40-45]. The Mittag–Leffler function $E_{\alpha,\beta}(z)$ is a generalization of the exponential function e^z appearing in normal diffusion. For any $\alpha\in(0,1)$, the function $E_{\alpha,1}(-\lambda t^\alpha)$ decays only polynomially like $\lambda^{-1}t^{-\alpha}$ as $\lambda, t\to\infty$ [22, equation (1.8.28), p. 43], which contrasts sharply with the exponential decay for $e^{-\lambda t}$ appearing in normal diffusion. This important feature directly translates into the limited smoothing property in space for the solution operators $F(t)$ [42] . As a result, in general, we cannot expect high-order approximations for subdiffusion problem (1.1) with nonsmooth data. In fact, our preceding study [14] shows an optimal convergence rate $O(h^{2-s})$ for piecewise linear finite element approximation, when $u^0\in\dot{H}^{-s}(\Omega)$.

2.1. A new splitting of the solution

For the simplicity of presentation, we consider the homogeneous subdiffusion equation with $f\equiv 0$ and $u^0\in L^2(\Omega)$. The argument could be extended to the weaker case $u^0\in H^s$ with $s\in[-1,0)$ by slight modifications. The case with nonsmooth $f\neq 0$ will be discussed in Section 4.

We split the integrand of (2.2) into two parts based on the following relation:

$$(z^\alpha + A)^{-1} = A^{-1} - z^\alpha(z^\alpha + A)^{-1}A^{-1}. \tag{2.4}$$

This leads to the following splitting of the solution operator:

$$F(t) := \frac{1}{2\pi \mathrm{i}} \int_{\Gamma_{\theta,\kappa}} e^{zt} z^{\alpha-1} (z^\alpha + A)^{-1} \,\mathrm{d}z = \frac{1}{2\pi \mathrm{i}} \int_{\Gamma_{\theta,\kappa}} e^{zt} \left[z^{\alpha-1} A^{-1} - z^{2\alpha-1} (z^\alpha + A)^{-1} A^{-1} \right] \mathrm{d}z$$
$$= \frac{t^{-\alpha}}{\Gamma(1-\alpha)} A^{-1} - \frac{1}{2\pi \mathrm{i}} \int_{\Gamma_{\theta,\kappa}} e^{zt} z^{2\alpha-1} (z^\alpha + A)^{-1} A^{-1} \,\mathrm{d}z.$$

In the case $f \equiv 0$ we obtain the following splitting of the solution:

$$u(t) = F(t)u^0 = u^s + u^r(t) \tag{2.5}$$

with

$$u^s = \frac{1}{\Gamma(1-\alpha)} A^{-1} t^{-\alpha} u^0 \quad \text{and} \quad u^r(t) = -\frac{1}{2\pi \mathrm{i}} \int_{\Gamma_{\theta,\kappa}} e^{zt} z^{2\alpha-1} (z^\alpha + A)^{-1} A^{-1} u^0 \,\mathrm{d}z,$$

denoting the singular and regular parts in this splitting, respectively. This splitting process can be continued by substituting relation (2.4) into the expression of the regular part repeatedly. Then we can obtain the following higher-order splitting:

$$u(t) = u^r(t) + \sum_{j=1}^{m} u_j^s \tag{2.6}$$

with

$$\begin{aligned} u_j^s &= (-1)^{j+1} \frac{t^{-j\alpha}}{\Gamma(1-j\alpha)} A^{-j} u^0, \quad \text{for } j = 1, 2, \ldots, m, \\ u^r(t) &= F^r(t)u^0 := \frac{(-1)^m}{2\pi \mathrm{i}} \int_{\Gamma_{\theta,\kappa}} e^{zt} z^{(1+m)\alpha-1} (z^\alpha + A)^{-1} A^{-m} u^0 \,\mathrm{d}z. \end{aligned} \tag{2.7}$$

Note that the singular part $u_j^s(t)$ is a solution of an elliptic problem, while the regular part $u^r(t)$ corresponds to the solution of a non-standard evolution problem with a relatively smooth initial value $A^{-m}u^0 \in \dot{H}^{2m}(\Omega)$. Since $(z^\alpha + A)^{-1}$ maps $\dot{H}^{2m}(\Omega)$ to $\dot{H}^{2m+2}(\Omega)$, it follows that the regular part $u^r(t)$ is in $\dot{H}^{2m+2}(\Omega)$ and

$$\|u^r(t)\|_{\dot{H}^{2m+2}(\Omega)} = \|F^r(t)u^0\|_{\dot{H}^{2m+2}(\Omega)} \le ct^{-(m+1)\alpha} \|u^0\|_{L^2(\Omega)} \tag{2.8}$$

In the next two subsections, we present error estimates for the Lagrange interpolation and the Ritz projection of functions in $\dot{H}^{2m+2}(\Omega)$, and then use the established results to prove high-order convergence of the finite element approximation to the regular part $u^r(t)$. The approximation to the singular part will be discussed in Section 4.

2.2. Finite element approximations to functions in $\dot{H}^{2m+2}(\Omega)$

We assume that the polygonal domain Ω is partitioned into a set $\mathcal{K}_h$ of shape-regular and locally quasi-uniform triangles with mesh size $h = \max_{K\in\mathcal{K}_h} \operatorname{diam}(K)$, and denote by X_h the Lagrange finite element space of degree $2m+1$ subject to the partition, i.e.,

$$X_h = \{v_h \in H_0^1(\Omega) : v_h|_K \in \mathbb{P}_{2m+1} \text{ for all } K \in \mathcal{K}_h\},$$

where $\mathbb{P}_{2m+1}$ denotes the space of polynomials of degree $2m+1$.

Let $P_h : L^2(\Omega) \to X_h$, $R_h : \dot{H}^1(\Omega) \to X_h$ and $A_h : X_h \to X_h$ be the L^2-orthogonal projection, the Ritz projection operator, and the discrete elliptic operator, respectively, defined by

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

We shall work with the following assumption on the triangulation of the domain.

Assumption 2.1. *We assume that the triangulation is locally refined towards the corners and edges of the domain, such that both the Lagrange interpolation $I_h : C(\overline{\Omega}) \to X_h$ and the Ritz projection $R_h : H_0^1(\Omega) \to X_h$ have optimal-order convergence, i.e.,*

$$\|v - I_h v\|_{L^2(\Omega)} + h\|v - I_h v\|_{H^1(\Omega)} \le ch^{2r+2}\|v\|_{\dot{H}^{2r+2}(\Omega)} \tag{2.9}$$

$$\|v - R_h v\|_{L^2(\Omega)} + h\|v - R_h v\|_{H^1(\Omega)} \le ch^{2r+2}\|v\|_{\dot{H}^{2r+2}(\Omega)} \tag{2.10}$$

for $v \in \dot{H}^{2r+2}(\Omega)$ and $0 \le r \le m$.

For example, in a two-dimensional polygonal domain Ω, Assumption 2.1 is satisfied by the following type of graded mesh (see Proposition A.1 in Appendix):

$$\hbar(x) \sim \begin{cases} |x - x_0|^{1-\gamma} h \quad \text{with } \gamma \in \left(0, \dfrac{\min(1, \pi/\theta)}{2m+1}\right) & \text{for } h_* \le |x - x_0| \le d_0 \\ h_* & \text{for } |x - x_0| \le h_*, \end{cases} \tag{2.11}$$

where $\hbar(x)$ denotes the spatially dependent diameter of triangles, x_0 is a corner of the polygon with interior angle θ, $h_* \sim h^{1/\gamma}$, and d_0 is a constant such that $D_0' = \{x \in \Omega : |x - x_0| < 2d_0\}$ is a sector centred at the corner x_0.

As a direct consequence of the approximation property (2.9)–(2.10), we have the following estimate in the negative Sobolev spaces.

Lemma 2.1. *Under Assumption 2.1 with $m \ge 1$, the following estimate holds*

$$\|R_h\phi - \phi\|_{\dot{H}^{-r}(\Omega)} \le ch^{r+1}\|R_h\phi - \phi\|_{\dot{H}^1(\Omega)}, \quad \text{with } 1 \le r \le 2m.$$

Proof. For any $\psi \in \dot{H}^r(\Omega)$ we let $w = A^{-1}\psi \in \dot{H}^{r+2}(\Omega)$. By the duality argument, we have

$$\begin{aligned} \|R_h\phi - \phi\|_{\dot{H}^{-r}(\Omega)} &= \sup_{\psi \in \dot{H}^r(\Omega)} \frac{\langle R_h\phi - \phi, \psi\rangle}{\|\psi\|_{\dot{H}^r(\Omega)}} = \sup_{\psi \in \dot{H}^r(\Omega)} \frac{(\nabla(R_h\phi - \phi), \nabla w)}{\|\psi\|_{\dot{H}^r(\Omega)}} \\ &= \sup_{\psi \in \dot{H}^r(\Omega)} \frac{(\nabla(R_h\phi - \phi), \nabla(w - I_h w))}{\|\psi\|_{\dot{H}^r(\Omega)}}. \end{aligned}$$

Then using (2.9) we derive

$$\begin{aligned} |(\nabla(R_h\phi - \phi), \nabla(w - I_h w))| &\le \|R_h\phi - \phi\|_{\dot{H}^1(\Omega)}\|w - I_h w\|_{\dot{H}^1(\Omega)} \\ &\le ch^{r+1}\|R_h\phi - \phi\|_{\dot{H}^1(\Omega)}\|w\|_{\dot{H}^{r+2}(\Omega)} \\ &= ch^{r+1}\|R_h\phi - \phi\|_{\dot{H}^1(\Omega)}\|\psi\|_{\dot{H}^r(\Omega)}. \end{aligned}$$

Then the desired result follows immediately. □

2.3. High-order approximation to the regular part $u^r(t)$

In order to approximate the regular part $u^r(t)$ in (2.5), we consider the following contour integral

$$u_h^r(t) = F_h^r(t)P_h u^0 := \frac{(-1)^m}{2\pi \mathrm{i}} \int_{\Gamma_{\theta,\kappa}} e^{zt} z^{(1+m)\alpha - 1}(z^\alpha + A_h)^{-1} A_h^{-m} P_h u^0 \, \mathrm{d}z. \tag{2.12}$$

We shall establish an error estimate for $u_h^r - u^r$ by using the following technical lemma.

Lemma 2.2. *The following estimate holds for $v \in H_0^1(\Omega)$ and $z \in \Sigma_\theta$ with $\theta \in (\frac{\pi}{2}, \pi)$:*

$$|z^\alpha|\|v\|_{L^2(\Omega)}^2 + \|\nabla v\|_{L^2(\Omega)}^2 \le c\big|z^\alpha\|v\|_{L^2(\Omega)}^2 + (\nabla v, \nabla v)\big|. \tag{2.13}$$

Proof. By [9, Lemma 7.1], we have that for any $z \in \Sigma_\theta$

$$|z|\|v\|_{L^2(\Omega)}^2 + \|\nabla v\|_{L^2(\Omega)}^2 \le c\big|z\|v\|_{L^2(\Omega)}^2 + (\nabla v, \nabla v)\big|.$$

Alternatively, let $\gamma = \|v\|^2_{L^2(\Omega)}$ and $\beta = \|\nabla v\|^2_{L^2(\Omega)} = (\nabla v, \nabla v)$ and $\arg(z) = \varphi$, we have

$$|z\gamma + \beta|^2 \geq (|z|\gamma\cos\varphi + \beta)^2 + (|z|\gamma\sin\varphi)^2.$$

Therefore, we derive

$$|z\gamma + \beta| \geq |z|\gamma\sin\varphi \quad \text{and} \quad |z\gamma+\beta|^2 \geq (\beta\cos\varphi + |z|\gamma)^2 + \beta^2\sin^2\varphi \geq \beta^2\sin^2\varphi.$$

Then for $\varphi \in [\pi-\theta, \theta]$, we have

$$2|z\gamma+\beta| \geq (|z|\gamma+\beta)\sin\varphi \geq (|z|\gamma+\beta)\sin\theta.$$

Meanwhile, for $\varphi \in [0, \pi-\theta]$, we have $\cos\varphi \geq \cos(\pi-\theta) > 0$.

$$|z\gamma+\beta| \geq |z|\gamma\cos\varphi + \beta \geq |z|\gamma\cos(\pi-\theta) + \beta \geq (|z|\gamma+\beta)\cos(\pi-\theta).$$

This completes the proof of the lemma. □

Let $w = (z^\alpha + A)^{-1}A^{-m}v$. Appealing again to Lemma 2.2, we obtain

$$|z^\alpha|\|A^m w\|^2_{L^2(\Omega)} + \|\nabla A^m w\|^2_{L^2(\Omega)} \leq c|((z^\alpha + A)A^m w, A^m w)| \leq c\|v\|_{L^2(\Omega)}\|A^m w\|_{L^2(\Omega)}.$$

Consequently

$$\|A^m w\|_{L^2(\Omega)} \leq c|z^\alpha|^{-1}\|v\|_{L^2(\Omega)} \quad \text{and} \quad \|\nabla A^m w\|_{L^2(\Omega)} \leq c|z^\alpha|^{-\frac{1}{2}}\|v\|_{L^2(\Omega)}. \tag{2.14}$$

In view of (2.14) and the resolvent estimate (2.3), we can bound $\|w\|_{\dot{H}^2(\Omega)}$ by

$$\begin{aligned}\|A^m w\|_{\dot{H}^2(\Omega)} &= \|A^{m+1}w\|_{L^2(\Omega)} = \|(-z^\alpha + z^\alpha + A)(z^\alpha + A)^{-1}v\|_{L^2(\Omega)} \\ &\leq c(\|v\|_{L^2(\Omega)} + |z^\alpha|\|A^m w\|_{L^2(\Omega)}) \leq c\|v\|_{L^2(\Omega)}.\end{aligned} \tag{2.15}$$

Next, we aim to develop an error estimate between $(z^\alpha + A)^{-1}A^{-m}u^0$ and its discrete analogue $(z^\alpha + A_h)^{-1}A_h^{-m}P_h u^0$ for $u^0 \in L^2(\Omega)$. We begin with the following technical lemma.

Lemma 2.3. *Let $u^0 \in L^2(\Omega)$ and we define $\{p_j\}_{j=1}^m$ such that*

$$p_1 = A^{-1}u^0 \quad \text{and} \quad p_j = A^{-1}p_{j-1} \quad \text{with } j = 1, 2, \ldots, m.$$

Moreover, we define $\{p_{j,h}\}_{j=1}^m \subset X_h$ such that

$$p_{1,h} = A_h^{-1}P_h u^0 \quad \text{and} \quad p_{j,h} = A_h^{-1}p_{j-1,h} \quad \text{with } j = 1, 2, \ldots, m.$$

Then there hold error estimates for $j = 1, 2 \ldots, m$

$$\|p_{j,h} - p_j\|_{L^2(\Omega)} + h\|\nabla(p_{j,h} - p_j)\|_{L^2(\Omega)} \leq ch^{2j}\|u^0\|_{L^2(\Omega)} \tag{2.16}$$

and

$$\|p_{j,h} - p_j\|_{H^{-s}(\Omega)} \leq ch^{2j+s}\|u^0\|_{L^2(\Omega)} \quad \text{with } 1 \leq s \leq 2m - 2j + 2. \tag{2.17}$$

Proof. Let $\sigma_j = p_j - p_{j,h}$. By the definition, we have $p_{1,h} = R_h p_1$ and hence derive

$$\|\sigma_1\|_{L^2(\Omega)} + h\|\nabla\sigma_1\|_{L^2(\Omega)} \leq ch^2\|u^0\|_{L^2(\Omega)}.$$

This and the negative norm estimate in Lemma 2.1 lead to

$$\|\sigma_1\|_{H^{-r}(\Omega)} \leq ch^{r+1}\|\sigma_1\|_{H^1(\Omega)} \leq ch^{r+2}\|u^0\|_{L^2(\Omega)} \text{ with } 1 \leq r \leq 2m.$$

Next, we prove (2.16) and (2.17) by mathematical induction. Assume that (2.16) and (2.17) holds for $j = k$. Then Moreover, p_j and $p_{j,h}$ respectively satisfy

$$\begin{aligned}(\nabla p_{k+1}, \nabla\varphi) &= (p_k, \varphi), \quad \forall\varphi \in H_0^1(\Omega), \\ (\nabla p_{k+1,h}, \nabla\varphi_h) &= (p_{k,h}, \varphi_h), \quad \forall\varphi_h \in X_h.\end{aligned}$$

Letting $\varphi = \varphi_h$ and subtracting these two identities yield the following error equation

$$(\nabla\sigma_{k+1}, \nabla\varphi_h) = (\sigma_k, \varphi_h), \quad \forall\varphi_h \in X_h. \tag{2.18}$$

Therefore, we have the following estimate

$$\begin{aligned}\|\nabla\sigma_{k+1}\|_{L^2(\Omega)}^2 &= (\nabla\sigma_{k+1},\nabla(p_{k+1}-R_hp_{k+1}))+(\sigma_k,R_hp_{k+1}-p_{k+1})+(\sigma_k,\sigma_{k+1})\\ &\le \|\nabla\sigma_{k+1}\|_{L^2(\Omega)}\|\nabla(p_{k+1}-R_hp_{k+1})\|_{L^2(\Omega)}\\ &\quad+\|\nabla\sigma_k\|_{L^2(\Omega)}\|p_{k+1}-R_hp_{k+1}\|_{H^{-1}(\Omega)}+\|\sigma_k\|_{H^{-1}(\Omega)}\|\nabla\sigma_{k+1}\|_{L^2(\Omega)}.\end{aligned}$$

According to (2.16), (2.17) with $j=k$ and $s=-1$, (2.10) and Lemma 2.1, we derive

$$\begin{aligned}&\|\nabla\sigma_{k+1}\|_{L^2(\Omega)}^2\\ &\le c\|\nabla(I-R_h)p_{k+1}\|_{L^2(\Omega)}^2+\|\nabla\sigma_k\|_{L^2(\Omega)}\|(I-R_h)p_{k+1}\|_{H^{-1}(\Omega)}+c\|\sigma_k\|_{H^{-1}(\Omega)}^2\\ &\le ch^{4k+2}\|u^0\|_{L^2(\Omega)}^2\end{aligned}\tag{2.19}$$

Next, we show the error estimate in $H^{-r}(\Omega)$ with $0\le r\le 2m-2k$ by duality argument. For any $\varphi\in\dot H^r(\Omega)$ we let $\phi=A^{-1}\varphi$. Then we observe

$$\begin{aligned}\|\sigma_{k+1}\|_{H^{-r}(\Omega)} &= \sup_{\varphi\in\dot H^r(\Omega)}\frac{\langle\sigma_{k+1},\varphi\rangle}{\|\varphi\|_{\dot H^r(\Omega)}}=\sup_{\varphi\in\dot H^r(\Omega)}\frac{(\nabla\sigma_{k+1},\nabla\phi)}{\|\varphi\|_{\dot H^r(\Omega)}}\\ &= \sup_{\varphi\in\dot H^r(\Omega)}\frac{(\nabla\sigma_{k+1},\nabla(\phi-R_h\phi))+(\sigma_k,R_h\phi-\phi)+(\sigma_k,\phi)}{\|\varphi\|_{\dot H^r(\Omega)}}\end{aligned}$$

Using (2.19) and (2.10), we derive

$$\begin{aligned}\sup_{\varphi\in\dot H^r(\Omega)}\frac{(\nabla\sigma_{k+1},\nabla(\phi-R_h\phi))}{\|\varphi\|_{\dot H^r(\Omega)}} &\le \sup_{\varphi\in\dot H^r(\Omega)}\frac{\|\nabla\sigma_{k+1}\|_{L^2(\Omega)}\|\nabla(I-R_h)\phi\|_{L^2(\Omega)}}{\|\varphi\|_{\dot H^r(\Omega)}}\\ &\le \sup_{\varphi\in\dot H^r(\Omega)}\frac{ch^{2k+1}\|u^0\|_{L^2(\Omega)}ch^{r+1}\|\phi\|_{\dot H^{r+2}(\Omega)}}{\|\varphi\|_{\dot H^r(\Omega)}}\le ch^{2k+2+r}.\end{aligned}$$

Meanwhile, by duality between $\dot H^{-1}(\Omega)$ and $\dot H^1(\Omega)$, we apply (2.10) and (2.16) with $j=k$ and $s=1$ to derive

$$\begin{aligned}\sup_{\varphi\in\dot H^r(\Omega)}\frac{(\sigma_k,R_h\phi-\phi)}{\|\varphi\|_{\dot H^r(\Omega)}} &\le \sup_{\varphi\in\dot H^r(\Omega)}\frac{\|\sigma_k\|_{\dot H^{-1}(\Omega)}\|(I-R_h)\phi\|_{\dot H^1(\Omega)}}{\|\varphi\|_{\dot H^r(\Omega)}}\\ &\le \sup_{\varphi\in\dot H^r(\Omega)}\frac{ch^{2k+1}\|u^0\|_{L^2(\Omega)}ch^{r+1}\|\phi\|_{\dot H^{r+2}(\Omega)}}{\|\varphi\|_{\dot H^r(\Omega)}}\le ch^{2k+2+r}.\end{aligned}$$

Similarly, by means of the duality between $\dot H^{-2-r}(\Omega)$ and $\dot H^{2+r}(\Omega)$, we apply again (2.16) with $j=k$ and $s=2+r$ to derive

$$\begin{aligned}\sup_{\varphi\in\dot H^r(\Omega)}\frac{(\sigma_k,\phi)}{\|\varphi\|_{\dot H^r(\Omega)}} &\le \sup_{\varphi\in\dot H^r(\Omega)}\frac{\|\sigma_k\|_{\dot H^{-r-2}(\Omega)}\|\phi\|_{\dot H^{2+r}(\Omega)}}{\|\varphi\|_{\dot H^r(\Omega)}}\\ &\le \sup_{\varphi\in\dot H^r(\Omega)}\frac{ch^{2k+r+2}\|u^0\|_{L^2(\Omega)}\|\varphi\|_{\dot H^r(\Omega)}}{\|\varphi\|_{\dot H^r(\Omega)}}\le ch^{2k+2+r}.\end{aligned}$$

This completes the proof of the lemma. □

Lemma 2.4. *Let $u^0\in L^2(\Omega)$, $z\in\Sigma_\theta$, $w=(z^\alpha+A)^{-1}p$ with $p=A^{-m}u^0$, and $w_h=(z^\alpha+A_h)^{-1}p_h$ with $p_h=A_h^{-m}P_hu^0$. Then there holds*

$$\|w_h-w\|_{L^2(\Omega)}+h\|\nabla(w_h-w)\|_{L^2(\Omega)}\le ch^{2m+2}\|u^0\|_{L^2(\Omega)}.\tag{2.20}$$

Proof. Let $e=w-w_h$ and $\sigma=p-p_h$. Then Lemma 2.3 implies the estimate

$$\|\sigma\|_{L^2(\Omega)}+h\|\nabla\sigma\|_{L^2(\Omega)}\le ch^{2m}\|u^0\|_{L^2(\Omega)}.\tag{2.21}$$

and the negative norm error estimate

$$\|\sigma\|_{H^{-r}(\Omega)} \le ch^{r+2m}\|u^0\|_{L^2(\Omega)}, \text{ with } 1 \le r \le 2. \tag{2.22}$$

Moreover, w and w_h respectively satisfy

$$z^\alpha(w,\varphi) + (\nabla w, \nabla\varphi) = (p,\varphi), \quad \forall \varphi \in H_0^1(\Omega),$$
$$z^\alpha(w_h,\varphi_h) + (\nabla w_h, \nabla\varphi_h) = (p_h,\varphi_h), \quad \forall \varphi_h \in X_h.$$

Subtracting these two identities yields the following error equation of e

$$z^\alpha(e,\varphi_h) + (\nabla e, \nabla\varphi_h) = (\sigma,\varphi_h), \quad \forall \varphi_h \in X_h. \tag{2.23}$$

This and Lemma 2.2 imply that

$$\begin{aligned}|z^\alpha|\|e\|_{L^2(\Omega)}^2 + \|\nabla e\|_{L^2(\Omega)}^2 &\le c\big|z^\alpha\|e\|_{L^2(\Omega)}^2 + (\nabla e, \nabla e)\big| \\ &= c\,|z^\alpha(e, w - R_h w) + (\nabla e, \nabla(w - R_h w)) - (\sigma, w_h - R_h w)|\end{aligned}$$

By using the Cauchy-Schwartz inequality and the duality between $\dot H^1(\Omega)$ and $\dot H^{-1}(\Omega)$, we arrive at

$$|z^\alpha|\|e\|_{L^2(\Omega)}^2 + \|\nabla e\|_{L^2(\Omega)}^2 \le c\big(|z^\alpha|\|w - R_h w\|_{L^2(\Omega)}^2 + \|\nabla(w - R_h w)\|_{L^2(\Omega)}^2 + \|\sigma\|_{H^{-1}(\Omega)}^2\big). \tag{2.24}$$

According to (2.14), (2.15) and (2.22), we derive

$$\begin{aligned}&|z^\alpha|\|e\|_{L^2(\Omega)}^2 + \|\nabla e\|_{L^2(\Omega)}^2 \\ &\le ch^{4m+2}\big(|z^\alpha|\|(z^\alpha + A)^{-1}u^0\|_{\dot H^1(\Omega)}^2 + \|(z^\alpha + A)^{-1}u^0\|_{\dot H^2(\Omega)}^2 + \|u^0\|_{L^2(\Omega)}^2\big) \\ &\le ch^{4m+2}\|u^0\|_{L^2(\Omega)}^2.\end{aligned} \tag{2.25}$$

This gives the desired bound on $\|\nabla e\|_{L^2(\Omega)}$. Next, we bound $\|e\|_{L^2(\Omega)}$ using a duality argument. For any fixed $\varphi \in L^2(\Omega)$, we set $\psi = (z^\alpha + A)^{-1}\varphi$. Then the preceding argument implies

$$|z^\alpha|\|\psi - R_h\psi\|_{L^2(\Omega)}^2 + \|\nabla(\psi - R_h\psi)\|_{L^2(\Omega)}^2 \le ch^2\|\varphi\|_{L^2(\Omega)}^2. \tag{2.26}$$

we have by duality

$$\|e\|_{L^2(\Omega)} = \sup_{\varphi\in L^2(\Omega)} \frac{|(e,\varphi)|}{\|\varphi\|_{L^2(\Omega)}} = \sup_{\varphi\in L^2(\Omega)} \frac{|z^\alpha(e,\psi) + (\nabla e, \nabla\psi)|}{\|\varphi\|_{L^2(\Omega)}}.$$

Then the desired estimate follows from (2.22), (2.23), (2.25) and (2.26) by

$$\begin{aligned}|z^\alpha(e,\psi) + (\nabla e, \nabla\psi)| &= |z^\alpha(e, \psi - R_h\psi) + (\nabla e, \nabla(\psi - R_h\psi)) + (\sigma, R_h\psi)| \\ &\le |z^\alpha(e, \psi - R_h\psi) + (\nabla e, \nabla(\psi - R_h\psi))| + |(\sigma, R_h\psi - \psi)| + |(\sigma,\psi)| \\ &\le |z^\alpha|^{\frac12}\|e\|_{L^2(\Omega)}|z^\alpha|^{\frac12}\|\psi - R_h\psi\|_{L^2(\Omega)} + \|\nabla e\|_{L^2(\Omega)}\|\nabla(\psi - R_h\psi)\|_{L^2(\Omega)} \\ &\quad + \|\sigma\|_{H^{-1}(\Omega)}\|\nabla(R_h\psi - \psi)\|_{L^2(\Omega)} + \|\sigma\|_{H^{-2}(\Omega)}\|\psi\|_{\dot H^2(\Omega)} \\ &\le ch^{2m+2}\|u^0\|_{L^2(\Omega)}\|\psi\|_{\dot H^2(\Omega)} \le ch^{2m+2}\|v\|_{L^2(\Omega)}\|\varphi\|_{L^2(\Omega)}.\end{aligned}$$

This completes proof of the lemma. □

Now we can state error estimates for the regular part.

Theorem 2.1. *Let u^r and u_h^r be the functions defined by (2.6) and (2.12), respectively. Then for $t > 0$, there holds:*

$$\|(u^r - u_h^r)(t)\|_{L^2(\Omega)} + h\|\nabla(u^r - u_h^r)(t)\|_{L^2(\Omega)} \le ch^{2m+2}t^{-(1+m)\alpha}\|u^0\|_{L^2(\Omega)}.$$

Proof. For $v \in L^2(\Omega)$, by the solution representations, the error $e_h(t)$ can be represented as

$$|(u^r - u_h^r)(t)| = \Big|\frac{1}{2\pi\mathrm{i}}\int_{\Gamma_{\theta,\kappa}} e^{zt}z^{(1+m)\alpha-1}(w_h(z) - w(z))\,\mathrm{d}z\Big|,$$

with $w(z) = (z^\alpha + A)^{-1}A^{-m}u^0$ and $w_h(z) = (z^\alpha + A_h)^{-1}A_h^{-m}P_hu^0$. By Lemma 2.4, and taking $\kappa = t^{-1}$ in the contour $\Gamma_{\theta,\kappa}$, we have

$$\|(u^r - u_h^r)(t)\|_{L^2(\Omega)} \le ch^{2m+2}\|u^0\|_{L^2(\Omega)} \int_{\Gamma_{\theta,\kappa}} e^{\Re(z)t}|z|^{(1+m)\alpha-1}\,|\mathrm{d}z| \le ch^{2m+2}t^{-(1+m)\alpha}\|u^0\|_{L^2(\Omega)}.$$

A similar argument also yields the $H^1(\Omega)$-estimate. □

Remark 2.1. *A slightly modification leads to the error estimate for very weaker initial data $u^0 \in \dot{H}^s(\Omega)$ with some $s \in [-1, 0]$. In particular let u^r and u_h^r be the functions defined by (2.6) and (2.12), respectively. Then for $t > 0$, there holds*

$$\|(u^r - u_h^r)(t)\|_{L^2(\Omega)} + h\|\nabla(u^r - u_h^r)(t)\|_{L^2(\Omega)} \le ch^{2m+2+s}t^{-(1+m)\alpha}\|u^0\|_{\dot{H}^s(\Omega)}. \tag{2.27}$$

Remark 2.2. *The argument could be further extended to rougher initial data, such as the Dirac delta function $u^0 = \delta_{x_*}$ in two dimension with a fixed $x_* \in \Omega$. Then we consider the splitting*

$$\begin{aligned} u^r(t) - u_h^r(t) &= F^r(t)u^0 - F_h^r(t)P_hu^0 \\ &= (F^r(t)u^0 - F^r(t)P_hu^0) + (F^r(t)P_hu^0 - F_h^r(t)P_hu^0). \end{aligned} \tag{2.28}$$

The first term could be bounded using the smoothing property (2.8) and the L^∞-stability of the L^2 projection (see [6])

$$\begin{aligned} \|F^r(t)u^0 - F^r(t)P_hu^0\|_{L^2(\Omega)} &= \sup_{\phi\in L^2(\Omega)} \frac{(F^r(t)u^0 - F^r(t)P_hu^0, \phi)}{\|\phi\|_{L^2(\Omega)}} \\ &\le \sup_{\phi\in L^2(\Omega)} \frac{\|(I-P_h)F^r(t)\phi\|_{L^\infty(\Omega)}}{\|\phi\|_{L^2(\Omega)}} \le C \sup_{\phi\in L^2(\Omega)} \frac{\|(I-I_h)F^r(t)\phi\|_{L^\infty(\Omega)}}{\|\phi\|_{L^2(\Omega)}} \\ &\le ch^{2m+1} \sup_{\phi\in L^2(\Omega)} \frac{\|F^r(t)\phi\|_{\dot{H}^{2m+2}(\Omega)}}{\|\phi\|_{L^2(\Omega)}} \le ch^{2m+1}t^{-(1+m)\alpha}, \end{aligned}$$

where we have used the L^∞ error estimate for the Lagrange interpolation (see Lemma A.2) in the second to last inequality.

Meanwhile, the second term in (2.28) could be bounded using the estimate (2.27) and the inverse inequality, i.e.,

$$\begin{aligned} \|F^r(t)P_hu^0 - F_h^r(t)P_hu^0\|_{L^2(\Omega)} &\le ch^{2m+2}t^{-(1+m)\alpha}\|P_hu^0\|_{L^2(\Omega)} \\ &= ch^{2m+2}t^{-(1+m)\alpha} \sup_{\phi\in L^2(\Omega)} \frac{|P_h\phi(x_*)|}{\|\phi\|_{L^2(\Omega)}} \le ch^{2m+1}t^{-(1+m)\alpha}. \end{aligned}$$

As a result, we have the following error estimate for Dirac delta initial condition in two dimension

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

This convergence rate is consistent with our numerical experiments, cf. Table 1–3.

3. Time discretization

In the preceding section, we have proposed a splitting of the exact solution into a time-dependent regular part plus several time-independent singular parts. Next, we shall develop and analyze a time stepping scheme for approximating the regular part using convolution quadrature.

We shall focus on time-stepping schemes with a uniform temporal mesh. Specifically, let $t_n = n\tau$, $n = 0, 1, \ldots, N$, be a uniform partition of the time interval $[0, T]$ with a time stepsize $\tau = N^{-1}T$, $N \in \mathbb{N}$, and recall that the generating function of BDF method of order k, $k = 1, \ldots, 6$, is defined

by

$$\delta_\tau(\zeta) := \frac{\delta(\zeta)}{\tau} \quad \text{with } \delta(\zeta) = \sum_{j=1}^{k} \frac{1}{j}(1-\zeta)^j. \tag{3.1}$$

The BDFk method is known to be $A(\vartheta_k)$-stable with angle $\vartheta_k = 90°$, $90°$, $86.03°$, $73.35°$, $51.84°$, $17.84°$ for $k = 1, 2, 3, 4, 5, 6$, respectively [11, pp. 251].

Then we apply the following convolution quadrature to approximate the semidiscrete solution (2.12):

$$U_h^{n,r} = \frac{(-1)^m}{2\pi \mathrm{i}} \int_{\Gamma_{\theta,\kappa}^\tau} e^{zt_n} \delta_\tau(e^{-z\tau})^{(1+m)\alpha-1} (\delta_\tau(e^{-z\tau})^\alpha + A_h)^{-1} A_h^{-m} P_h u^0 \,\mathrm{d}z. \tag{3.2}$$

where the the contour $\Gamma_{\theta,\kappa}^\tau$ is $\Gamma_{\theta,\kappa}^\tau := \{z \in \Gamma_{\theta,\kappa} : |\Im(z)| \le \frac{\pi}{\tau}\}$ oriented with an increasing imaginary part. The evaluation of the contour integral in (3.2) is equivalent to solving the following time-stepping scheme for $U_h^{n,r}$:

$$\tau^{-\alpha} \sum_{j=0}^{n} \omega_j^{(\alpha)} U_h^{n-j,r} + A_h U_h^{n,r} = (-1)^m \tau^{-(1+m)\alpha} \omega_n^{(1+m)\alpha-1} A_h^{-m} P_h u^0, \quad \text{for } 0 \le n \le N. \tag{3.3}$$

Here the quadrature weights $\big(\omega_j^{(\beta)}\big)_{j=0}^\infty$ are given by the coefficients in the following power series expansion

$$\delta_\tau(\zeta)^\beta = \frac{1}{\tau^\beta} \sum_{j=0}^{\infty} \omega_j^{(\beta)} \zeta^j. \tag{3.4}$$

with the generating function (3.1). Generally, those weights can be evaluated efficiently via recursion or discrete Fourier transform [40, 43].

Note that the time stepping scheme (3.3) begins with $n = 0$, which is different from the usual time stepping schemes for evolution problems. The idea is closely related to correct the initial steps of the regular time stepping scheme [34, 7, 17, 47]. See a brief explanation in [17, Appendix A].

The next Lemma shows the equivalence between the convolution quadrature (3.2) and the time stepping scheme (3.3).

Lemma 3.1. *The function $U_h^{n,r}$ given by the contour integral (3.2) is the solution of the time stepping scheme (3.3) for all $0 \le n \le N$.*

Proof. We begin with the time stepping scheme (3.3). By multiplying both sides of the relation (3.3) by ζ^n, summing over n from 0 to ∞ and collecting terms, we obtain

$$\begin{aligned}\sum_{n=0}^{\infty} \zeta^n \Big(\tau^{-\alpha} \sum_{j=0}^{n} \omega_j^{(\alpha)} U_h^{n-j,r}\Big) + A_h \sum_{n=0}^{\infty} U_h^{n,r} \zeta^n &= (-1)^m \tau^{-1} A_h^{-m} P_h u^0 \Big(\tau^{1-(1+m)\alpha} \sum_{n=0}^{\infty} \omega_n^{(1+m)\alpha-1} \zeta^n\Big) \\ &= (-1)^m \tau^{-1} A_h^{-m} P_h u^0 \delta_\tau(\zeta)^{(1+m)\alpha-1}.\end{aligned}$$

For any sequence $(v^n)_{n=1}^\infty$, we denotes its generating function by $\widetilde{v}(\xi) = \sum_{n=0}^\infty v^n \zeta^n$. The the leading term in the above relation can be written as

$$\begin{aligned}\sum_{n=0}^{\infty} \zeta^n \Big(\tau^{-\alpha} \sum_{j=0}^{n} \omega_j^{(\alpha)} U_h^{n-j,r}\Big) &= \tau^{-\alpha} \sum_{j=0}^{\infty} \omega_j^{(\alpha)} \zeta^j \Big(\sum_{n=j}^{\infty} U_h^{n-j,r} \zeta^{n-j}\Big) \\ &= \delta_\tau(\zeta)^\alpha \widetilde{U}_h^r(\zeta).\end{aligned}$$

Therefore we obtain

$$\widetilde{U}_h^r(\zeta) = (-1)^m \tau^{-1} \delta_\tau(\zeta)^{(1+m)\alpha-1} (\delta_\tau(\zeta)^\alpha + A_h)^{-1} A_h^{-m} P_h u^0.$$

Since $\widetilde{U}_h^r(\xi)$ is analytic with respect to ζ in the unit disk on the complex plane $\mathbb{C}$, thus Cauchy's integral formula and the change of variables $\zeta = e^{-z\tau}$ lead to the following representation for arbitrary $\varrho \in (0,1)$

$$\begin{aligned} U_h^{n,r} &= \frac{1}{2\pi \mathrm{i}} \int_{|\zeta|=\varrho} \zeta^{-n-1} \widetilde{U}_h^r(\zeta)\,\mathrm{d}\zeta = \frac{\tau}{2\pi \mathrm{i}} \int_{\Gamma^\tau} e^{zt_n} \widetilde{U}_h^r(e^{-z\tau})\,\mathrm{d}z \\ &= \frac{(-1)^m}{2\pi \mathrm{i}} \int_{\Gamma^\tau} e^{zt_n} \delta_\tau(e^{-z\tau})^{(1+m)\alpha-1} (\delta_\tau(e^{-z\tau})^\alpha + A_h)^{-1} A_h^{-m} P_h u^0\,\mathrm{d}z \end{aligned} \tag{3.5}$$

where Γ^τ is given by $\Gamma^\tau := \{z = -\frac{\log \varrho}{\tau} + \mathrm{i}y : y \in \mathbb{R} \text{ and } |y| \le \frac{\pi}{\tau}\}$.

Note that $\delta_\tau(e^{-z\tau})^{(1+m)\alpha-1}(\delta_\tau(e^{-z\tau})^\alpha + A_h)^{-1}$ is analytic for $z \in \Sigma_{\theta,\kappa}^\tau$, which is a region enclosed by Γ^τ, $\Gamma_{\theta,\kappa}^\tau$ and the two lines $\Gamma_\pm^\tau := \mathbb{R} \pm \mathrm{i}\frac{\pi}{\tau}$ (oriented from left to right). Using the periodicity of $e^{-z\tau}$ and Cauchy's theorem, we deform the contour Γ^τ to $\Gamma_{\theta,\kappa}^\tau$ in the integral (3.5) to obtain the desired representation (3.2). □

Finally, we study the error of convolution approximation. To this end, we need the following lemma on the sectorial property and approximation property of the generating function $\delta_\tau(\zeta)$. See the detailed proof in [17, Lemma B.1].

Lemma 3.2. *For any ε, there exists $\theta_\varepsilon \in (\frac{\pi}{2}, \pi)$ such that for any $\theta \in (\frac{\pi}{2}, \theta_\varepsilon)$, there exist positive constants c, c_1, c_2 (independent of τ) such that*

$$\begin{aligned} &c_1|z| \le |\delta_\tau(e^{-z\tau})| \le c_2|z|, \quad \delta_\tau(e^{-z\tau}) \in \Sigma_{\pi-\vartheta_k+\varepsilon}, \\ &|\delta_\tau(e^{-z\tau}) - z| \le c\tau^k |z|^{k+1}, \quad |\delta_\tau(e^{-z\tau})^\alpha - z^\alpha| \le c\tau^k|z|^{k+\alpha}, \quad \forall\, z \in \Gamma_{\theta,\kappa}^\tau, \end{aligned}$$

where $\sigma > 0$ and the contour $\Gamma_{\theta,\kappa}^\tau \subset \mathbb{C}$ is defined by

$$\Gamma_{\theta,\kappa}^\tau := \{z = \rho e^{\pm \mathrm{i}\theta} : \rho \ge \kappa, |\Im(z)| \le \tfrac{\pi}{\tau}\} \cup \left\{z = \kappa e^{\mathrm{i}\varphi} : |\varphi| \le \theta\right\}.$$

Theorem 3.1. *Let $U_h^{n,r}$ be the function defined by the convolution quadrature (3.2), and u_h^r be the function defined by the contour integral (2.12). Then we have*

$$\|U_h^{n,r} - u_h^r(t_n)\|_{L^2(\Omega)} \le c\tau^k t_n^{-k-m\alpha} \|u^0\|_{\dot{H}^s(\Omega)} \quad \text{for any } s \in [-1, 0].$$

Proof. Let $K(z) = z^{(1+m)\alpha-1}(z^\alpha + A_h)^{-1}$. Then we may split the error as

$$\begin{aligned} U_h^{n,r} - u_h^r(t_n) &= \frac{(-1)^m}{2\pi \mathrm{i}} \int_{\Gamma_{\theta,\kappa}^\tau} e^{zt_n} \Big(K(z) - K(\delta_\tau(e^{-z\tau}))\Big) A_h^{-m} P_h u^0\,\mathrm{d}z \\ &\quad + \frac{(-1)^m}{2\pi \mathrm{i}} \int_{\Gamma_{\theta,\kappa} \setminus \Gamma_{\theta,\kappa}^\tau} e^{zt_n} K(z) A_h^{-m} P_h u^0\,\mathrm{d}z =: I_1 + I_2. \end{aligned}$$

Using the resolvent estimate (2.3) and Lemma 3.2, we derive

$$\|K(z) - K(\delta_\tau(e^{-z\tau}))\|_{L^2(\Omega)\to L^2(\Omega)} \le c\tau^k |z|^{m\alpha+k-1}.$$

As a result, we choose $\kappa = t_n^{-1}$ in the contour $\Gamma_{\theta,\kappa}^\tau$, we obtain an estimate for I_1:

$$\begin{aligned} \|I_1\|_{L^2(\Omega)} &\le c\tau^k \|A_h^{-m} P_h u^0\|_{L^2(\Omega)} \Big(\int_\kappa^\infty e^{\kappa t_n \cos\theta} \rho^{m\alpha+k-1}\,\mathrm{d}\rho + \int_{-\theta}^{\theta} e^{\kappa t_n \cos\varphi} \kappa^{m\alpha+k}\,\mathrm{d}\varphi \Big) \\ &\le c\tau^k (t_n^{-m\alpha-k} + \kappa^{m\alpha+k}) \|A_h^{-m} P_h u^0\|_{L^2(\Omega)} \le c\tau^k t_n^{-k-m\alpha} \|A_h^{-m} P_h u^0\|_{L^2(\Omega)} \\ &\le c\tau^k t_n^{-k-m\alpha} \|u^0\|_{\dot{H}^s(\Omega)}, \end{aligned}$$

for any $s \in [-1, 0]$. The last inequality follows from the stability of P_h in $\dot{H}^s(\Omega)$ for $s \in [-1, 0]$:

$$\|A_h^{-m} P_h u^0\|_{L^2(\Omega)} \le c\|A_h^{-1} P_h u^0\|_{\dot{H}^1(\Omega)} \le c\|P_h u^0\|_{\dot{H}^{-1}(\Omega)} \le c\|u^0\|_{\dot{H}^{-1}(\Omega)}.$$

Meanwhile, for the term I_2, we apply the resolvent estimate (2.3) and Lemma 3.2 to derive

$$\begin{aligned}\|I_1\|_{L^2(\Omega)} &\le c\int_{\frac{\pi}{\tau\sin\theta}}^{\infty} e^{-c\rho t_n}\rho^{m\alpha-1}\|A_h^{-m}P_hu^0\|_{L^2(\Omega)}\,\mathrm{d}\rho\\ &\le c\tau^k\int_{\frac{\pi}{\tau\sin\theta}}^{\infty} e^{-c\rho t_n}\rho^{m\alpha+k-1}\|A_h^{-m}P_hu^0\|_{L^2(\Omega)}\,\mathrm{d}\rho\\ &\le c\tau^k t_n^{-m\alpha-k}\|A_h^{-m}P_hu^0\|_{L^2(\Omega)} \le c\tau^k t_n^{-m\alpha-k}\|u^0\|_{\dot H^{-1}(\Omega)}.\end{aligned}$$

This completes the proof of the theorem. □

In view of Remarks 2.1–2.2 and Theorem 3.1, we have the following error estimate for the fully discrete solution.

Corollary 3.1. *Assume that $u^0\in\dot H^s(\Omega)$ with some $s\in[-1,0]$, and u is the solution to (1.1) with $f=0$. Let $U_h^{n,r}$ be the function defined by the convolution quadrature (3.2). Suppose that $\phi_{j,h}\in X_h$ is an approximation to $A^{-j}v$. Then the fully discrete solution*

$$U_h^n=\sum_{j=1}^{m}(-1)^{j+1}\frac{t^{-j\alpha}}{\Gamma(1-j\alpha)}\phi_{j,h}+U_h^{n,r} \tag{3.6}$$

satisfies the following error estimate

$$\|U_h^n-u(t_n)\|_{L^2(\Omega)}\le c\Big(h^{2m+2+s}t_n^{-(1+m)\alpha}+\tau^k t_n^{-k-m\alpha}\Big)\|v\|_{\dot H^s(\Omega)}+c\sum_{j=1}^{m}t_n^{-j\alpha}\|\phi_{j,h}-A^{-j}v\|_{L^2(\Omega)}.$$

4. Discussion on the approximation to the singular part $\sum_{j=1}^m u_j^s(t)$

The singular part $\sum_{j=1}^m u_j^s(t)$ of the solution in (2.6), with

$$u_j^s(t)=(-1)^{j+1}\frac{t^{-j\alpha}}{\Gamma(1-j\alpha)}A^{-j}u^0, \tag{4.1}$$

should be approximated separately. Since $u_j^s(t)$ can be computed by solving several elliptic equations, its computational cost is much smaller than the computation of the regular part (which requires solving an evolution problem; see the full discretization in the next section). Therefore, in general, the singular part can be solved by a much smaller mesh size without significantly increasing the overall computational cost.

In the following, we discuss several cases in which the singular part can be solved with high-order accuracy without using smaller meshes.

Example 4.1 (Piecewise smooth initial data)**.** If the initial value $u^0\in L^2(\Omega)$ is globally discontinuous (therefore nonsmooth) but piecewise smooth, separated by a smooth closed interface $\Gamma\subset\Omega$, then one can approximate $q_j=A^{-j}u^0$ by $q_{j,h}=A_h^{-j}P_hu^0$ using isoparametric finite element method with triangulations fitting the interface. The computation of $q_{j,h}=A_h^{-j}P_hu^0$ is equivalent to solving the following several standard elliptic equations:

$$A_hq_{k,h}=q_{k-1,h},\quad k=1,\dots,j,\quad\text{with}\quad q_{0,h}=P_hu^0.$$

By denoting $\Omega=\Omega_1\cup\Omega_2\cup\Gamma$, where Ω_1 and Ω_2 are separated by a smooth interface Γ, the following high-order convergence can be achieved for isoparametric finite elements of degree $2m+1$ fitting the interface Γ:

$$\|q_{j,h}-q_j\|_{L^2}\le Ch^{2m+2}\|u^0\|_{H^{2m+2}_{\text{piecewise}}(\Omega)}, \tag{4.2}$$

where $\dot H^{2m+2}_{\text{piecewise}}(\Omega)=\big\{g\in L^2(\Omega):g|_{\Omega_j}\in\dot H^{2m+2}(\Omega_j)\text{ for }j=1,2\big\}$. This shows that the singular part in (4.1) can be approximated with high-order accuracy for piecewise smooth initial data.

The error estimate in (4.2) can be proved by using the following result (for isoparametric finite elements of degree $2m+1$ fitting the interface):

$$\|A_h^{-1}f - A^{-1}f\|_{L^2} \le Ch^{2m+2}\|f\|_{\dot{H}^{2m+2}_{\text{piecewise}}(\Omega)}, \tag{4.3}$$

which was originally proved in [33] for a bounded smooth domain Ω which contains the interface Γ. If Ω is a polygon which contains the interface Γ, then the interface is away from the corners of the polygon (therefore the functions in $\dot{H}^{2r+2}_{\text{piecewise}}(\Omega)$ are locally in the classical Sobolev space H^{2r+2} near the interface), it follows that the error estimates in [33] near the interface still hold for functions in $\dot{H}^{2r+2}_{\text{piecewise}}(\Omega)$. Therefore, the combination of Proposition A.1 and the error estimates in [33] yields (4.3) for the triangulations satisfying (2.11). Since

$$\|q_k\|_{\dot{H}^{2m+2}_{\text{piecewise}}(\Omega)} = \|A^{-1}q_{k-1}\|_{\dot{H}^{2m+2}_{\text{piecewise}}(\Omega)} \le C\|q_{k-1}\|_{\dot{H}^{2m}_{\text{piecewise}}(\Omega)},$$

iterating this inequality yields that $\|q_k\|_{\dot{H}^{2m+2}_{\text{piecewise}}(\Omega)} \le C\|u^0\|_{H^{2m+2}_{\text{piecewise}}(\Omega)}$. By using this regularity and (4.3), we have

$$\begin{aligned}
\|q_{k,h} - q_k\|_{L^2} &= \|A_h^{-1}q_{k-1,h} - A^{-1}q_{k-1}\|_{L^2} \\
&\le \|A_h^{-1}(q_{k-1,h} - P_hq_{k-1})\|_{L^2} + \|A_h^{-1}P_hq_{k-1} - A^{-1}q_{k-1}\|_{L^2} \\
&\le \|q_{k-1,h} - P_hq_{k-1}\|_{L^2} + \|A_h^{-1}P_hq_{k-1} - A^{-1}q_{k-1}\|_{L^2} \\
&\le \|q_{k-1,h} - q_{k-1}\|_{L^2} + \|q_{k-1} - P_hq_{k-1}\|_{L^2} + \|A_h^{-1}P_hq_{k-1} - A^{-1}q_{k-1}\|_{L^2} \\
&\le \|q_{k-1,h} - q_{k-1}\|_{L^2} + Ch^{2m+2}\|q_{k-1}\|_{\dot{H}^{2m+2}_{\text{piecewise}}(\Omega)} + Ch^{2m+2}\|q_{k-1}\|_{\dot{H}^{2m+2}_{\text{piecewise}}(\Omega)} \\
&\le \|q_{k-1,h} - q_{k-1}\|_{L^2} + Ch^{2m+2}\|u^0\|_{H^{2m+2}_{\text{piecewise}}(\Omega)}.
\end{aligned}$$

By iterating this inequality for $k = 1, \dots, j$, and using the following basic result:

$$\|q_{0,h} - q_0\|_{L^2} = \|P_hu^0 - u^0\|_{L^2} \le Ch^{2m+2}\|u^0\|_{H^{2m+2}_{\text{piecewise}}(\Omega)},$$

we obtain the high-order convergence in (4.2).

Example 4.2 (Dirac–Delta point source)**.** If the initial value is a Dirac–Delta point source centered at some interior point $x_0 \in \Omega$, i.e., $u^0 = \delta_{x_0}$, then the function

$$w_1 = A^{-1}u^0 - \hat{q}_1, \quad \text{with} \quad \hat{q}_1(x) = \frac{1}{2\pi}\ln|x - x_0|,$$

is the solution of the following boundary value problem:

$$\begin{cases} -\Delta w_1 = 0 & \text{in } \Omega, \\ w_1 = -\hat{q}_1 & \text{for } x \in \partial\Omega, \end{cases} \tag{4.4}$$

Let χ be a smooth cut-off function such that $\chi = 1$ in a neighborhood of the boundary $\partial\Omega$ and $\chi = 0$ in a neighborhood of x_0. Then $\chi\hat{q}_1 \in C^\infty$ and

$$\begin{cases} -\Delta(w_1 - \chi\hat{q}_1) = \Delta(\chi\hat{q}_1) \in C_0^\infty(\Omega) \subset \dot{H}^{2m}(\Omega) & \text{in } \Omega, \\ w_1 - \chi\hat{q}_1 = 0 & \text{on } \partial\Omega. \end{cases}$$

This implies that $w_1 - \chi\hat{q}_1 \in A^{-1}\dot{H}^{2m}(\Omega) = \dot{H}^{2m+2}(\Omega)$. Since the explicit expression of $\Delta(\chi\hat{q}_1)$ is known, we can approximate $w_1 - \chi\hat{q}_1$ by the finite element function $A_h^{-1}\Delta(\chi\hat{q}_1)$ and, correspondingly, approximate $q_1 = A^{-1}u^0$ by $q_{1,h} = \hat{q}_1 + \chi\hat{q}_1 + A_h^{-1}\Delta(\chi\hat{q}_1)$. The error of this approximation can be estimated as follows:

$$\|q_{1,h} - q_1\|_{L^2(\Omega)} \le Ch^{2m+2}\|w_1 - \chi\hat{q}_1\|_{\dot{H}^{2m+2}(\Omega)} \le Ch^{2m+2}.$$

Since w_1 is in $H^{2m+2}(\Omega)$, it follows that $A^{-1}w_1$ can be approximated by $A_h^{-1}w_1$ with an error bound of $O(h^{2m+2})$. Therefore, in order to compute $q_2 = A^{-2}u^0 = A^{-1}\hat{q}_1 + A^{-1}w_1$ with an error bound of $O(h^{2m+2})$, it suffices to approximate $A^{-1}\hat{q}_1$ with the desired accuracy. This can be done

similarly as the approximation of $A^{-1}u^0$ by utilizing the following fact: The function

$$\hat{q}_2(x) = -\frac{1}{2\pi}|x-x_0|^2 \ln|x-x_0| + c|x-x_0|^2$$

satisfies the equation $-\Delta\hat{q}_2 = \hat{q}_1$. Therefore, the function $w_2 = A^{-1}\hat{q}_1 - \hat{q}_2$ is the solution of the following boundary value problem:

$$\begin{cases} -\Delta w_2 = 0 & \text{in } \Omega, \\ w_2 = -\hat{q}_2 & \text{for } x \in \partial\Omega, \end{cases} \tag{4.5}$$

which is in the same form as (4.4). Therefore, $A^{-1}\hat{q}_1$ can be computed with high-order accuracy similarly as the above-mentioned computation of $A^{-1}u^0$. Repeating this process will yield high-order approximations to $q_j = A^{-j}u^0$ with the following error bound:

$$\|q_{j,h} - q_j\|_{L^2(\Omega)} \le Ch^{2m+2}.$$

This shows that the singular part in (4.1) can be approximated with high-order accuracy if the initial value is a Dirac–Delta point source.

Example 4.3 (Dirac measure concentrated on an interface)**.** If the initial value is a Dirac measure concentrated on an oriented interface $\Gamma \subset \Omega$, i.e., $u^0 = \delta_\Gamma$ with

$$\langle u^0, v\rangle = \langle \delta_\Gamma, v\rangle := \int_\Gamma v \,\mathrm{d}s \qquad \text{for all } v \in \dot{H}^{\frac{1}{2}+\mu}(\Omega),\ \mu > 0.$$

Then the function $A^{-1}u^0$ can be approximated by $A_h^{-1}P_hu^0$ with error of $O(h^{2m+2})$ in $L^2(\Omega)$ by using a locally refined mesh towards the interface Γ; see e.g., [32, Theorem 4.8]. Similarly,

$$\begin{aligned}\|A^{-2}u^0 - A_h^{-1}P_hu^0\|_{L^2} &\le \|A^{-2}u^0 - A_h^{-1}P_hA^{-1}u^0\|_{L^2} + \|A_h^{-1}P_h(A^{-1}u^0 - A_h^{-1}P_hu^0)\|_{L^2} \\ &\le \|(A^{-1} - A_h^{-1}P_h)A^{-1}u^0\|_{L^2} + Ch^{2m+2}.\end{aligned}$$

Since $A^{-1}u^0$ is more regular than $u^0 = \delta_\Gamma$, the locally refined mesh also yields optimal-order approximation

$$\|(A^{-1} - A_h^{-1}P_h)A^{-1}u^0\|_{L^2} \le Ch^{2m+2}.$$

The approximation to $A^{-j}u^0$ is similar. The details are omitted.

Therefore, the singular part in (4.1) can be approximated with high-order accuracy if the initial value is a Dirac measure concentrated on an interface.

5. Extension to subdiffusion with nonsmooth source

In this section, we consider the subdiffusion problem with an inhomogeneous source term $g(t)f(x)$ where g and f are respectively temporally and spatially dependent functions, i.e.,

$$\partial_t^\alpha u(t) + Au(t) = g(t)f \quad 0 < t \le T, \quad \text{with } u(0) = 0. \tag{5.1}$$

By means of Laplace transform, the solution to (5.1) could be represented by

$$u(t) = \frac{1}{2\pi\mathrm{i}}\int_{\Gamma_{\theta,\kappa}} e^{zt}\hat{g}(z)(z^\alpha + A)^{-1}f \,\mathrm{d}z,$$

where $\hat{g}(z)$ denotes the Laplace transform of g. Using the identity (2.4), we have the splitting

$$u(t) = \sum_{j=1}^{m} u_j^s(t) + u^r(t) \tag{5.2}$$

where

$$u_j^s(t) = \sum_{j=0}^{m-1} \frac{(-1)^j}{2\pi \mathrm{i}} A^{-(j+1)} f \int_{\Gamma_{\theta,\kappa}} e^{zt} \hat{g}(z) z^{j\alpha} \,\mathrm{d}z =: \sum_{j=0}^{m-1} \Big((-1)^j A^{-(j+1)} f \Big) G_j(t)$$

$$u^r(t) = \frac{(-1)^m}{2\pi \mathrm{i}} \int_{\Gamma_{\theta,\kappa}} e^{zt} z^{m\alpha} \hat{g}(z) (z^\alpha + A)^{-1} A^{-m} f \,\mathrm{d}z.$$

Next, we briefly introduce the approximation to $u^r(t)$. Using the argument in Section 2.2, we apply the semidiscrete finite element method:

$$u_h^r(t) = \frac{(-1)^m}{2\pi \mathrm{i}} \int_{\Gamma_{\theta,\kappa}} e^{zt} z^{m\alpha} \hat{g}(z) (z^\alpha + A_h)^{-1} A_h^{-m} P_h f \,\mathrm{d}z.$$

By assuming that $g \in C^{\lfloor m\alpha \rfloor + 1}[0,T]$, then the Taylor expansion and Lemmas 2.3 and 2.4 imply the following error estimate

$$\|u^r(t) - u_h^r(t)\|_{L^2(\Omega)} \le c h^{2m+2} \|f\|_{L^2(\Omega)} \Big(\sum_{\ell=0}^{\lfloor m\alpha \rfloor} |g^{(\ell)}(0)| t^{-m\alpha+\ell} + \int_0^t |g^{\lfloor m\alpha \rfloor + 1}(t-s)| s^{-m\alpha + \lfloor m\alpha \rfloor} \,\mathrm{d}s \Big)$$

We then apply convolution quadrature to discretize in the time variable. Let $\delta(\cdot)$ be the generating function of BDFk method defined in (3.1). By assuming that $g \in C^K[0,T]$ with $K = \lfloor (m-1)\alpha \rfloor + k$, we apply the Taylor expansion to derive

$$\begin{aligned} u_h^r(t) = &\sum_{\ell=0}^{K} \frac{(-1)^m [g^{(\ell)}(0)]}{2\pi \mathrm{i}} \int_{\Gamma_{\theta,\kappa}} e^{zt} z^{m\alpha-\ell} (z^\alpha + A_h)^{-1} A_h^{-m} P_h f \,\mathrm{d}z \\ &+ (-1)^m \frac{1}{2\pi \mathrm{i}} \int_{\Gamma_{\theta,\kappa}} e^{zt} z^{m\alpha} \widehat{R}_K(z) (z^\alpha + A_h)^{-1} A_h^{-m} P_h f \,\mathrm{d}z \end{aligned}$$

where $R_K(t) = \frac{t^K}{K!} * g^{(K+1)}$ denotes the remainder of the Taylor series. Then we consider the time stepping approximation by convolution quadrature:

$$\begin{aligned} U_h^{r,n} = &\sum_{\ell=0}^{K} \frac{(-1)^m [g^{(\ell)}(0)]}{2\pi \mathrm{i}} \int_{\Gamma_{\theta,\kappa}^\tau} e^{zt_n} \delta_\tau(e^{-z\tau})^{m\alpha-\ell} (\delta_\tau(e^{-z\tau})^\alpha + A_h)^{-1} A_h^{-m} P_h f \,\mathrm{d}z \\ &+ (-1)^m \frac{1}{2\pi \mathrm{i}} \int_{\Gamma_{\theta,\kappa}^\tau} e^{zt} \delta_\tau(e^{-z\tau})^{m\alpha} \widetilde{R}_K(\delta_\tau(e^{-z\tau})) (\delta_\tau(e^{-z\tau})^\alpha + A_h)^{-1} A_h^{-m} P_h f \,\mathrm{d}z \end{aligned} \tag{5.3}$$

where $\widetilde{R}_K(\xi) = \sum_{\ell=0}^{\infty} R_K(t_\ell) \xi^\ell$. Note that the fully discrete scheme could be solved via a time stepping manner. Then using the argument in Section 3, we have the error estimate

$$\|u_h^r(t_n) - U_h^{r,n}\|_{L^2(\Omega)} \le c\tau^k \Big(\sum_{\ell=0}^{K} |g^{(\ell)}(0)| t_n^{\ell-k-(m-1)\alpha} + \int_0^t |g^{(K+1)}(s)| (t-s)^{\lfloor (m-1)\alpha \rfloor - (m-1)\alpha} \,\mathrm{d}s \Big) \|f\|_{L^2(\Omega)}.$$

Similarly, we can approximate the function $G_j(t)$ in $u_j^s(t)$ by using convolution quadrature generated by BDFk. Then we only need to solve an elliptic problem $A^{-(j+1)} f$ in $u_j^s(t)$ accurately, see Example 4.1-4.3.

Moreover, the above argument could be further generalized to the problem

$$\partial_t^\alpha u(t) + Au(t) = \sum_{i=1}^{B} g_i(t) f_i \quad 0 < t \le T, \quad \text{with } u(0) = 0, \tag{5.4}$$

where g_i and f_i are respectively temporally and spatially dependent functions for all $i = 1, \ldots, B$.

6. Numerical experiments

In the section, we present numerical experiments to support the theoretical analysis and to illustrate the high-order convergence of the proposed method for nonsmooth initial data. Throughout, we consider a two-dimensional subdiffusion model (1.1) in a unit square domain $\Omega = (0,1)^2 \subset \mathbb{R}^2$. In our computation, the spatial mesh size be $h_j = h_0/2^j$, and step size be $\tau_j = \tau^0/2^j$, where h_0 and τ^0 will be specified later. The errors are computed by the Cauchy difference

$$E_{h_j} = \|u_{\tau_{\rm ref},h_j} - u_{\tau_{\rm ref},h_{j+1}}\|_{L^2(\Omega)}, \qquad E_{\tau_j} = \|u_{\tau_j,h_{\rm ref}} - u_{\tau_{j+1},h_{\rm ref}}\|_{L^2(\Omega)}, \tag{6.1}$$

and the convergence orders are computed by using the following formulae:

$$\begin{aligned} &\text{spatial convergence order} = -(\log(E_{h_{j+1}}) - \log(E_{h_j}))/\log 2, \\ &\text{temporal convergence order} = -(\log(E_{\tau_{j+1}}) - \log(\tau_{h_j}))/\log 2. \end{aligned} \tag{6.2}$$

Let r be the degree of finite elements in the spatial discretization, and k be the order of the time-stepping method. We illustrate the convergence of the time discretization for $k = 1, 2, 3, 4$ by fixing $m = 1$ and $r = 3$, and illustrate the convergence of the spatial discretization with different m ($m = 0, 1$) and r ($r = 1, 2, 3$) by fixing $k = 4$. All the examples are performed by Firedrake [41], and the meshes are generated by Gmsh [10].

Example 6.1 (Dirac delta initial value)**.** In the first example, we test the very weak initial condition $u^0 = \delta_{x_0}$, where δ_{x_0} denotes the Dirac delta measure concentrated at the single point $x_0 = (0.5 + \epsilon, 0.5 + \epsilon)$ with $\epsilon = 10^{-4}$. Here, a perturbation is given to move the source away from the vertex of the meshes.

To test the temporal convergence order of the fully discrete solution (3.6) for different k, we set $\tau_j = \tau^0/2^j$ with $\tau_0 = 1/32$ and a fixed spatial meshes $h_{\rm ref} = 1/512$. The results of the L^2-errors are presented in Figure 1, and confirm kth-order convergence for the BDFk method.

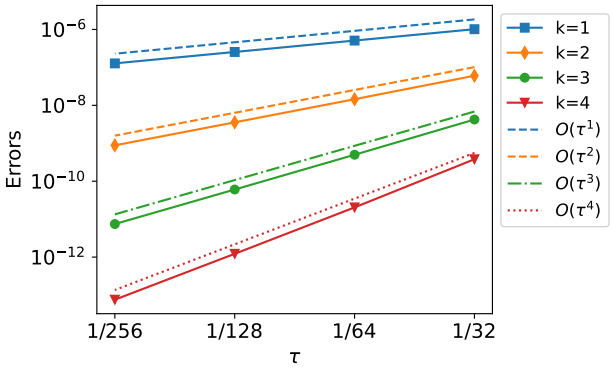
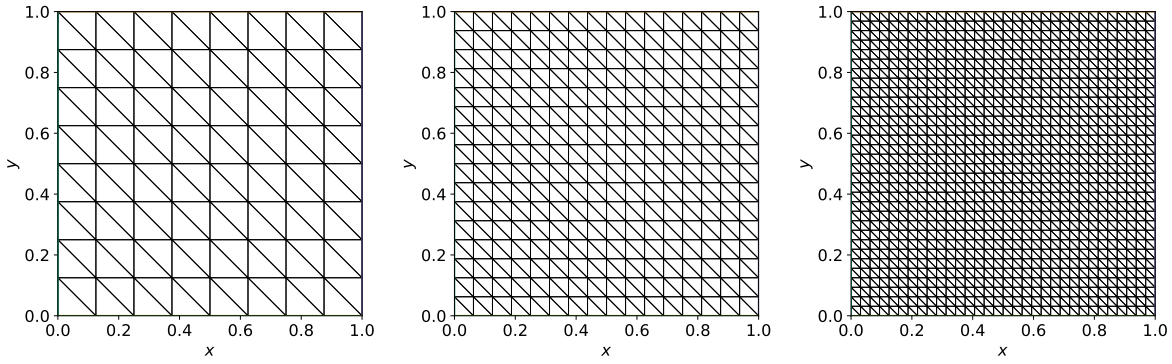


FIGURE 1. Example 6.1: Errors of time discretization with $\alpha = 0.6$, $m = 1$, $r = 3$. The dashed lines are $O(\tau^k)$.

To test the high-order convergence in space, we set $h_j = h_0/2^j$ with $h_0 = 1/16$ and $\tau_{\rm ref} = 1/1024$. The meshes are refined by subdividing the triangles into four congruent sub-triangles (cf. Figure 2). In Table 1, we test the convergence of the standard Galerkin finite element method, i.e. $m = 0$, using piecewise r-th order polynomials, for both subdiffusion equation ($\alpha = 0.6$) and normal diffusion equation ($\alpha = 1$). The empirical convergence for the fractional subdiffusion equation is always first-order, while that for the normal diffusion equation is of order $r + 1$. This interesting phenomenon is attributed to the infinite smoothing effect of the normal diffusion and the limited smoothing property of the fractional subdiffusion, cf. (1.4). Note that the Dirac delta function is in $\dot{H}^{-1-\mu}(\Omega)$ with any $\mu > 0$ and hence the solution to the subdiffusion equation (1.1) belongs to $\dot{H}^{1-\mu}(\Omega)$. In

order to improve the convergence, we apply the splitting strategy (2.6) with $m = 1$, approximate the regular part using the fully discrete scheme (3.3), and compute the singular part using the method provided in Section 4.2. In Table 2, we present the errors for both the regular part and the singular part. The convergence for the regular part could be improved to $O(h^{\min(3,r+1)})$ and the singular part could be approximated with order $O(h^{\min(4,r+1)})$. This convergence could be further improved by splitting one more singular term. See Table 3 for the approximation with $m = 2$. These results fully supports our theoretical findings and the necessity of the proposed splitting method.



(a) $j = 0$ (b) $j = 1$ (c) $j = 2$

FIGURE 2. The meshes for Example 6.1.

TABLE 1. Example 6.1: comparison of $\alpha = 0.6$ and $\alpha = 1$, with $m = 0$.

	$\alpha = 0.6$						$\alpha = 1$					
	$r = 1$		$r = 2$		$r = 3$		$r = 1$		$r = 2$		$r = 3$	
h_j	E_{h_j}	conv.	E_{h_j}	conv.	E_{h_j}	conv.	E_{h_j}	conv.	E_{h_j}	conv.	E_{h_j}	conv.
1/32	1.59e-04	-	1.02e-04	-	6.60e-05	-	1.79e-10	-	1.22e-13	-	1.40e-15	-
1/64	7.86e-05	1.02	4.92e-05	1.05	3.06e-05	1.11	4.58e-11	1.97	1.25e-14	3.28	8.40e-17	4.06
1/128	3.86e-05	1.03	2.30e-05	1.10	1.36e-05	1.18	1.15e-11	1.99	1.46e-15	3.09	4.87e-18	4.11
1/256	1.87e-05	1.04	1.04e-05	1.15	6.32e-06	1.10	2.89e-12	2.00	1.80e-16	3.03	2.80e-19	4.12
theor. conv.		1.00		1.00		1.00		2.00		3.00		4.00

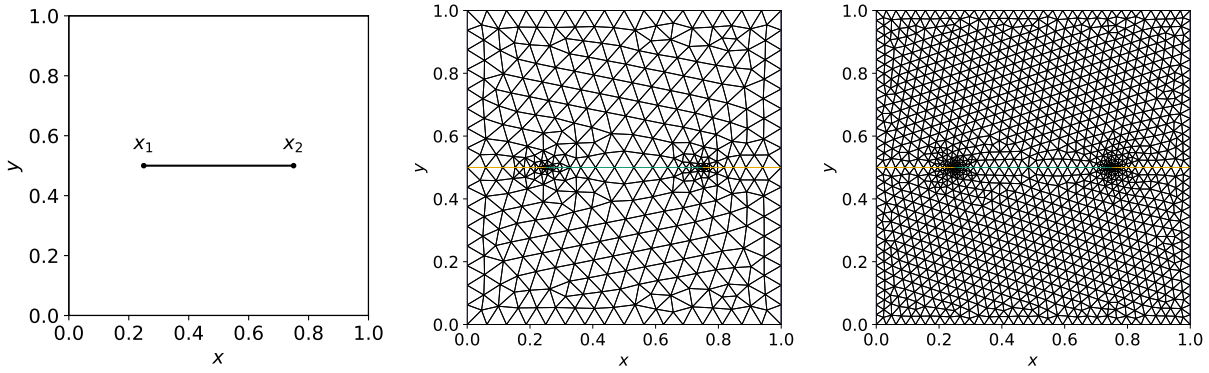
TABLE 2. Example 6.1: improved spatial convergence using (3.6) with $m = 1$.

$\alpha = 0.6$	regular part						singular part					
	$r = 1$		$r = 2$		$r = 3$		$r = 1$		$r = 2$		$r = 3$	
h_j	E_{h_j}	conv.	E_{h_j}	conv.	E_{h_j}	conv.	E_{h_j}	conv.	E_{h_j}	conv.	E_{h_j}	conv.
1/32	3.06e-06	-	2.61e-08	-	2.53e-09	-	2.89e-04	-	3.92e-05	-	7.85e-06	-
1/64	7.71e-07	1.99	3.54e-09	2.88	3.16e-10	3.00	8.45e-05	1.77	5.65e-06	2.80	5.51e-07	3.83
1/128	1.95e-07	1.98	4.76e-10	2.89	4.00e-11	2.98	2.24e-05	1.92	7.32e-07	2.95	3.39e-08	4.02
1/256	4.97e-08	1.97	6.42e-11	2.89	5.33e-12	2.91	5.68e-06	1.98	9.25e-08	2.98	2.09e-09	4.02
theor. conv.		2.00		3.00		3.00		2.00		3.00		4.00

TABLE 3. Example 6.1: improved spatial convergence using (3.6) with $m = 2$.

$\alpha = 0.6$	regular part						singular part					
	$r = 1$		$r = 2$		$r = 3$		$r = 1$		$r = 2$		$r = 3$	
h_j	E_{h_j}	conv.	E_{h_j}	conv.	E_{h_j}	conv.	E_{h_j}	conv.	E_{h_j}	conv.	E_{h_j}	conv.
1/32	5.69e-07	-	2.08e-09	-	3.29e-11	-	2.77e-04	-	3.77e-05	-	7.49e-06	-
1/64	1.43e-07	1.99	2.61e-10	3.00	1.98e-12	4.05	8.11e-05	1.77	5.41e-06	2.80	5.27e-07	3.83
1/128	3.60e-08	1.99	3.26e-11	3.00	1.14e-13	4.12	2.14e-05	1.92	7.01e-07	2.95	3.25e-08	4.02
1/256	9.13e-09	1.98	4.08e-12	3.00	7.10e-15	4.00	5.44e-06	1.98	8.86e-08	2.98	2.00e-09	4.02
theor. conv.		2.00		3.00		4.00		2.00		3.00		4.00

Example 6.2 (Dirac measure concentrated on an interface)**.** In the second example, we test the initial condition $u^0 = \delta_\Gamma$, where δ_Γ denotes the Dirac measure concentrated on an oriented interface $\Gamma = \overrightarrow{x_1x_2}$ with $x_1 = (0.25, 0.75)$, $x_2 = (0.75.0.5)$, cf. Figure 3 (a).



(a) Line segment Γ (b) $r = 2, j = 1$ (c) $r = 2, j = 2$

FIGURE 3. Example 6.2: Line segment Γ and the graded mesh for the singular part.

In order to reduce the computational cost, we use quasi-uniform meshes and locally graded meshes for the time-dependent regular part and the steady singular part, respectively. To generate the quasi-uniform meshes, we generate the initial mesh with mesh size $h_0 = 1/8$ by Gmsh, and refine the mesh several times to reach the mesh size $h_j = h_0/2^j$. For the singular part, we generate the jth-level graded meshes with the local cell diameter $\hbar(x)$ in sub-domain $B(x_i, d_0)$ as

$$\hbar(x) \backsim \begin{cases} |x - x_i|^{1-\gamma} h_j, & \text{for} \quad h_\star \le |x - x_i| \le d_0, \\ h_*, & \text{for} \quad |x - x_i| \le h_* \backsim h_j^{1/\gamma}. \end{cases} \tag{6.3}$$

where $\gamma \in (0, 1/r)$. The graded mesh for approximating the singular part are presented in Figure 3 (b) and (c). Note that the refinement strategy used here is different from the method proposed in [32], where the local mesh size for the jth-level graded meshes in the neighborhood of x_0 and x_1 is

$$\hbar(x) \backsim \begin{cases} |x - x_i|(1 - c_p)h_j, & \text{for} \quad h_\star \le |x - x_i| \le d_0, \\ h_\star, & \text{for} \quad |x - x_i| \le h_\star \backsim \kappa_p^j h_j, \end{cases} \tag{6.4}$$

where $c_p = 2^{-r/a}$ with $a \in (0, 1)$ [32, Algorithm 4.1]. As proved in [32, Theorem 3.8], the solution of the Possion equation with line Dirac source belongs to weighted Sobolev space

$$\mathcal{K}^{l+1}_{a+1}(B(x_i, d)\backslash\Gamma) := \{v : \rho^{|s|-a-1} D^s v \in L^2(B(x_i, d)\backslash\Gamma), \forall |s| \le l + 1\} \tag{6.5}$$

for any $l \geq 1$ and $a \in (0,1)$ in the neighborhood of x_1 and x_2. Though the refine methods given above are different, both of them can resolve the singularity around the end point of Γ and obtain optimal convergence order.

To test the temporal convergence order, we let the step sizes be $\tau_j = \tau_0/2^j$ with $\tau_0 = 1/32$ and fixed the spatial mesh size $h_{\text{ref}} = h_6 = 1/512$. As shown in Figure 4, the convergence order of BDFk scheme is $O(\tau^k)$, which agrees well with our theoretical result in Corollary 3.1.

To test the convergence in space, we first compare the numerical results of $\alpha = 0.6$ and $\alpha = 1$ using the standard finite element method (i.e., $m = 0$) with uniform meshes. As shown in Table 4, the convergence for the fractional diffusion is at most second-order even if we use high-order finite element methods, while the convergence order of the normal diffusion is $r + 1$. In order to improve the convergence, we apply the splitting method with $m = 1$. The empirical errors of regular part and singular part are presented in Table 5. Our numerical experiments indicate the optimal convergence rate for the P^r finite element method with $r = 2, 3$.

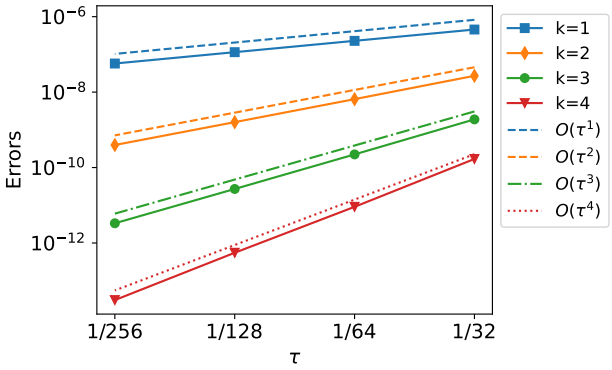


FIGURE 4. Example 6.2: Errors of time discretization with $m = 1$ and $r = 3$.

TABLE 4. Example 6.2: Comparison of $\alpha = 0.6$ and $\alpha = 1$, with $m = 0$.

	$\alpha = 0.6$						$\alpha = 1$					
	$r = 1$		$r = 2$		$r = 3$		$r = 1$		$r = 2$		$r = 3$	
h_j	E_{h_j}	conv.	E_{h_j}	conv.	E_{h_j}	conv.	E_{h_j}	conv.	E_{h_j}	conv.	E_{h_j}	conv.
1/32	3.35e-05	-	4.40e-06	-	1.42e-06	-	3.90e-11	-	2.35e-14	-	1.30e-16	-
1/64	8.99e-06	1.90	1.10e-06	2.00	3.56e-07	2.00	9.85e-12	1.98	2.71e-15	3.12	8.14e-18	4.00
1/128	2.39e-06	1.91	2.75e-07	2.00	8.91e-08	2.00	2.47e-12	2.00	3.31e-16	3.03	5.09e-19	4.00
1/256	6.31e-07	1.92	6.88e-08	2.00	2.23e-08	2.00	6.18e-13	2.00	4.12e-17	3.01	3.18e-20	4.00
theor. conv.		2.00		2.00		2.00		2.00		3.00		4.00

To summarize, the numerical experiments demonstrate that the splitting strategy substantially enhances the convergence of numerical schemes, as indicated in Corollary 3.1. When performing simulations, we advise selecting the parameter m based on the available computational resources and the required precision of the approximation. For instance, if the initial data is a Dirac measure focused on an interface (Example 6.2), setting $m = 1$ would be appropriate. Employing the P^3 finite element method for spatial discretization, coupled with the BDF4 convolution quadrature for temporal discretization, yields a numerical scheme that is fourth-order accurate in both space and time. This level of accuracy is typically more than adequate for practical applications.

TABLE 5. Example 6.2: Improved spatial convergence by using (3.6) with $m = 1$.

$\alpha = 0.6$	regular part						singular part					
	$r = 1$		$r = 2$		$r = 3$		$r = 1$		$r = 2$		$r = 3$	
h_j	E_{h_j}	conv.	E_{h_j}	conv.	E_{h_j}	conv.	E_{h_j}	conv.	E_{h_j}	conv.	E_{h_j}	conv.
1/32	5.66e-07	-	3.95e-09	-	5.03e-11	-	3.33e-05	-	4.07e-07	-	1.15e-08	-
1/64	1.42e-07	2.00	4.96e-10	2.99	3.41e-12	3.88	8.97e-06	1.89	5.26e-08	2.95	7.82e-10	3.87
1/128	3.55e-08	2.00	6.21e-11	3.00	2.29e-13	3.90	2.38e-06	1.91	6.46e-09	3.03	4.86e-11	4.01
1/256	8.88e-09	2.00	7.77e-12	3.00	1.52e-14	3.91	5.16e-07	2.21	8.57e-10	2.92	3.04e-12	4.00
theor. conv.		2.00		3.00		4.00		2.00		3.00		4.00

7. Conclusions

We have constructed a new splitting of the solution to the subdiffusion equation, which allows us to develop high-order finite element approximations in case of nonsmooth initial data. In this method, the solution is split into a time-dependent smooth part plus a time-independent nonsmooth part. We have developed high-order spatial and time discretizations to approximate the smooth part of the solution, and proved that the proposed fully discrete finite element method approximates the regular part of the solution to high-order accuracy for nonsmooth initial data in $L^2(\Omega)$. Moreover, we have illustrated how to approximate the time-independent nonsmooth part through several examples of initial data, including piecewise smooth initial data, Dirac–Delta point source, and Dirac measure concentrated on an interface. More generally, the time-independent nonsmooth part can be approximated by using smaller mesh size without increasing the overall computational cost significantly. This is possible as the nonsmooth part is time-independent and therefore avoids the time-stepping procedure. We have also illustrated the effectiveness of the proposed method through several numerical examples.

Acknowledgements

The research of B. Li is partially supported by the National Natural Science Foundation of China (NSFC project 12231003) and Hong Kong Research Grants Council (GRF Project No. 15300519). The research of Z. Yang is partially supported by an internal grant of The Hong Kong Polytechnic University (Project ID: P0031035, Work Programme: ZZKQ). The research of Z. Zhou is partially supported by Hong Kong Research Grants Council (Project No. 25300818) and an internal grant of The Hong Kong Polytechnic University (Project ID: P0031041, Work Programme: ZZKS).

Appendix: On the triangulation satisfying Assumption 2.1

In this Appendix we show that the graded mesh defined in (2.11), for a two-dimensional polygonal domain Ω, satisfies Assumption 2.1.

In terms of the notation introduced in (2.11) and the subsequent text, we divide the domain $D_0 = \{x \in \Omega : |x - x_0| < d_0\}$ into $D_0 = D_* \cup (\cup_{j=1}^{J} D_j)$ with

$$D_j := \{x \in \Omega : 2^{-j-1} d_0 \le |x - x_0| < 2^{-j} d_0\} \quad \text{and} \quad D_* := \{x \in \Omega : |x - x_0| \le 2^{-J} d_0 = h_*\}.$$

Let $d_j = 2^{-j} d_0$ and $h_j = \max_{x \in D_j} h(x) \sim d_j^{1-\gamma} h$, and denote by

$$D_j' := \{x \in \Omega : 2^{-j-2} d_0 \le |x - x_0| < 2^{-j+1} d_0\}$$

a neighborhood of D_j. Then the following lemma provides a regularity estimate near the corner.

Lemma A.1. *If $A=-\Delta$ and $v\in \dot{H}^{2r+2}(\Omega)$ for some $r\geq 0$, then*

$$\|v\|_{H^s(D_j)}\leq Cd_j^{1-s+\min(1,\pi/\theta)}\|v\|_{\dot{H}^{2r+2}(\Omega)} \quad \text{for } 0\leq s\leq 2r+2.$$

Proof. For $v\in \dot{H}^{2r+2}(\Omega)$, the following weighted regularity result is known (for example, see [29, Proof of Lemma 5.1, inequality (5.8)]):

$$|v|_{H^{k+1}(D_j)}\leq C\sum_{i=0}^{k-1} d_j^{-i}\|\Delta v\|_{H^{k-1-i}(D_j')}+Cd_j^{-k}\|\nabla v\|_{L^2(D_j')} \quad 1\leq k\leq 2r+1. \tag{A.1}$$

By using the singular expansions and local energy estimate, i.e.,

$$v|_{D_0}\in O(|x-x_0|^{\pi/\theta})+H^2(D_0) \quad \text{and} \quad \|\nabla v\|^2_{L^2(D_j)}\leq Cd_j^{-2}\|v\|^2_{L^2(D_j')}+Cd_j^2\|f\|^2_{L^2(D_j')},$$

we further obtain

$$|v|_{H^{k+1}(D_j)}\leq C\sum_{i=0}^{k-1} d_j^{-i}\|\Delta v\|_{H^{k-1-i}(D_j')}+Cd_j^{-k+\min(1,\pi/\theta)}\|f\|_{L^2(\Omega)} \quad 1\leq k\leq 2r+1.$$

If $\|\Delta v\|_{H^s(D_j')}\leq Cd_j^{-s+\min(1,\pi/\theta)}$ for $0\leq s\leq 2r$, then the inequality above yields

$$|v|_{H^s(D_j)}\leq Cd_j^{1-s+\min(1,\pi/\theta)}, \quad 2\leq s\leq 2r+2.$$

By using the singular expansions and local energy estimate in (A.1), we find that the inequality above also holds for $s=0,1$. Since $d_j\leq C$, it follows that $\|v\|_{H^s(D_j')}\leq Cd_j^{-s+\min(1,\pi/\theta)}$ for $0\leq s\leq 2r+2$. This is the same estimate as that for Δv, but the range of s increases by 2. Therefore, for any function $f\in L^2(\Omega)$, by picking up the regularity from $\Delta^{-l}f$ to $\Delta^{-l-1}f$ for $l=0,1,\ldots,r$, we obtain

$$|\Delta^{-r-1}f|_{H^s(D_j)}\leq Cd_j^{1-s+\min(1,\pi/\theta)}\|f\|_{L^2(\Omega)} \quad \text{for} \quad 0\leq s\leq 2r+2.$$

Replacing $\Delta^{-r-1}f$ by v in the estimate above, we obtain the result of Lemma A.1. □

Then the following proposition confirms the approximation properties (2.9)–(2.10).

Proposition A.1. *Let $A=-\Delta$. If the diameter of triangles satisfies condition (2.11) near every corner of the domain, then (2.9)–(2.10) holds.*

Proof. Let $v\in \dot{H}^{2r+2}(\Omega)$ and $s=2r+1$, with $0\leq r\leq m$. We consider the following decomposition:

$$\begin{aligned}\|v-I_hv\|^2_{H^1(D_0)}&=\|v-I_hv\|^2_{H^1(D_*)}+\sum_{j=1}^J\|v-I_hv\|^2_{H^1(D_j)}\\ &\leq Ch_*^{2\min(1,\pi/\theta)}\|\Delta v\|^2_{L^2(\Omega)}+C\sum_{j=1}^J h_j^{2s}\|v\|^2_{H^{s+1}(D_j')},\end{aligned} \tag{A.2}$$

and

$$\begin{aligned}\|v-I_hv\|^2_{L^2(D_0)}&=\|v-I_hv\|^2_{L^2(D_*)}+\sum_{j=1}^J\|v-I_hv\|^2_{L^2(D_j)}\\ &\leq Ch_*^{2+2\min(1,\pi/\theta)}\|\Delta v\|^2_{L^2(\Omega)}+C\sum_{j=1}^J h_j^{2s+2}\|v\|^2_{H^{s+1}(D_j')},\end{aligned} \tag{A.3}$$

where $h_j=\max_{x\in D_j}\hbar(x)$, and we have used the following result:

$$\|v-I_hv\|_{H^1(D_*)}\leq Ch_*^{\min(1,\pi/\theta)}\|\Delta v\|_{L^2(\Omega)}. \tag{A.4}$$

In the case $\pi/\theta < 1$, this result follows from

$$\|v - I_h v\|_{H^1(D_*)} \le C h_*^{\min(1,\pi/\theta)} \|v\|_{B_{2,\infty}^{\min(1,\pi/\theta)}(\Omega)} \le C h_*^{\min(1,\pi/\theta)} \|\Delta v\|_{L^2(\Omega)},$$

where $B_{2,\infty}^{\min(1,\pi/\theta)}(\Omega)$ is the Besov space. The last inequality is a consequence of the singularity expansion

$$v|_{D_0} = c|x - x_0|^{\pi/\theta} \sin(\arg(x - x_0)) + w \text{ for some } c \in \mathbb{R} \text{ and } w \in H^2(D_0),$$

where $|c| \le \|\Delta v\|_{L^2(\Omega)}$. In the case $\pi/\theta > 1$, (A.4) follows from the standard estimate for the Lagrange interpolation, i.e.,

$$\|v - I_h v\|_{H^1(D_*)} \le C h_* \|v\|_{H^2(\Omega)} \le C h_* \|\Delta v\|_{L^2(\Omega)}.$$

By substituting the result of Lemma A.1 into (A.2) and using the condition $\gamma_j < \frac{\min(1,\pi/\theta_j)}{r}$, we obtain

$$\begin{aligned}
\|v - I_h v\|_{H^1(D_0)}^2 &\le C h_*^{2\min(1,\pi/\theta)} \|v\|_{\dot{H}^{2m+2}(\Omega)}^2 + \sum_j C h_j^{2s} d_j^{-2s+2\min(1,\pi/\theta_j)} \|v\|_{\dot{H}^{2r+2}(\Omega)}^2 \\
&\le C h^{2s} \|v\|_{\dot{H}^{2r+2}(\Omega)}^2 + \sum_j C d_j^{(1-\gamma_j)2s-2s+2\min(1,\pi/\theta_j)} h^{2s} \|v\|_{\dot{H}^{2r+2}(\Omega)}^2 \\
&\le C h^{2s} \|v\|_{\dot{H}^{2r+2}(\Omega)}^2 + \sum_j C d_j^{2s\left(\frac{\min(1,\pi/\theta_j)}{s} - \gamma_j\right)} h^{2s} \|v\|_{\dot{H}^{2r+2}(\Omega)}^2 \\
&\le C h^{2s} \|v\|_{\dot{H}^{2r+2}(\Omega)}^2.
\end{aligned}$$

This proves that, by substituting $s = 2r + 1$ into the inequality above,

$$\|v - I_h v\|_{H^1(\Omega)} \le C h^{2r+1} \|v\|_{\dot{H}^{2r+2}(\Omega)}.$$

Similarly, by substituting the result of Lemma A.1 into (A.3), we obtain

$$\|v - I_h v\|_{L^2(\Omega)} \le C h^{r+1} \|v\|_{\dot{H}^{2r+2}(\Omega)}.$$

This proves the desired estimate in (2.9).

By the optimal H^1-norm approximation property of the Ritz projection, we have

$$\|v - R_h v\|_{H^1(\Omega)} \le C \|v - I_h v\|_{H^1(\Omega)} \le C h^{2r+1} \|v\|_{\dot{H}^{2r+2}(\Omega)}.$$

By a standard duality argument, we obtain

$$\|v - R_h v\|_{L^2(\Omega)} \le C h \|v - R_h v\|_{H^1(\Omega)} \le C h^{2r+2} \|v\|_{\dot{H}^{2r+2}(\Omega)}.$$

This proves the desired estimate in (2.10). □

Lemma A.2. *If the mesh size satisfies condition (2.11), then the following estimate holds:*

$$\|v - I_h v\|_{L^\infty(\Omega)} \le c h^{2r+1} \|v\|_{\dot{H}^{2r+2}(\Omega)} \quad \text{for } 0 \le r \le m. \tag{A.5}$$

Proof. The basic L^∞ estimates of the Lagrange interpolation says that

$$\|v - I_h v\|_{L^\infty(D_j)} \le c h_j^{2r+1} \|v\|_{H^{2r+2}(D_j')}.$$

Since $\|v\|_{H^{2r+2}(D_j')} \le c d_j^{-2r-1+\min(1,\pi/\theta)} \|v\|_{\dot{H}^{2r+2}(\Omega)}$, it follows that

$$\begin{aligned}
\|v - I_h v\|_{L^\infty(D_j)} &\le c d_j^{-2r-1+\min(1,\pi/\theta)} h_j^{2r+1} \|v\|_{\dot{H}^{2r+2}(\Omega)} \\
&\le c d_j^{-2r-1+\min(1,\pi/\theta)} d_j^{(1-\gamma)(2r+1)} h^{2r+1} \|v\|_{\dot{H}^{2r+2}(\Omega)} \\
&\le c d_j^{\min(1,\pi/\theta)-(2r+1)\gamma} h^{2r+1} \|v\|_{\dot{H}^{2r+2}(\Omega)}.
\end{aligned}$$

Then (A.5) follows from the condition $\gamma < \min(1, \pi/\theta)/(2m + 1) \le \min(1, \pi/\theta)/(2r + 1)$ in (2.11). □

References

[1] E. E. Adams and L. W. Gelhar. Field study of dispersion in a heterogeneous aquifer: 2. spatial moments analysis. *Water Res. Research*, 28(12):3293–3307, 1992.

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

[3] L. Banjai and M. López-Fernández. Efficient high order algorithms for fractional integrals and fractional differential equations. *Numer. Math.*, 141:289–317, 2019.

[4] L. Banjai and C. G. Makridakis. A posteriori error analysis for approximations of time-fractional subdiffusion problems. *Math. Comp.*, 91(336):1711–1737, 2022.

[5] B. Cockburn and K. Mustapha. A hybridizable discontinuous Galerkin method for fractional diffusion problems. *Numer. Math.*, 130(2):293–314, 2015.

[6] M. Crouzeix and V. Thomée. The stability in l_p and w_p^1 of the l_2-projection onto finite element function spaces. *Math. Comp.*, pages 521–532, 1987.

[7] E. Cuesta, C. Lubich, and C. Palencia. Convolution quadrature time discretization of fractional diffusion-wave equations. *Math. Comp.*, 75(254):673–696, 2006.

[8] S. Franz and N. Kopteva. Pointwise-in-time a posteriori error control for higher-order discretizations of time-fractional parabolic equations. *J. Comput. Appl. Math.*, 427:Paper No. 115122, 18, 2023.

[9] H. Fujita and T. Suzuki. Evolution problems. In *Handbook of numerical analysis, Vol. II*, Handb. Numer. Anal., II, pages 789–928. North-Holland, Amsterdam, 1991.

[10] C. Geuzaine and J.-F. Remacle. Gmsh: A 3-d finite element mesh generator with built-in pre- and post-processing facilities. *International Journal for Numerical Methods in Engineering*, 79(11):1309–1331, 2009.

[11] E. Hairer and G. Wanner. *Solving Ordinary Differential Equations. II*. Springer-Verlag, Berlin, second edition, 1996. Stiff and differential-algebraic problems.

[12] Y. Hatano and N. Hatano. Dispersive transport of ions in column experiments: An explanation of long-tailed profiles. *Water Res. Research*, 34(5):1027–1033, 1998.

[13] B. Jin. *Fractional Differential Equations*. Springer, Switzerland, 2021.

[14] B. Jin, R. Lazarov, J. Pasciak, and Z. Zhou. Galerkin FEM for fractional order parabolic equations with initial data in H^{-s}, $0 \le s \le 1$. In *Numerical analysis and its applications*, volume 8236 of *Lecture Notes in Comput. Sci.*, pages 24–37. Springer, Heidelberg, 2013.

[15] B. Jin, R. Lazarov, and Z. Zhou. Error estimates for a semidiscrete finite element method for fractional order parabolic equations. *SIAM J. Numer. Anal.*, 51(1):445–466, 2013.

[16] B. Jin, R. Lazarov, and Z. Zhou. Numerical methods for time-fractional evolution equations with nonsmooth data: a concise overview. *Comput. Methods Appl. Mech. Engrg.*, 346:332–358, 2019.

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

[18] B. Jin and Z. Zhou. *Numerical treatment and analysis of time-fractional evolution equations*, volume 214 of *Applied Mathematical Sciences*. Springer, Cham, [2023] ©2023.

[19] S. Karaa. Semidiscrete finite element analysis of time fractional parabolic problems: a unified approach. *SIAM J. Numer. Anal.*, 56(3):1673–1692, 2018.

[20] S. Karaa, K. Mustapha, and A. K. Pani. Finite volume element method for two-dimensional fractional subdiffusion problems. *IMA J. Numer. Anal.*, 37(2):945–964, 2017.

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

[22] A. A. Kilbas, H. M. Srivastava, and J. J. Trujillo. *Theory and applications of fractional differential equations*, volume 204 of *North-Holland Mathematics Studies*. Elsevier Science B.V., Amsterdam, 2006.

[23] N. Kopteva. Error analysis of an L2-type method on graded meshes for a fractional-order parabolic problem. *Math. Comp.*, 90:19–40, 2021.

[24] N. Kopteva. Pointwise-in-time a posteriori error control for time-fractional parabolic equations. *Appl. Math. Lett.*, 123:Paper No. 107515, 8, 2022.

[25] N. Kopteva and X. Meng. Error analysis for a fractional-derivative parabolic problem on quasi-graded meshes using barrier functions. *SIAM J. Numer. Anal.*, 58:1217–1238, 2020.

[26] N. Kopteva and M. Stynes. A posteriori error analysis for variable-coefficient multiterm time-fractional subdiffusion equations. *J. Sci. Comput.*, 92(2):Paper No. 73, 23, 2022.

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

[28] K. N. Le, W. McLean, and K. Mustapha. Numerical solution of the time-fractional Fokker-Planck equation with general forcing. *SIAM J. Numer. Anal.*, 54(3):1763–1784, 2016.

[29] B. Li. Maximum-norm stability of the finite element method for the Neumann problem in nonconvex polygons with locally refined mesh. *Math. Comp.*, 2022, DOI: 10.1090/mcom/3724.

[30] B. Li, Y. Lin, S. Ma, and Q. Rao. An exponential spectral method using vp means for semilinear subdiffusion equations with rough data. *SIAM J. Numer. Anal.*, 61:2305–2326, 2023.

[31] B. Li and S. Ma. Exponential convolution quadrature for nonlinear subdiffusion equations with nonsmooth initial data. *SIAM J. Numer. Anal.*, 60(2):503–528, 2022.
[32] H. Li, X. Wan, P. Yin, and L. Zhao. Regularity and finite element approximation for two-dimensional elliptic equations with line Dirac sources. *J. Comput. Appl. Math.*, 393:Paper No. 113518, 16, 2021.
[33] J. Li, J. M. Melenk, B. Wohlmuth, and J. Zou. Optimal a priori estimates for higher order finite elements for elliptic interface problems. *Appl. Numer. Math.*, 60:19–37, 2010.
[34] C. Lubich, I. H. Sloan, and V. Thomée. Nonsmooth data error estimates for approximations of an evolution equation with a positive-type memory term. *Math. Comp.*, 65(213):1–17, 1996.
[35] R. Metzler and J. Klafter. The random walk's guide to anomalous diffusion: a fractional dynamics approach. *Phys. Rep.*, 339(1):1–77, 2000.
[36] K. Mustapha. FEM for time-fractional diffusion equations, novel optimal error analyses. *Math. Comp.*, 87(313):2259–2272, 2018.
[37] K. Mustapha, B. Abdallah, and K. M. Furati. A discontinuous Petrov–Galerkin method for time-fractional diffusion equations. *SIAM J. Numer. Anal.*, 52:2512–2529, 2014.
[38] K. Mustapha and W. McLean. Uniform convergence for a discontinuous Galerkin, time-stepping method applied to a fractional diffusion equation. *IMA J. Numer. Anal.*, 32:906–925, 2011.
[39] R. R. Nigmatullin. The realization of the generalized transfer equation in a medium with fractal geometry. *Phys. Stat. Solid. B*, 133(1):425–430, 1986.
[40] I. Podlubny. *Fractional differential equations*, volume 198 of *Mathematics in Science and Engineering*. Academic Press, Inc., San Diego, CA, 1999.
[41] F. Rathgeber, D. A. Ham, L. Mitchell, M. Lange, F. Luporini, A. T. T. Mcrae, G.-T. Bercea, G. R. Markall, and P. H. J. Kelly. Firedrake: Automating the finite element method by composing abstractions. *ACM Transactions on Mathematical Software*, 43(3):1–27, 2017.
[42] K. Sakamoto and M. Yamamoto. Initial value/boundary value problems for fractional diffusion-wave equations and applications to some inverse problems. *J. Math. Anal. Appl.*, 382(1):426–447, 2011.
[43] E. Sousa. How to approximate the fractional derivative of order $1 < \alpha \leq 2$. *Internat. J. Bifur. Chaos Appl. Sci. Engrg.*, 22(4):1250075, 13, 2012.
[44] M. Stynes, E. O'Riordan, and J. L. Gracia. Error analysis of a finite difference method on graded meshes for a time-fractional diffusion equation. *SIAM J. Numer. Anal.*, 55:1057–1079, 2017.
[45] V. Thomée. *Galerkin Finite Element Methods for Parabolic Problems*. Springer-Verlag, Berlin, second edition, 2006.
[46] L. von Wolfersdorf. On identification of memory kernels in linear theory of heat conduction. *Math. Methods Appl. Sci.*, 17(12):919–932, 1994.
[47] K. Wang and Z. Zhou. High-order time stepping schemes for semilinear subdiffusion equations. *SIAM J. Numer. Anal.*, 58(6):3226–3250, 2020.
[48] L. Wang and J. Liu. Total variation regularization for a backward time-fractional diffusion problem. *Inverse Problems*, 29(11):115013, 22, 2013.
[49] Z. Zhang and Z. Zhou. Numerical analysis of backward subdiffusion problems. *Inverse Problems*, 36(10):105006, 27, 2020.

Department of Applied Mathematics, The Hong Kong Polytechnic University, Kowloon, Hong Kong
Email address: bygli@polyu.edu.hk

Department of Applied Mathematics, The Hong Kong Polytechnic University, Kowloon, Hong Kong
Email address: zongze.yang@polyu.edu.hk

Department of Applied Mathematics, The Hong Kong Polytechnic University, Kowloon, Hong Kong
Email address: zhizhou@polyu.edu.hk